National Snow and Ice Data Center

Advancing knowledge of Earth's frozen regions

CIRES

University of Colorado **Boulder**

Function Follows Form:

A Practical Guide to Research Data Curation

Julia Collins

National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO

UCAR SEA 2016
5 April 2016

National Snow and Ice Data Center

Advancing knowledge of Earth's frozen regions

CIRES

University of Colorado **Boulder**

Function Follows Form:

A Practical Guide to Research Data Curation

Julia Collins

National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO

UCAR SEA 2016
5 April 2016

About the title…

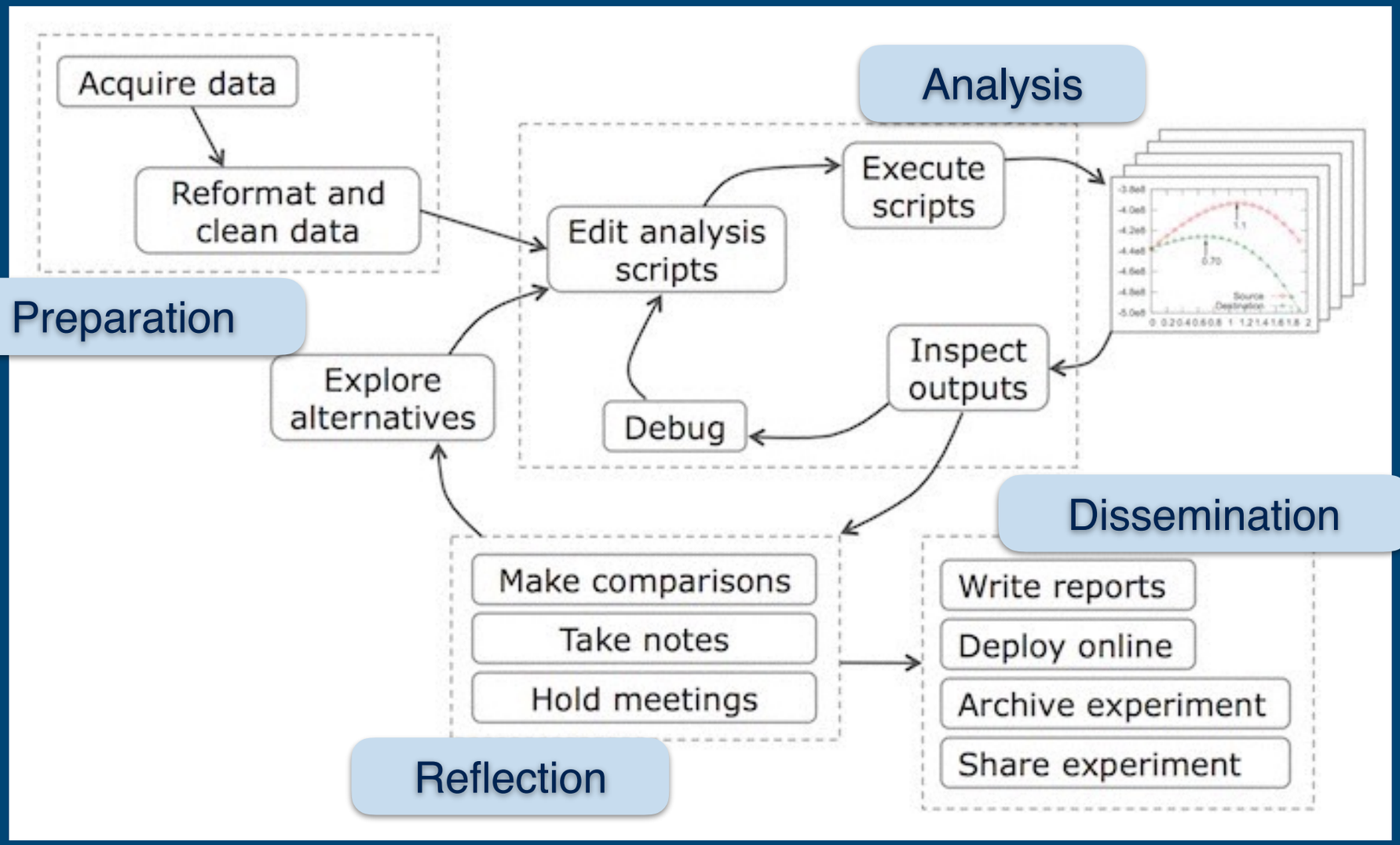
[…] form ever follows function, and this is the law. Where function does not change, form does not change.

