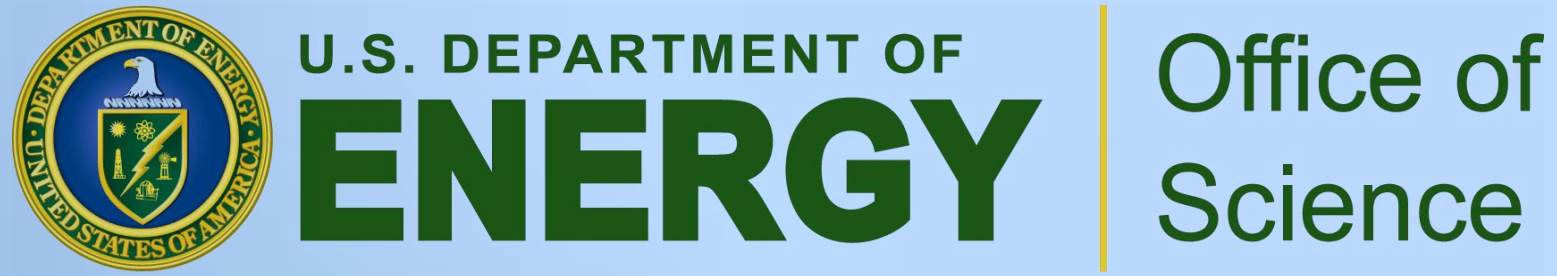


# SURVIVING THE DELUGE (PART I): LEARNING FROM AN ANALYSIS OF DATA EXTRACTION PATTERNS FROM THE ARM CLIMATE RESEARCH FACILITY ARCHIVE

ORNL, ARM Archive: Raymond McCord, Stefanie Hall, Giri Palanisamy, W. Christopher Lenhardt; Mission Research: Sean Moore

