Non-Robotic Science Autonomy Development

A white paper for the 2023-2032 Planetary Science and Astrobiology Decadal Survey

Lead Author:

Bethany P. Theiling, NASA Goddard Space Flight Center, 1-301-614-6909,
bethany.p.theiling@nasa.gov

Co-authors:

William Brinckerhoff¹, Julie Castillo-Rogez², Luoth Chou^{1,3}, Victoria Da Poian^{1,4},
Heather Graham^{1,5}, Sona S. Hosseini², Eric Lyness^{1,6}, James MacKinnon¹, Marc Neveu^{1,7},
Kaizad Raimalwala⁸, Barbara Thompson¹

Endorsers:

Rashied Amini², Ricardo Arevalo⁷, Abigail Azari⁹, Edward Cloutis¹⁰, Jamie Cook¹, Ryan M.
Danell^{1,11}, John C. Day², Caitriona M. Jackman¹², Kelly E. Miller¹³, Gordon Osinski¹⁴, Chad
Pozarycki^{1,15}, Lynnae Quick¹, Pablo Sobron¹⁶, Jennifer Stern¹, Melissa Trainer¹

Affiliations

¹NASA Goddard Space Flight Center, ²NASA-JPL/Caltech, ³Georgetown University, ⁴Southeastern
Universities Research Association, ⁵Catholic University of America, ⁶Microtel LLC, ⁷University of
Maryland College Park, ⁸Mission Control Space Services, Inc., ⁹Space Sciences Laboratory,
University of California, Berkeley, ¹⁰University of Winnipeg, ¹¹Danell Consulting, Inc., ¹²Dublin
Institute for Advanced Studies, Dublin, Ireland, ¹³Southwest Research Institute, ¹⁴University of
Western Ontario, ¹⁵Center for Research and Exploration in Space Science and Technology II
(CRESST II), ¹⁶Impossible Sensing

Key Points

- 1) Future planetary missions, especially those to the outer solar system, face significant challenges to increase sampling, reduce uncertainties, and manage and transmit increasing data volumes with limited data link rates.
- 2) Autonomous platforms enable instruments to perform inter-calibrations, sample validation, and discriminate data transmission, which reduces measurement uncertainties for optimal science return.
- 3) Science autonomy is necessary to achieve science goals for planetary missions under extreme conditions, short mission timeframes, and long delays in communication, and is specifically discussed in 3 of the Planetary Mission Concept Studies informing this Decadal Survey and the proposed Europa Lander.
- 4) Science autonomy will enable missions that are otherwise not possible, such as a sub-ice shell ocean submersible or a Venus lander.

1. Motivation for Advancing Science Autonomy

Significant advancements in autonomy have been made across both current and planned Planetary Science, Heliophysics, Astrophysics, and Earth Science missions over the past decade. The 2015 NASA Technology Roadmaps detail the need for advancing autonomy to support future missions ranging from Discovery to Flagship classes, and to planetary targets ranging from the Moon to Mars, Venus, and Europa. These mission projections draw from recommendations for advancing mission autonomy made throughout the 2013-2022 Planetary Decadal Survey.

Much of the mission autonomy development before and since these reports has focused on *robotic autonomy*, the onboard processing of raw or low-level data products that enables a spacecraft and/or flight instrument(s) to proceed safely and efficiently with mission objectives using minimal human interaction [1]. **We use the term *non-robotic science autonomy* to refer to the ability of a science instrument to analyze its own data in order to calibrate itself, optimize operational parameters based on real-time findings, and ultimately make mission-level decisions based on scientific observations and determine which data products to prioritize and send back first.** Science autonomy also includes data processing software that could be used for rapid data interpretation by scientists. Four of the Planetary Mission Concept Studies (PMCS) in preparation for this Decadal Survey, as well as the Europa Lander concept, target planetary environments from which communications are limited in data link rates and/or in time, including outer solar system targets, subsurface oceans, and hot and/or highly irradiated surfaces. Indeed, 4 of these mission concepts, including the PMCS Mercury lander, Intrepid moon rover, and Venus Flagship, as well as the Europa Lander concept describe needs for autonomy. Beyond the coming decade, future submarine missions beneath ice shells of ocean worlds would only be enabled by the ability to operate and make decisions autonomously.