Louis Sullivan, American architect

https://en.wikipedia.org/wiki/Form_follows_function

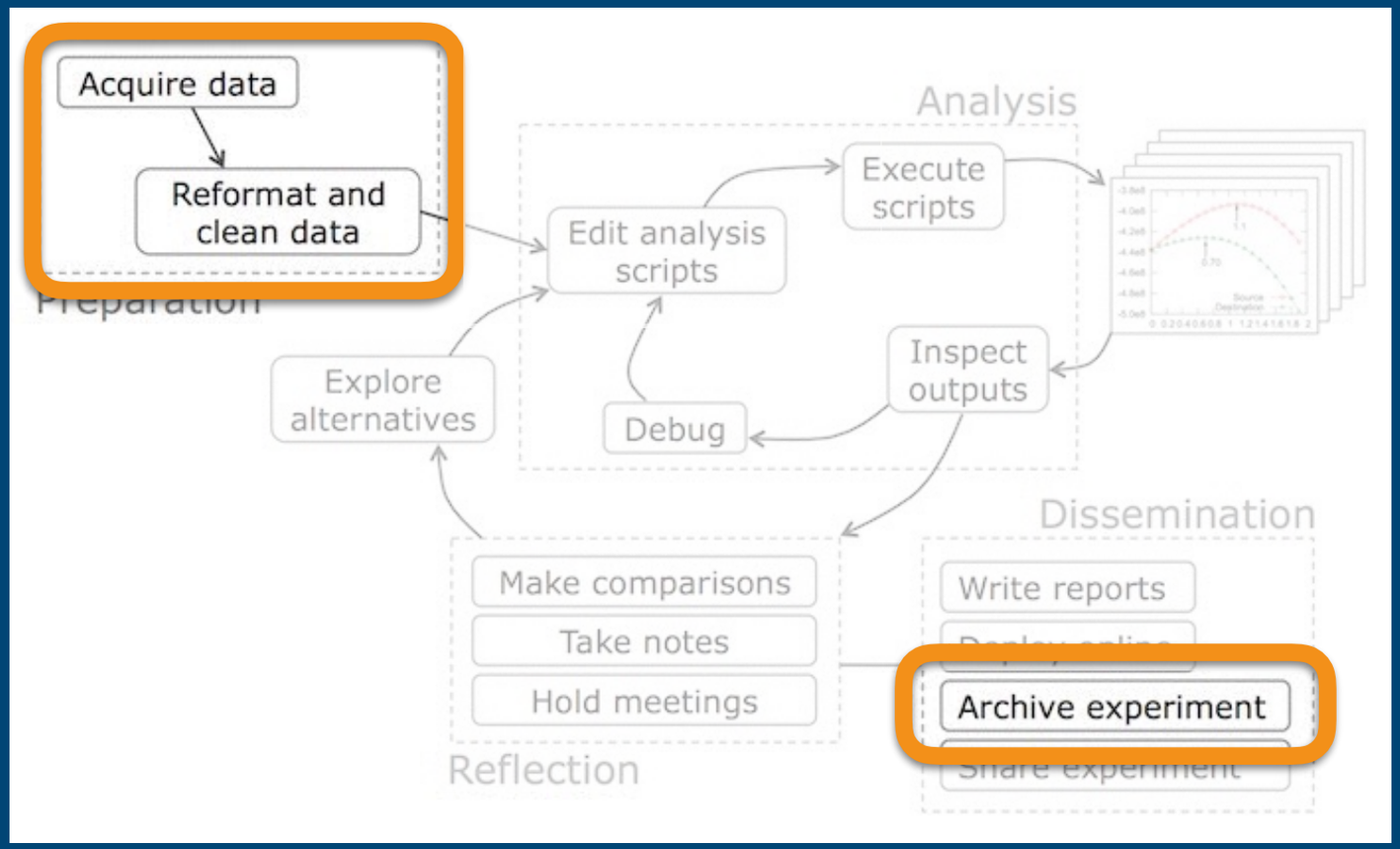
Photo: http://calumet412.com/page/159

Data Science workflow