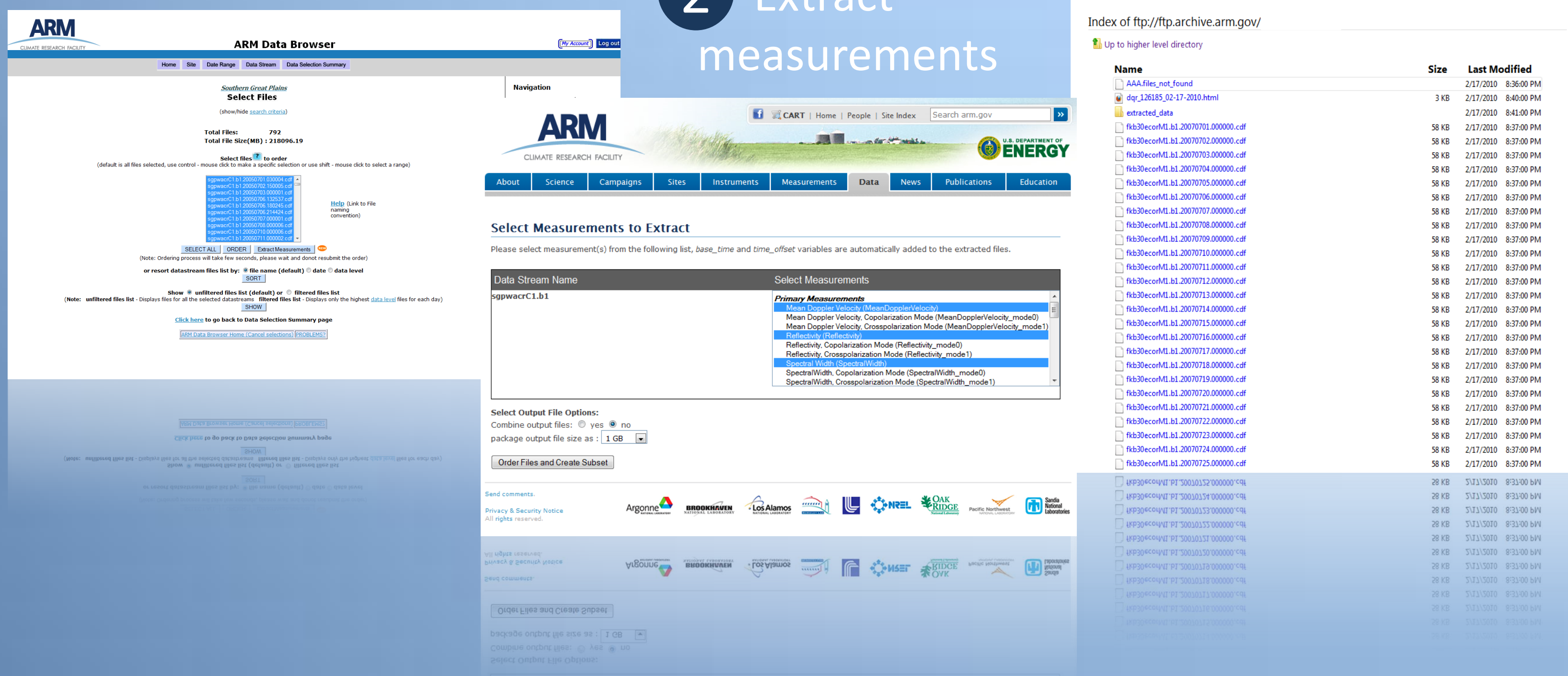


## WHAT IS DATA EXTRACTION? DATA EXTRACTION IS...