There are two broad categories of non-robotic science autonomy: flight instrument autonomy and data interpretation autonomy. Flight instrument autonomy deals with an instrument's ability to **autonomously collect** and selectively **transmit** data to Earth. Instruments capable of autonomous data collection, both robotically and in terms of decision-making (what samples to analyze, when, for how long, and fidelity of transmitted data) would, for example, greatly enhance the science return for missions in extreme environments, and are being planned for e.g., the proposed Europa Lander mission [2]. Autonomy in terms of data transmission would address the

challenge of instruments generating far more raw data than can be downlinked to Earth. Two current strategies for reducing transmission volumes are compression and segmentation. High compression ratios can introduce severe artifacts. Subtle features in e.g., spectra can be lost, and if the most important or valued measurements are only a small fraction of the total data collected, then compressing and preserving the whole population is often done at the expense of the vital sub-population. Segmentation (using an algorithm to separate out a key segment or subset of data) requires schemes that reliably do not remove critical data. **Therefore, autonomous instruments must be developed that can discriminate for high-reward data to be transmitted.** Machine learning can be used to develop intelligent algorithms that are able to identify and transmit the

Figure 1. Target selections by AEGIS for MSL's ChemCam identify top-ranked analytical targets in green, reject blue-outlined targets, and retain red-outlined targets. [3]

most interesting data first. Such intelligent instruments may enable development and flight of instruments and analytical techniques that had previously been ruled out as too data-intensive. However, development of intelligent instruments will require analog planetary data collection, training data sets, and rigorous testing of the algorithm(s) using an instrument of equivalent technology readiness level. The second category of non-robotic science autonomy focuses on data interpretation. Currently, much of the mission data collected requires some level of processing and interpretation by individuals or working groups of engineers and/or scientists. However, such methods require significant personnel time and work efforts by one or more experts, which may not always be feasibly supported throughout the lifetime of a mission. The Autonomous Exploration for Gathering Increased Science (AEGIS) is a current use of machine learning algorithms to automate interpretation of visual images to assist in subsequent sample selection for the ChemCam instrument onboard the Mars Science Laboratory (MSL) (Figure 1) [3], which has resulted in a significant increase in sampling and analysis (Figure 3). We discuss this case study in Section 3. While data processing and analysis by experts is necessary for scientific advancement (and discoveries continue for years beyond a prime mission), science return would be enhanced by automating certain tasks.

Both types of science autonomy are challenged and motivated by ever-increasing data volumes as instruments evolve. More mature and sophisticated instruments have been / are being developed to answer and satisfy the ever more challenging data processing requirements growing along with mission ambitions. As an example, data volume for mass spectrometers is growing by orders of magnitude over the past decade (Table 1), while data transmission rates will increase by at most one order of magnitude due to fundamental limits of physics. This difference between anticipated data volume and transmission implies that as much as 90% of the data generated by e.g., mass spectrometers on future missions would not be transmitted to Earth.

2. Benefits of Science Autonomy

Instrumental autonomy for sample selection and certain routine science measurement tasks (traditionally done by humans) could not only improve sampling cadence for remote and *in situ* planetary missions, but also would enable science activities in locations where explicit

Table 1. Mass spectrometer data volumes for various missions over time.

Mass spectrometer configuration	Mission	Approx. data points/second
Quadrupole	MSL	50
Ion trap	ExoMars	50,000
Orbitrap	Future	5,000,000

human direction is difficult or impossible. This capability will be necessary as we plan missions deeper in the Solar System and/or in extreme environments (e.g., sub-ice shell oceans), where data transfer rates are substantially outpaced by data volumes, making supervision and planning of every analysis increasingly challenging. Science autonomy could also open new capabilities for short-lived missions such as Venus surface operations (as brief as a few hours) or atmospheric decent probes.