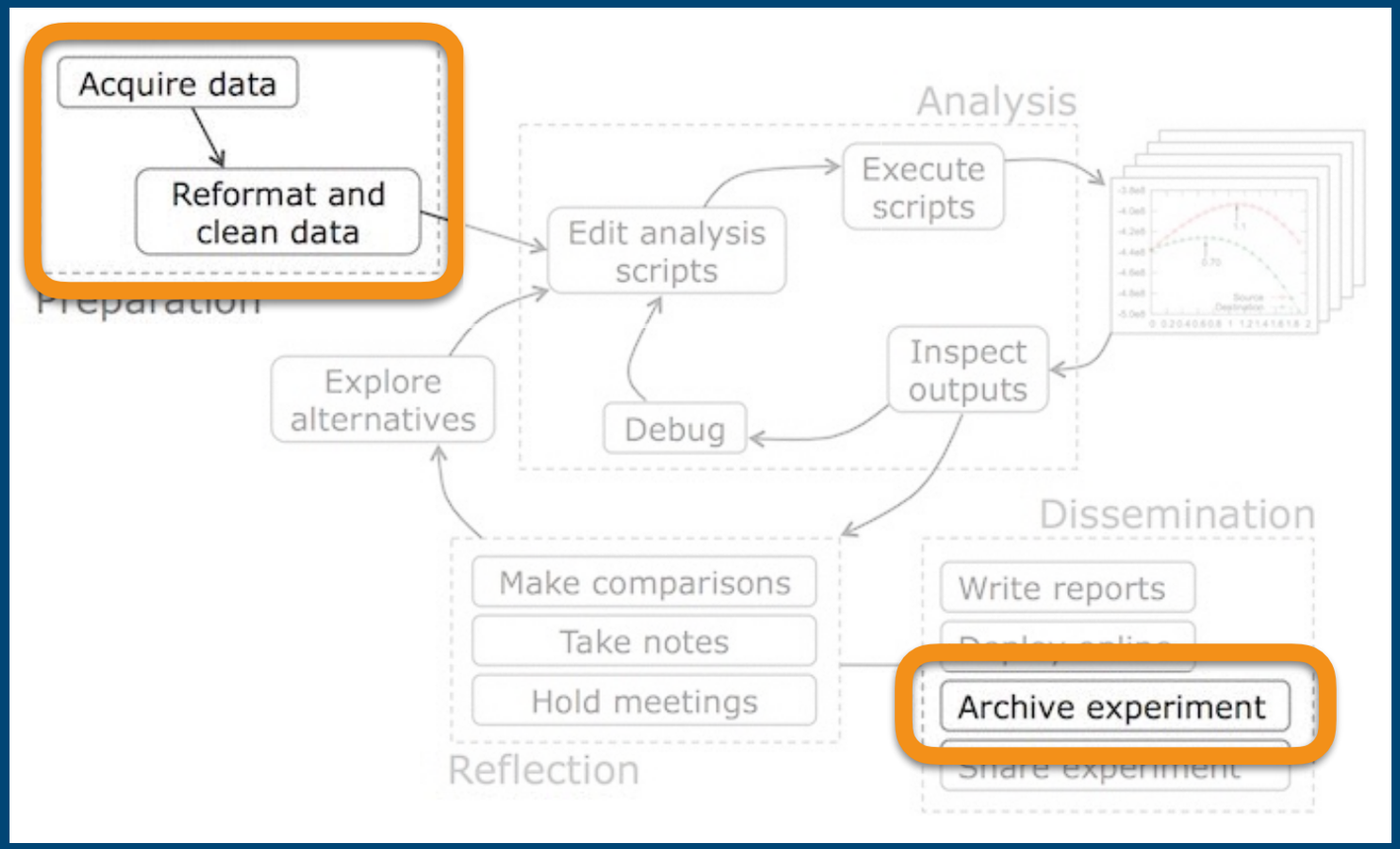
Guo, P. (2013, October 30). Data Science Workflow: Overview and Challenges. Retrieved from http://cacm.acm.org/blogs/blog-cacm/169199-data-science-workflow-overview-and-challenges/fulltext