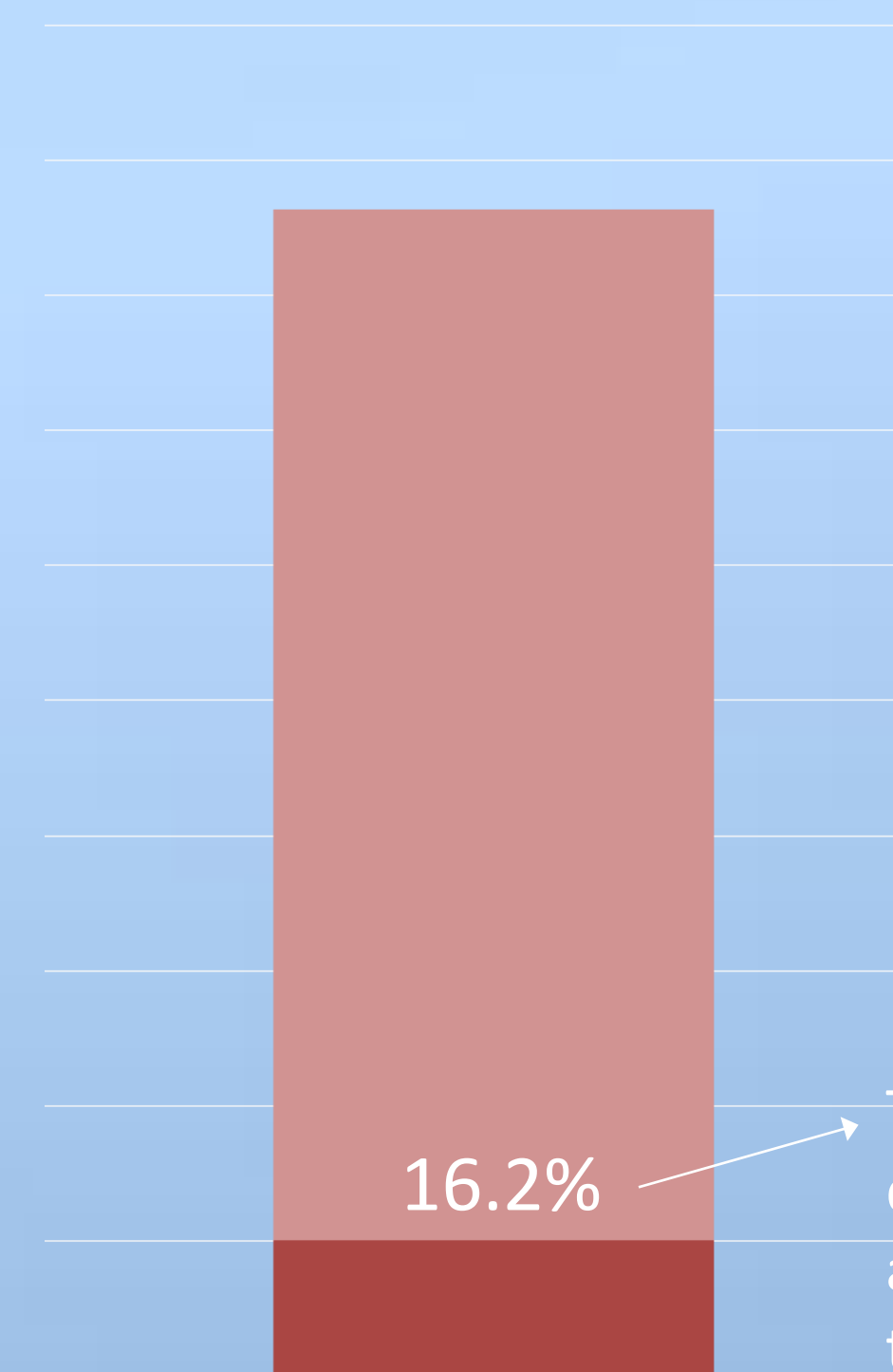
- Easy to use
- Built in to normal file ordering process
- Available from all User Interfaces