The potential benefits of science autonomy could be roughly quantified, in part, by its impact on ‘data economy’ (minimizing cost-per-byte) for data downlink, which is substantially higher for outer planets missions because the cost increases with distance from Earth or a relay station. This cost-effectiveness related to science autonomy (data economy) across missions to Venus, Earth, the Moon, Mars, Ocean Worlds (e.g., Europa and Enceladus), small solar system bodies, and the heliosphere was discussed in detail at the *2018 Workshop on Autonomy for Future NASA Science Missions*. In addition to the substantial distance to outer planets and their moons, outer planets missions require massive and energy-consuming antennae to transmit data, especially in the absence of relay stations. Further, outer planet spacecraft often have narrow communication windows, which may be complicated by powerful radiation environments (e.g., Jupiter’s magnetic field). On an instrument scale: The Mars Organic Molecule Analyzer (MOMA) onboard the ExoMars Rover will obtain ~1000 mass spectra per experiment on Mars, constrained by the maximum amount of data that can be sent back to Earth for a particular sol. The same instrument deployed to a more distant world would return far fewer spectra due to downlink rates and limited communication windows. However, if an instrument could prioritize transmitting the most interesting results first (~top 10%), the data economy of the mission would increase by $\approx 10\times$. For most mission scenarios to the outer solar system, it will be impractical or even impossible to send all the data back to Earth. Increased autonomy would enable instruments to discriminate between e.g., calibration data, data below detection limits, data containing instrumental artifacts, and the most interesting data (which would at least demonstrate measurement of some parameter above the detection limit). **Onboard autonomy algorithms can be informed by laboratory analog-based analyses on engineering test units (analogous to a flight instrument) using a comprehensive (and unbiased) dataset for the training phase and cross-validation with a known dataset during the testing phase.**

3. Science Autonomy in Current and Future Planetary Science Missions

Efforts to increase science autonomy in terms of sampling, data transmission, and data interpretation have had some investment, but little infusion, relative to robotic autonomy. Most notably, current efforts to develop science autonomy relevant to outer planets missions are lacking due to the substantially smaller data volumes that can be afforded during the mission. Most autonomy development uses machine learning to recognize patterns and develop predictive algorithms for data, while attempting to maintain the fidelity (low error) of the original data. These machine learning training efforts (e.g., supervised and reinforcement learning) typically require massive data sets and thus, their development for outer planets missions has received less attention due to the smaller data volumes expected. While small data volumes are riskier for machine learning and data-training, efforts to train small datasets more characteristic of outer solar system missions will ultimately be required for successful science return from those missions. Recent advances in data science have already begun to refine machine learning algorithms for smaller datasets [4]. **As machine learning tools and predictive algorithms advance, mission concepts and science goals previously considered too risky or impossible due to data concerns or instrument power can be explored [5].** In the following discussion, we illustrate several examples of current advances in

science autonomy and directions for future advancement of science autonomy related to inner and outer solar system targets.

3.1. Inner Solar System Targets

3.1.1. The Moon

Short durations for lunar surface missions (hours-weeks) require operations teams to maximize science return with limited data and/or sample collection under tight downlink or timing constraints, necessitating rapid tactical decision-making processes. As an example, autonomous geologic scene classification and novelty detection would increase the chances of detecting valuable novel / sparse features (e.g., meteorites, outcrop, pyroclastic deposits, secondary craters) from visual imagery that may otherwise be missed in scenarios that prioritize driving and other mission needs. A rover equipped with this capability could identify features of interest from its multi-spectral imager navigation camera, and command the spectral imager to scan that feature without time-consuming human-initiated targeting decisions. The Autonomous Soil Assessment System – Contextualizing Rocks, Anomalies and Terrains in Exploratory Robotic Science (ASAS-CRATERS), currently in development, uses deep learning models to classify terrain and detect novel geological features by using only a rover’s navigation sensors to enable improved geologic contextualization and analysis (Figure 2) [6]. The ASAS-CRATERS system can be used to autonomously identify, rank, and select features of interest for onboard instrument targeting, and/or downlink prioritization. This level of autonomous processing and decision-making can also be applied to data from specific instruments, such as classification of spectra or spectral images, or identification of anomalous patterns in mass spectrometry. Techniques like Bayesian inference can encode scientific hypotheses to enable intelligent decision-making based on results of this autonomous analysis [7, 8].

Figure 2. Labelled images demonstrating a mock-up of possible ASAS-CRATERS classification scheme for lunar surface features. [6] Images from the Yutu-1 PCAM dataset.

3.1.2. Mars

Mars missions have benefited from short communication times with Earth (7 min delay), solar power, and high data transmission to Earth due to the relay capabilities of Mars’s orbiting satellites. There are several motivating factors that drive the need for autonomy in Mars science operations. In traditional Mars rover operations, visual surface characterization and subsequent analysis and decision-making takes place in day-long tactical cycles [9]. Therefore, Mars operational and science teams can analyze the data and update operations based on that data within a sol. However, decision-making necessarily requires a team (from the Payload Downlink Lead (PDL) to the Science Operations Working Group (SOWG)) dedicated entirely to this purpose; each proposed operational change would then require approvals across the mission organization before proceeding. In contrast, autonomous capabilities