Data lifecycle elements



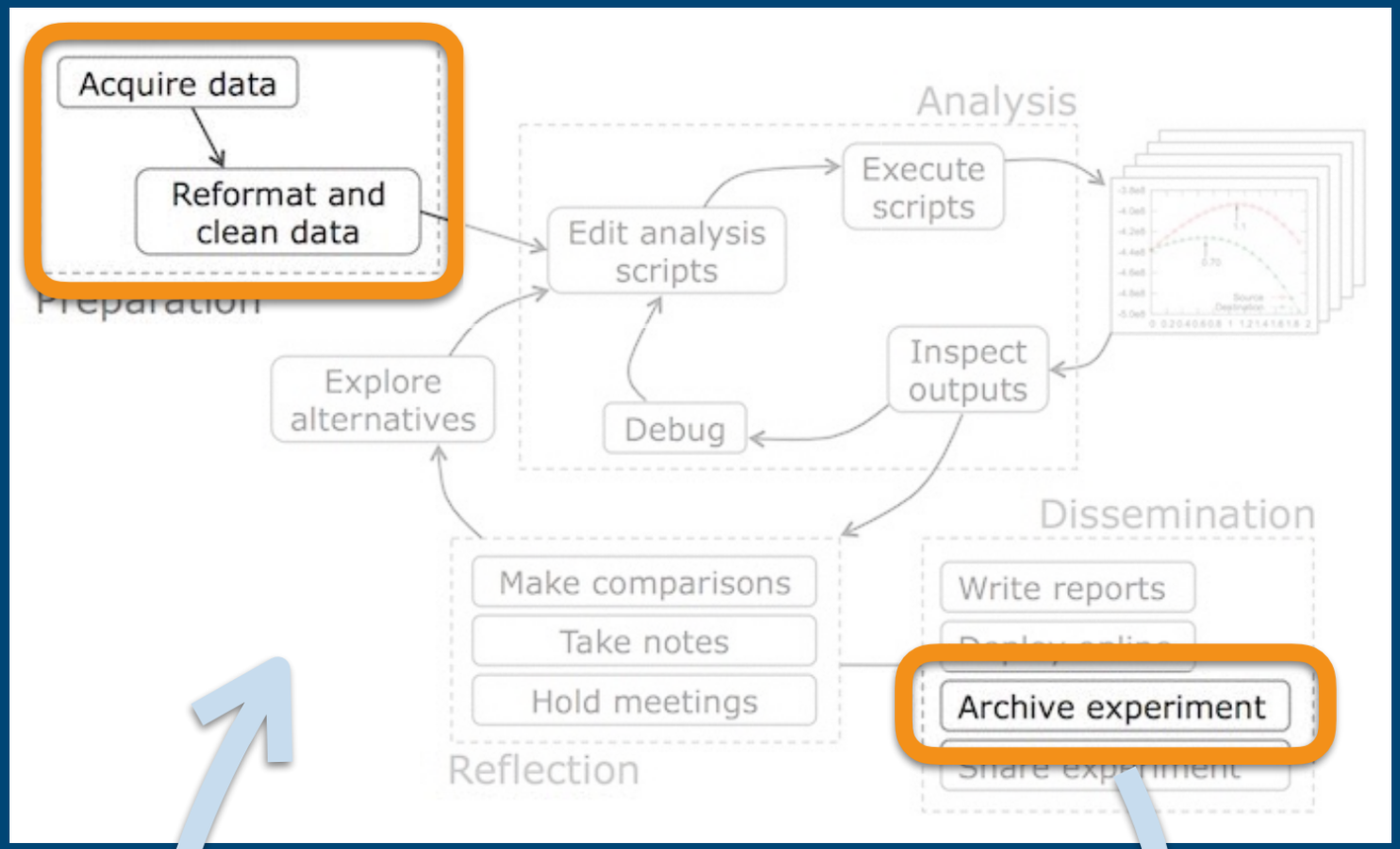
Data lifecycle elements

Act I



Data lifecycle elements

Act 1



Act 2

[Discover and] re-use data

Act 1: Field Work

Act 1: Data compilation

```
1995081102   1995081018   BDE14  BD  E   1995   222 8   18  5.78   85.40
1995081102   1995081018   BDE15  BD  E   1995   222 8   18  6.05   81.69
1995081102   1995081018   BDE16  BD  E   1995   222 8   18  6.13   -999.10
1995081102   1995081018   BDE17  BD  E   1995   222 8   18  6.05   -999.10
1995081102   1995081018   BDE18  BD  E   199                       2.69
1995081102   1995081018   BDE20  BD  E   1995   222 8   18  .96    -999.10
1995081102   1995081018   BDE21  BD  E   1995   222 8   18  5.96   81.09
```