- 1 Select files to order
- 2 Extract measurements
- 3 Retrieve extracted data files



- A valuable tool for all data users

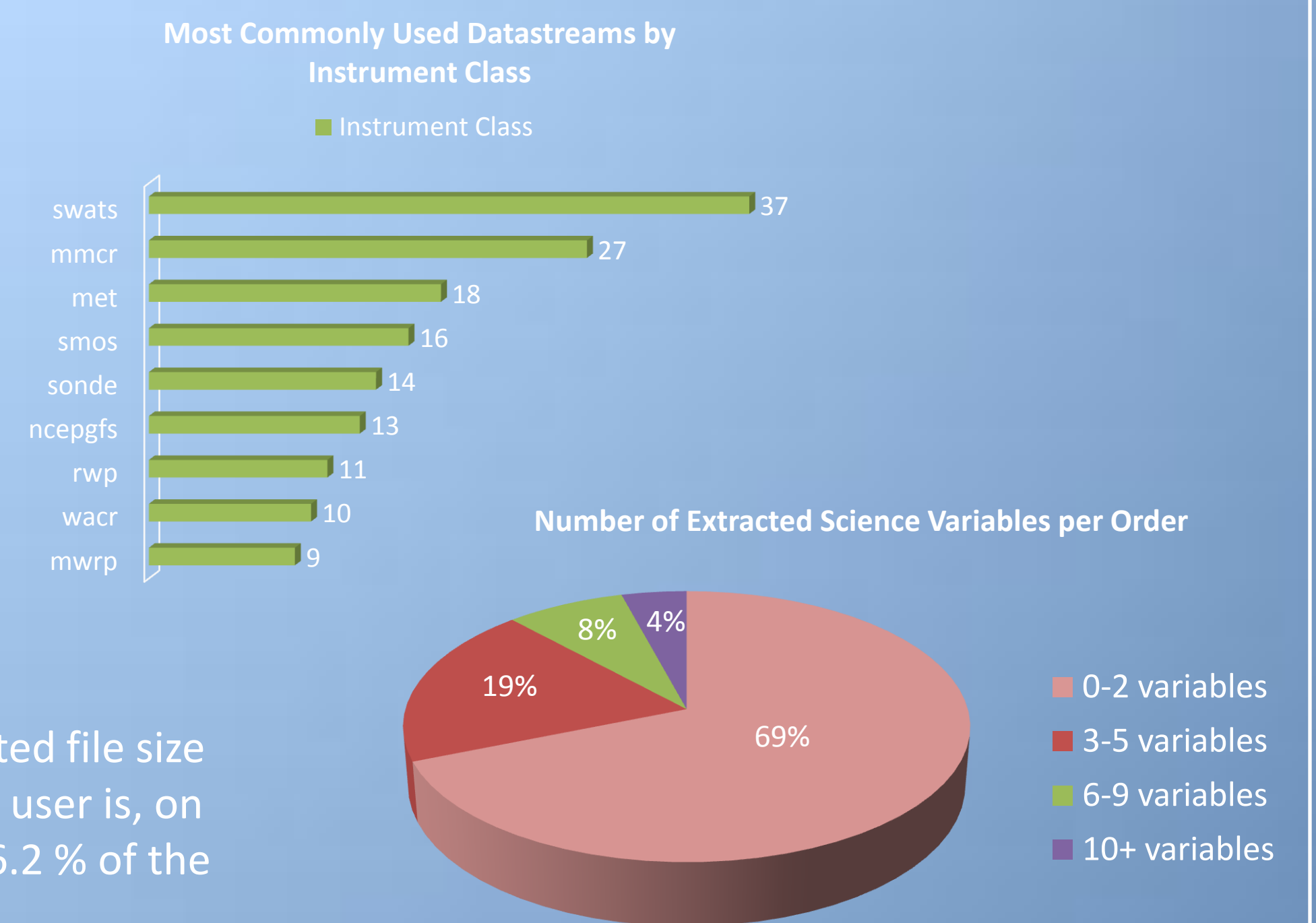
File Size Reduction resulting from Data Extraction



