1. **Presentation Title: Rapid Data Understanding Using the Complex Data Explorer (CODEX)**
2. **Presenter:** 
   * *Jack Lightholder, Data Scientist, JPL, jack.a.lightholder@jpl.nasa.gov, (908) 268-5006*
3. **Short Biography of Presenter:** *(100-200 Words)*

Jack Lightholder is a data scientist at NASA’s Jet Propulsion Laboratory in Pasadena, California. As a member of the Machine Learning and Instrument Autonomy Group, Jack supports a variety of missions in concept, development and operational phases. Jack is the principal investigator for The Operational Recommendations for Capturing History and Infusing Data Science (ORCHIDS) system. Previously, Jack lead the Dust Devils Microgravity Research Group, an astrophysics research group investigating the phenomenon of dust electrification and its relation to coagulation of dust particles. The group was awarded contracts through NASA’s Reduced Gravity office to conduct microgravity research flights to further study the phenomena.

1. **Other Authors:**

* *Josh Rodriguez, Front End Developer, JPL,* [Joshua.D.Rodriguez@jpl.nasa.gov](mailto:Joshua.D.Rodriguez@jpl.nasa.gov), (818) 393-4765
* Rob Tapella, User Interface Designer, JPL, [Robert.k.tapella@jpl.nasa.gov](mailto:Robert.k.tapella@jpl.nasa.gov), (818) 393-8030
* Lukas Mandrake, Data Scientist, JPL, [mandrake@jpl.nasa.gov](mailto:mandrake@jpl.nasa.gov), (818) 354-1705

1. **Point-of-Contact if other than Presenter:**

**N/A**

1. **Keywords:** *(short list of focus areas addressed in your presentation)*

***Data Exploration, Mission Operations, Machine Learning, Data Science, Telemetry, Science Data System, Anomaly Detection***

1. **Has this abstract been approved by your organization? \***

**Yes**

1. **If accepted, will you allow your presentation to be included on the GSAW website?**

**Yes**

1. **Abstract:** *(200 Words to Two Pages)*

Modern science datasets from missions like OCO-3 and telemetry records in operations may have 500+ simultaneous measurements at each of millions of time samples. Scientists would often like to look through the record and discover not only expected trends but ones they did not initially guess, while mission operations personnel perform the same task under serious time pressure, should an anomaly occur. In both cases, the optimal environment for this rapid exploration large data would be one where visualizations were clear, interactive, and responsive, permitting the investigator to “play” with the data and gain rapid insight, falsify hypothesis, and make discoveries. Machine Learning (ML) has proven invaluable in providing some of these key data insights, but to do so in a statistically robust and reliable manner requires a data science professional and a lot of custom Python code, losing any sense of interaction and play. CODEX addresses these concerns by providing a desktop-like environment with standard scientific graph types that are robust to rapid, powerful exploration. For example, all graphs are linked such that selecting data on one graph reveals those same points in all others regardless of type including histograms and heat maps. Machine Learning comes into play as well, allowing users to request high-level directions like “Show me more like this region,” “How many groups best describe all of these observations,” and “Predict this value based on these 500 other values, and show me the small set of variables that were responsible for 90% of the prediction.” All of this power normally requires considerable tuning of complex parameters in the ML algorithms; CODEX achieves this tuning visually, by showing demonstrations of what would happen for various hyper parameters and permitting the user to select visually the desired outcome. These capabilities significantly shorten analysis time, leading to richer human understanding of complex data sets, without writing a line of code.

\* Aerospace Corporation employees are required to have abstracts and presentations cleared through the Office of Technical Relations (OTR) process