One year per file

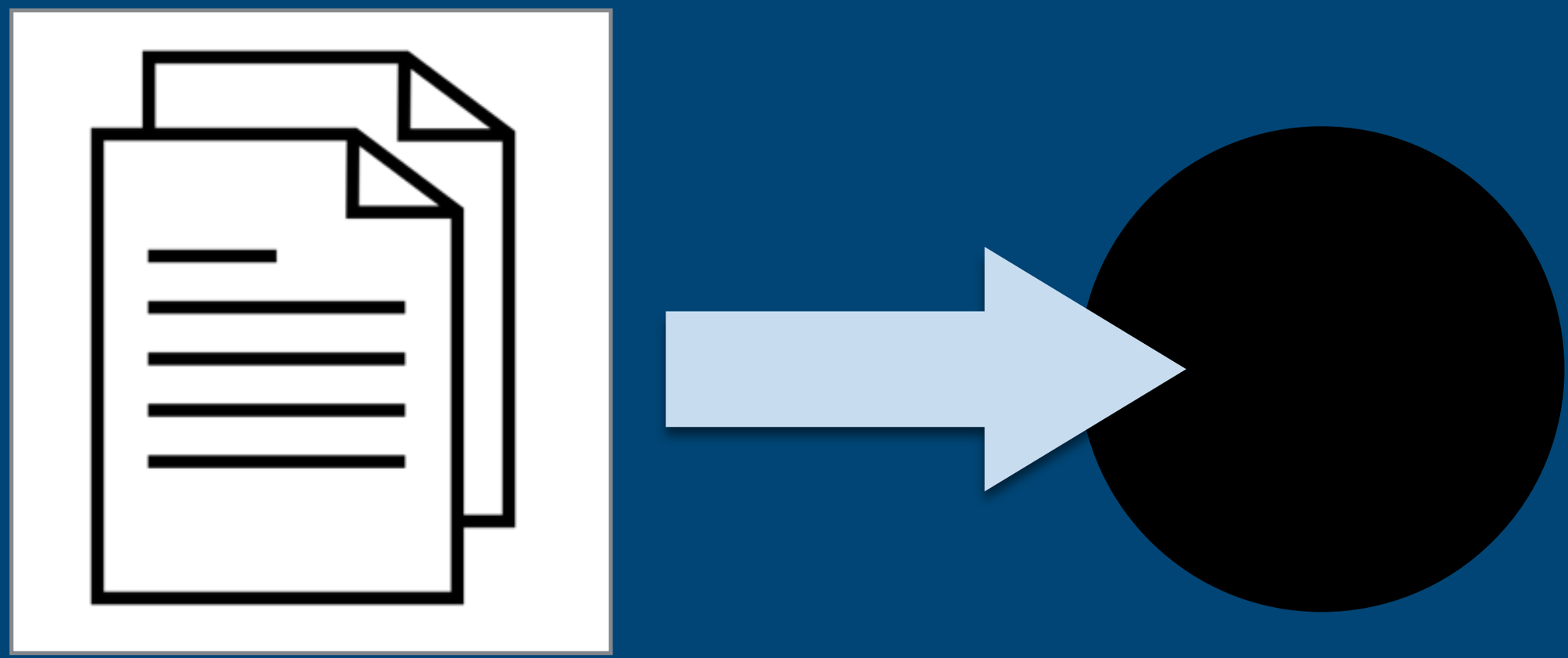
Temperature

Site, community, treatment