Pre-extraction size (KB)  
Post-extraction size (KB)

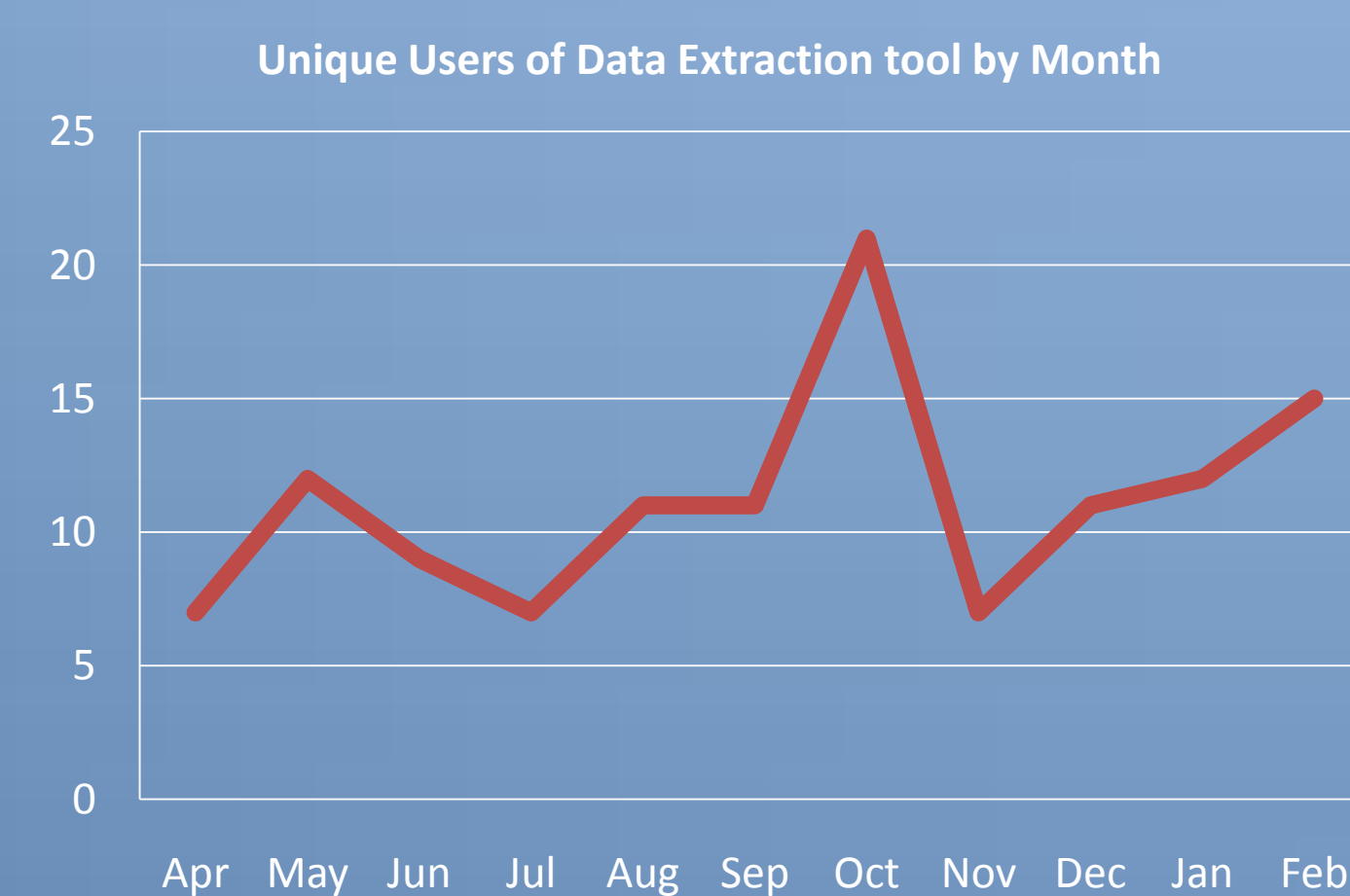
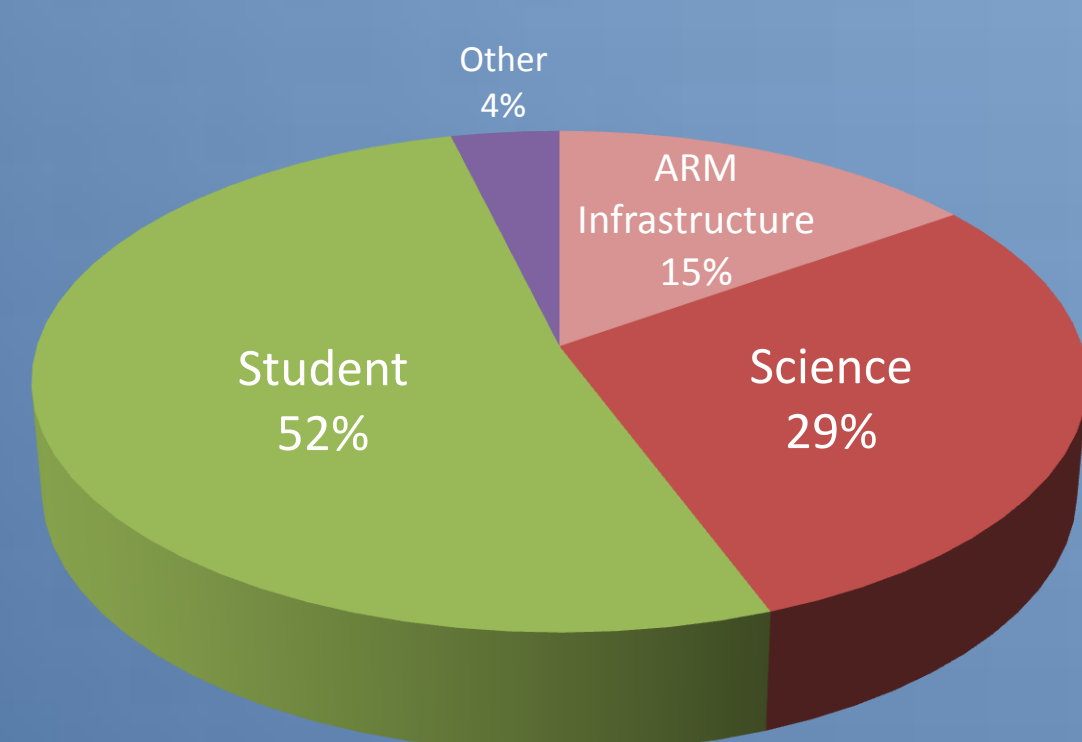
The post-extracted file size delivered to the user is, on average, only 16.2 % of the total file size!