Figure 3. Deployment of AEGIS to ChemCam sampling targets on MSL resulted in a significant increase in sampling and data collection [3].

would enable real-time decision-making onboard the instrument that work toward achieving established science goals, while removing some level of need for human interaction. This allows for less demanding constraints on the operational and science teams, and would increase the number of useful data collected and transmitted. Current Mars missions have already enhanced their autonomous capabilities by using predictive algorithms based on machine learning. The AEGIS system is a recent example of the increased science return resulting from science autonomy development [3]. The AEGIS software, which was developed for the Mars Exploration Rover (MER) Opportunity in 2010 to autonomously select targets for detailed imaging, has been modified for use as a target selection tool for ChemCam on MSL. AEGIS uses “intelligent selection” to prioritize targets for chemical analysis (Figure 1) and early results demonstrate a significant increase in the targets analyzed after deployment (Figure 3) [3].

Science autonomy would also enhance surface reconnaissance prior to human exploration. While the bandwidth to Mars is relatively high, the Mars Design Reference Mission conceptualized at the *2018 Workshop on Autonomy for Future NASA Science Missions* suggested that future missions would consist of many small rovers in order to make the necessary survey of a potential astronaut landing zone. However, a fleet of rovers and other landed craft would exceed the bandwidth available at Mars, and a dedicated science and engineering team for each rover would be prohibitively expensive and time-consuming. Therefore, each rover or a fixed base lander would require science autonomy software to efficiently survey the area necessary and provide the science team with only the most relevant and important data.

3.2. Outer Solar System Targets (Ocean Worlds)

3.2.1. Europa

Planned and potential missions to Europa must grapple with longer communication delays due to the distance from Earth, orbital pathways, and the lack of communication infrastructure, as well as a highly radiative environment due to Jupiter’s powerful magnetosphere. The radiation environment significantly reduces mission timelines; 3.5 years for the Europa Clipper flyby mission, and 16 days (22 days contingency) for the proposed Europa Lander mission. These challenges significantly increase the cost per byte and necessitate plans for increasing data economy. Nevertheless, the potential for Europa to possess a habitable ocean and possibly evidence of current or extant life drives mission and technology development. For example, expected data rates for instruments on board the Europa Lander are expected to range from 1 kbit per second for long duration monitoring to tens of kbits per second for triggered high resolution measurements. A preliminary concept of operations from e.g., the ultrahigh resolution Characterization of Ocean Residues and Life Signatures (CORALS) mass spectrometer would expect 3 to 4 GB of data volume produced per sample (with little to no data compression). Because of the large data volume of the CORALS instrument, each sample spectrum would require substantial data reduction, pre-processing, and compression before transmitting to Earth. Data compression could result in substantial reduction of the data quality and mass resolution, each of which are critical features of the CORALS design. Therefore, any loss in data quality due to compression negates the significant analytical advantage of the CORALS instrument over other lower-resolution mass spectrometers. The Europa Lander Science Definition Team (SDT) report [10] indicates that “sample acquisition is anticipated to last 5 hours” and that “the science and engineering teams have 8 and 16 hours, respectively, to plan and generate commands for the subsequent [24 hrs], which includes making the decision on the activities for the next [24 hrs].” A single Earth day therefore could result in Gb of raw, unprocessed data, transmitting at a rate of 80 kbps that would need to be processed and analyzed before science and engineering teams could make operational decisions and send updated commands. **As a result, the**

short mission duration, abbreviated analysis times, and large data volumes necessitate development and deployment of science autonomy to maximize science return.

Figure 4. Imaging coverage for Europa’s surface using the EIS NAC (Europa Clipper illustrates >95% mapping at ≤ 50 m/pixel. [11]

m/pixel, with some areas at ≤ 1 m/pixel. Machine learning could be used to predict surface characteristics, including topographic relief, for areas with lower resolution imaging (Figure 4: [11]). Not only does this advance surface navigation and landing site selection for a lander mission, this would also advance knowledge of the distribution and formation of cryotectonic and cryovolcanic features on Europa’s surface.