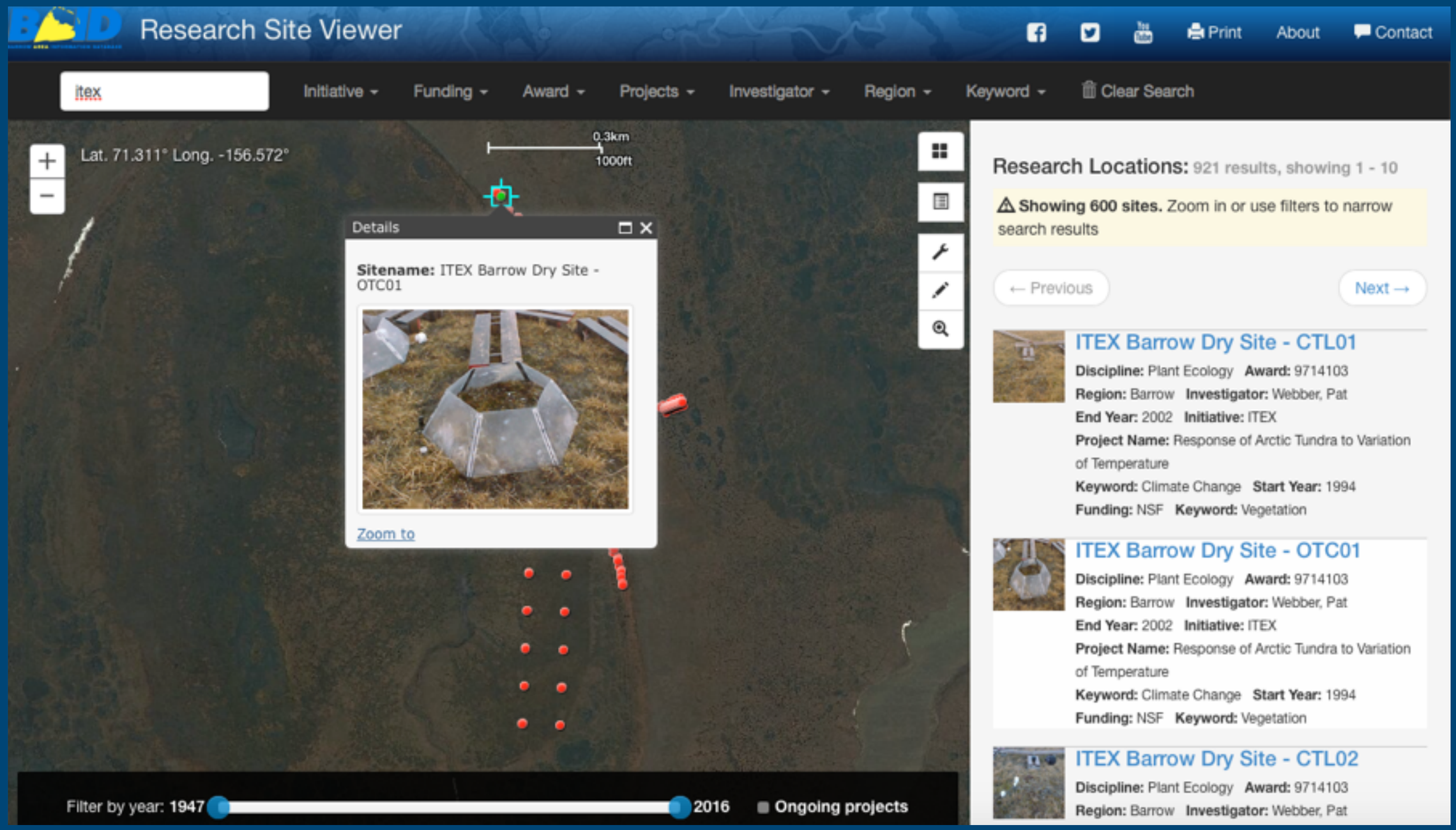
Relative humidity

Act 1: Archive