- Intended to reduce the burden of “big” data products



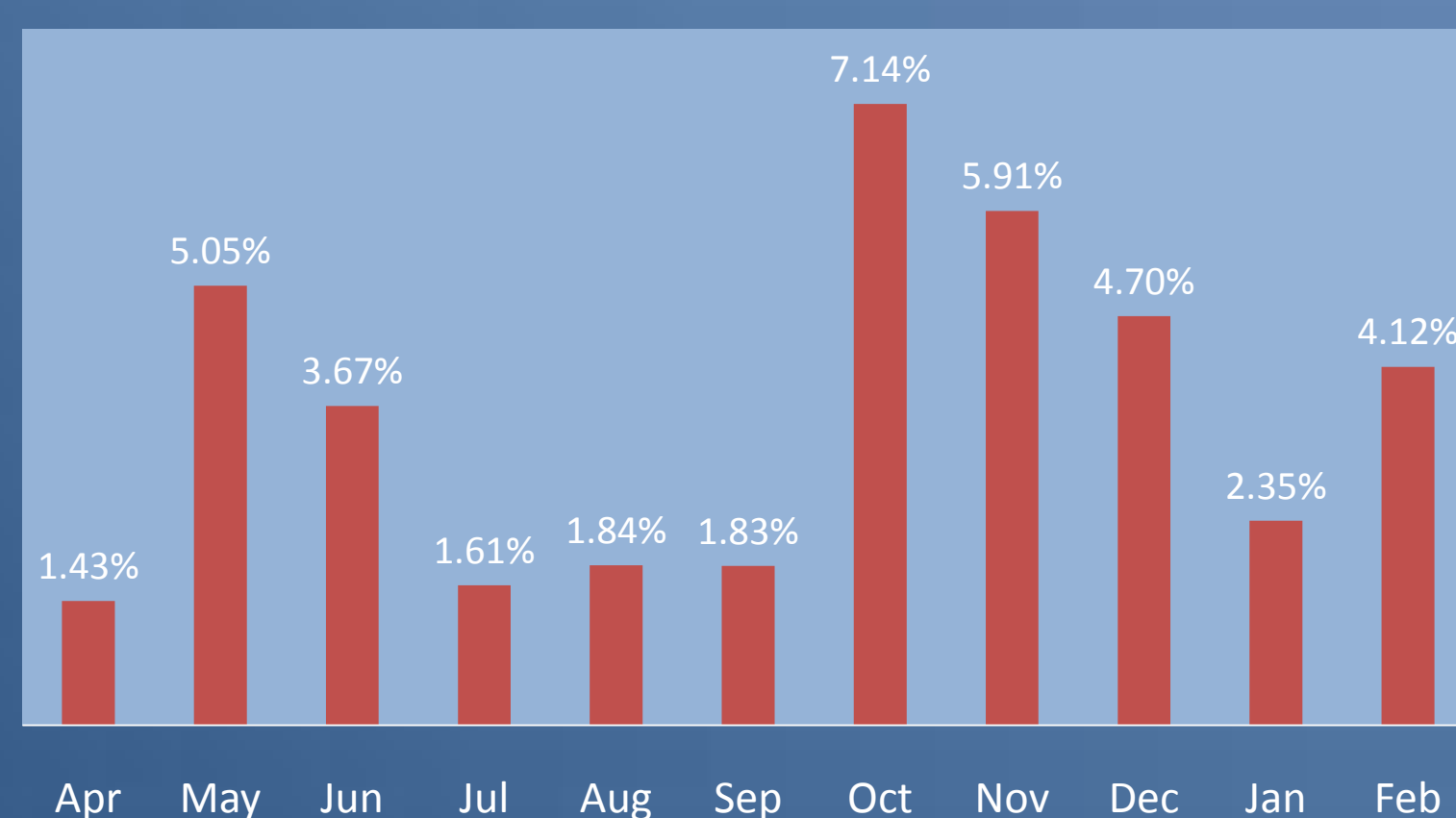
- Most frequently used to extract only a few scientific variables

