challenges to Internet research

- domain-specific knowledge required to understand data and format
- limited programming ability by many researchers
- ad-hoc code difficult to reuse

data platform integration

- create a single common interface
- provide APIs rather than flat files
- dynamic documentation/examples (Jupyter Notebooks)