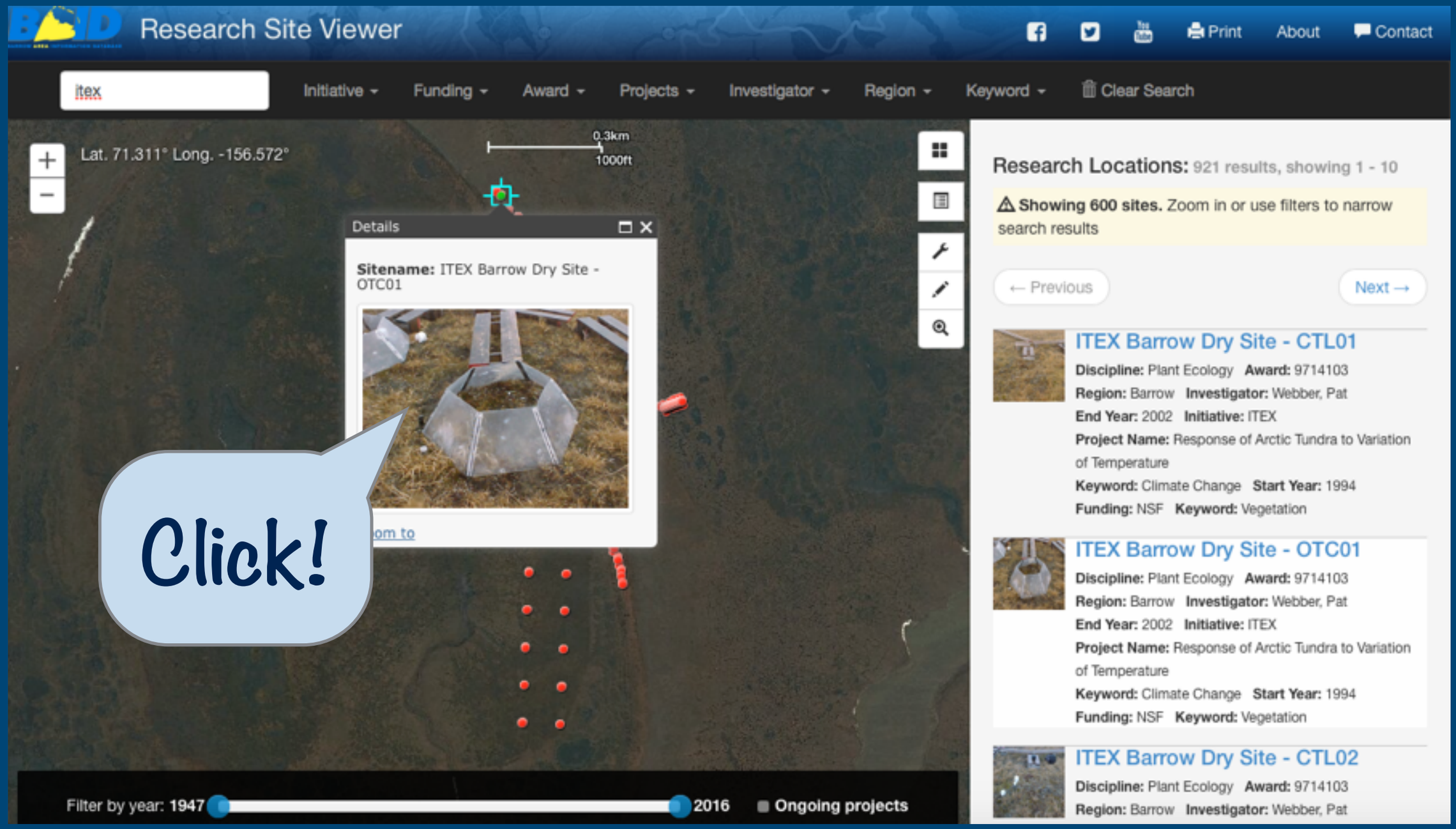
PI transfers data to archive as per funding agency requirements