- Used primarily by students and scientists

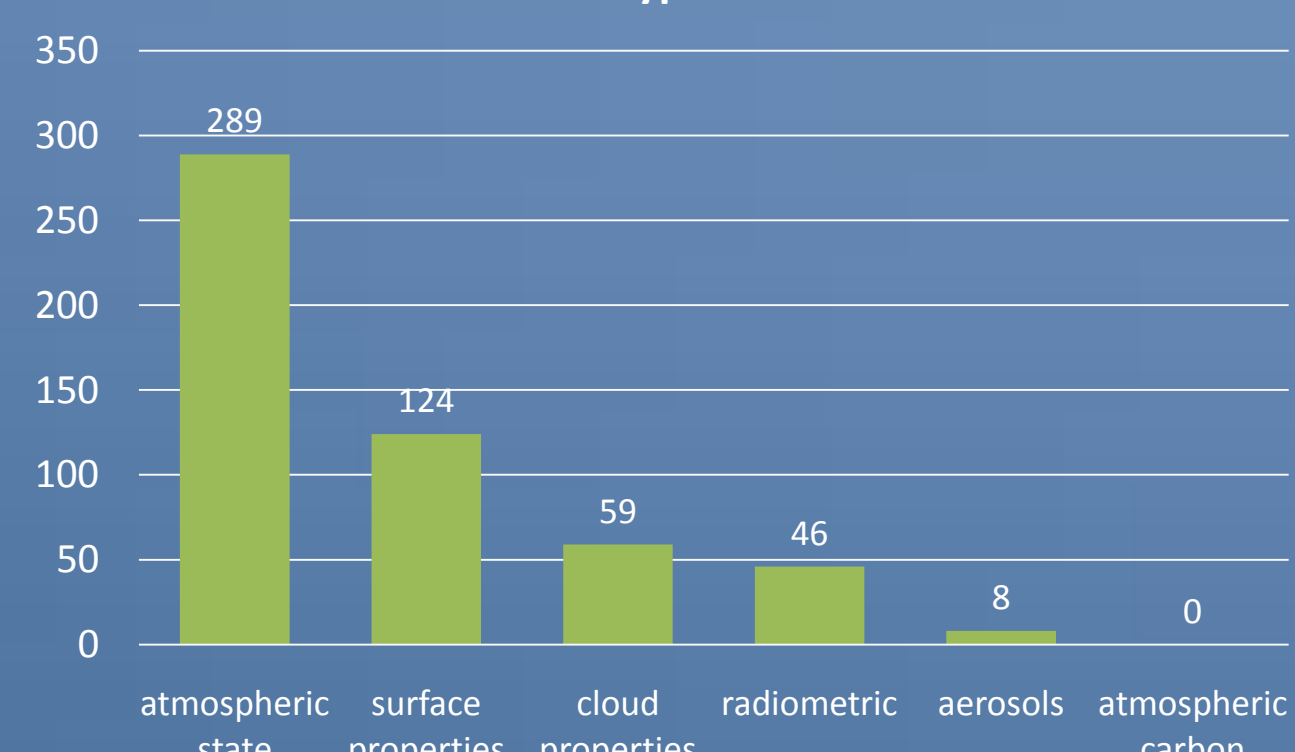


- Under-utilized overall

Percentage of Data Requests using Extraction by Month



Number of Extracted Variables by Measurement Type



- In need of YOUR feedback

**YOU TELL US!**

How can the ARM Archive improve upon the customization of data products?

Sub-sampling / Striding

Multi-input extraction  
Across data streams

User-written extraction  
and analysis programs

Dynamic analysis  
environment

Handwritten feedback: *###*

Handwritten feedback: *###*

Handwritten feedback: *###*

Handwritten feedback: *###*

Other (write in):

Blank area for handwritten feedback.