### A Future Data Analysis Scenario

1. A scientist has a hypothesis and wishes to test it.
2. The relevant data sets are already on the ground, having been obtained by NASA spacecraft and Earth-based facilities over a time span of many decades.
3. The scientist quickly codes up a data processing algorithm, building upon a large library of tested, documented, open-source building blocks contributed by her fellow scientists.
4. She prepares a virtual machine (VM) that contains the program and any relevant support files, and submits the VM to a cloud computing service.
5. An arbitrary number of instances of the VM execute in the cloud, assimilate the results, and deliver them back to her quickly.
6. Using visualization tools based on other open-source algorithms, she explores the results and reaches her conclusions.
7. She publishes her results as a journal article, plus supplemental data and VMs. Results are immediately available to the rest of the community.

#### Approaches

- Develop automated tools and templates that ensure every data set is accompanied by suitable documentation and metadata.
- Employ intelligent “scraping” of new publications.
- Widespread availability of open-source software will prevent NASA-funded scientists from constantly needing to “re-invent the wheel”.
- Eventually, the accumulated tools will allow most scientists to focus on research, not programming.
- NASA should continue to encourage the sharing of open source software by its scientists.
- NASA should reward (and fund) scientists for well-designed, well-documented libraries.
- Scientists have no incentive to share their tools if publications are the only metric of success.
- We should consider identifying a limited set of “recommended” programming languages.
- Discourage or perhaps even forbid the use of proprietary languages (e.g., IDL and Matlab).
- Long-term costs to NASA are very high.
- Proprietary languages cannot be used in VMs.
- Proprietary software is limited in its flexibility.
- Find ways to harness the latest developments in computing, e.g., neural nets, GPU programming, and whatever else may be coming in the decades ahead.

#### Lessons Learned 1979–2016 and a Vision for 2020–2050

**Archival Data and Computational Power in Planetary Astronomy:**

Mark R. Showalter, Matthew S. Tiscareno, and Robert S. French, SETI Institute, Mountain View, CA

### Lessons Learned

- **1979**
  - **Hardware Solution (Best):** Develop better spacecraft for Earth-based ground-based observations.
  - **Lesson:** Scientists have no incentive to share their tools.

- **2017**
  - **Cloud Computing:** Has the potential to revolutionize planetary data analysis and modeling.
  - **Lesson:** Highly parallel problems (e.g., process every image that...) can execute very quickly.

- **2050**
  - **Scientific Research Will Always Be a Human-Centered Enterprise.**
  - **HAL Will Not Help Us.**
  - **Our Goal Should Be to Minimize Obstacles Along the Path from Hypothesis to Conclusion.**

### Approaches

- NASA archiving must be comprehensive, including data from Earth-based telescopes, lab experiments and published theoretical models.
- All data sets must include raw and calibrated data, plus a rich and complete set of metadata.
- **Cloud Computing:** Has the potential to revolutionize planetary data analysis and modeling.
- **Lesson:** Highly parallel problems (e.g., process every image that...) can execute very quickly.

**Approaches:**

- **Hardware Solution (Best):** Develop better spacecraft for Earth-based ground-based observations.
- **Lesson:** Scientists have no incentive to share their tools.
- **Cloud Computing:** Has the potential to revolutionize planetary data analysis and modeling.
- **Lesson:** Highly parallel problems (e.g., process every image that...) can execute very quickly.

**Approaches:**

- NASA archiving must be comprehensive, including data from Earth-based telescopes, lab experiments and published theoretical models.
- All data sets must include raw and calibrated data, plus a rich and complete set of metadata.
- **Cloud Computing:** Has the potential to revolutionize planetary data analysis and modeling.
- **Lesson:** Highly parallel problems (e.g., process every image that...) can execute very quickly.

**Approaches:**

- NASA archiving must be comprehensive, including data from Earth-based telescopes, lab experiments and published theoretical models.
- All data sets must include raw and calibrated data, plus a rich and complete set of metadata.
- **Cloud Computing:** Has the potential to revolutionize planetary data analysis and modeling.
- **Lesson:** Highly parallel problems (e.g., process every image that...) can execute very quickly.