



**Climate Program Remote Sensing Workshop
February 12-14, 2018**

SPEAKER TITLES/ABSTRACTS

Maggie Johnson
SAMSI

“A Notional Framework for a Theory of Data Systems”

Modern, large scale data analysis typically involves the use of massive data stored on different computers that do not share the same file system. Computing complex statistical quantities, such as those that characterize spatial or temporal statistical dependence, requires information that crosses the boundaries imposed by this partitioning of the data. To leverage the information in these distributed data sets, analysts are faced with a trade-off between various costs (e.g., computational, transmission, and even the cost building an appropriate data system infrastructure) and inferential uncertainties (bias, variance, etc.) in the estimates produced by the analysis. In this talk we introduce a framework for quantifying this trade-off by optimizing over both statistical and data system design aspects of the problem. We illustrate with a simple example, and discuss how it may be extended to more complex settings.

This is joint work with Amy Braverman (JPL) and Brian Reich (NCSU)