3.2.2. *Enceladus*

The Enceladus Orbilander PMCS illustrates a potential data return scenario from the Saturn system. In this concept, the instruments that generate the most data are the radar and reconnaissance imagers during the mission’s orbital phase (125 Gb for the imager), nanopore sequencer during the landed phase (4 Gb), and to a lesser extent two mass spectrometers during both mission phases (≈ 0.5 -1 Gb each). Communications can take place at ≈ 35 kbps, 12h/Earth day while in orbit and ≈ 3 h/day once landed. The Orbilander mission concept is not data-constrained due to several-fold schedule margins during the 3-year prime science phase at Enceladus, which allows for ample communication time. However, the schedule margin is critical because the rate of data return is lower than that of data generation. In orbit, the spacecraft flies through the plume over the south polar terrain every 12 hours. Assuming a radar data generation rate equivalent to that of REASON onboard Europa Clipper of 24 Gb/flyby yields a data generation: return ratio of $\approx (24 \times 10^9 \text{ bits/flyby} \times 2 \text{ flybys/day}) / (35000 \text{ bps} \times 12 \text{ h communication/day} \times 3600 \text{ seconds/hour}) = 30$, neglecting all other data. Estimates on the reconnaissance imaging generation: return ratio are substantially lower (0.45) (see Enceladus PMCS by MacKenzie et al.).

A separate architecture considered during the study, though initially promising, failed to close due to unmanageable data constraints, which overall could benefit from science autonomy. That architecture involved a decimeter-scale, battery-powered lander solely carrying a nanopore sequencer to attempt to confirm any biosignatures detected in the plume sampled by its carrier relay orbiter. (The sequencer must be run from the surface as it requires more sample than can be gathered in months of plume fly-throughs.) Although mass, power, sample acquisition, and analysis time could all be accommodated during the Europa Lander-like length of its landed mission, the 4 Gb of data generated could not be returned fast enough with the lander’s antenna. It may be that developments in science autonomy enable such an architecture in the future.

4. Concluding Remarks

New frontiers of scientific exploration bring new challenges. In addition to the well-known mechanical and operational challenges, there are also constraints posed by data collection and transmission. Many missions have severe limitations on the volume of data that can be transmitted, requiring innovative new strategies to optimize data economy. Fortunately, there are methods that go beyond the classical compression and data partitioning schemes; novel machine learning methods can form the basis of an onboard data budget by making informed decisions based on the situation. Science autonomy can also involve prioritization of targets and data collection. The key challenges of data collection and transmission are ideally addressed in concert, as they are not fully separable problems. They require us to use all of the knowledge we have from prior studies, as well as an assessment of all the potential characteristics of the environment to be studied.

Science autonomy expands our mission discovery capabilities; in some cases, it allows us to access new scientific information and explore frontiers that would not be possible otherwise. However, a successful science autonomy strategy requires investment and development to ensure that the methods are robust and reliable. The following recommendations ensure that state-of-the-art methods are integrated with physical knowledge so we take full advantage of exciting new frontiers of exploration.

Recommendations for Advancing Science Autonomy

Based on the discussion above, we make the following recommendations:

- 1) Support technology development in conjunction with machine learning to determine whether software can be developed to support more accurate data compression and/or segmentation, lower data transfer requirements, and discrimination of calibration data or data that is lower than detection limits from data above detection limits.
- 2) Support laboratory analog testing of technology in conjunction with machine learning applications to determine whether software can be developed that accurately identifies chemical compositions.
- 3) Support field testing of technology in conjunction with machine learning applications to determine whether software can be developed that accurately classifies visual images, chemical compositions, and planetary features such as (but not limited to) seismicity, magnetism, and radiation in Earth-analog environments.
- 4) Development of open-access training datasets to support machine learning efforts.

References

[1] Gao and Chien, 2017. Science Robotics 2: eaan5074. [2] Tan-Wang & Sell, 2019. AbSciCon Abstracts with Programs. [3] Francis et al., 2017. Space Robotics 2: eaan4582. [4] Li et al., 2019. Int. J. Machine Learning and Cybernetics 10.10: 2805-2822. [5] Azari et al., 2020. Integrating machine learning for planetary science: perspectives for the next decade. White Paper for the Planetary Science and Astrobiology Decadal Survey 2023-2032. [6] Raimalwala et al., 2020. Lunar Surface Science Workshop 2020, Denver, CO. [7] Arora et al., 2017. ArXiv:1712.09716 [Cs]. <http://arxiv.org/abs/1712.09716> [8] Candela et al., 2017. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), 3811-3818. [9] Gaines et al., 2016. Proc. 4th Workshop on Planning and Robotics, 115-125. [10] Hand et al., 2017. Europa Lander Science Definition Team. [11] Turtle et al., 2019. LPSC Abstracts with Programs.