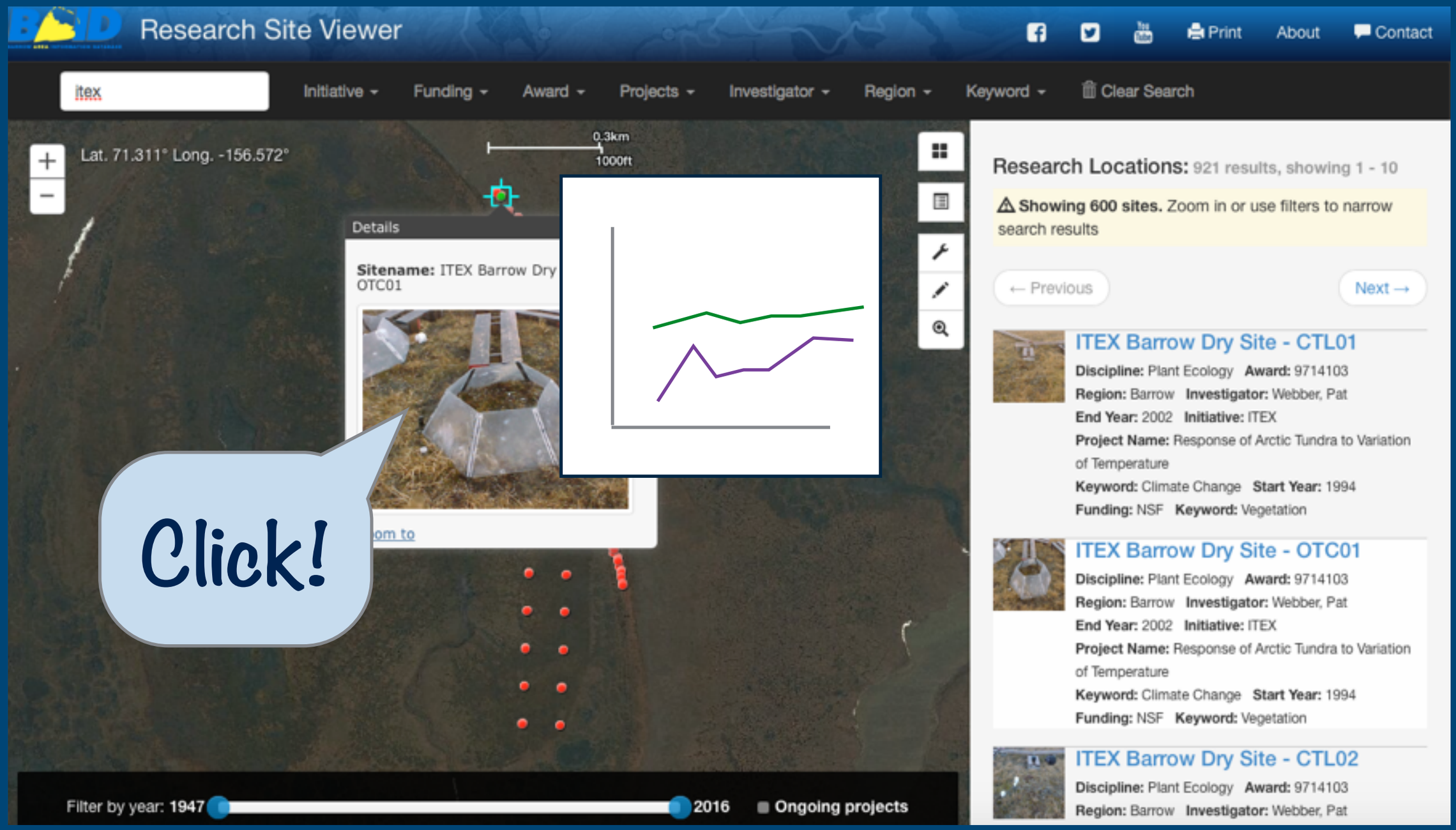
Act 2: New application



Act 2: New application



Act 2: New application



That's a nice story, but…

How is this relevant to me?

- When you find yourself in the data set creation pipeline, do the right thing.
- Educate domain scientists regarding data structure pros and cons
- Keep in mind the need for care and feeding of previously-created (and archived) data. Multi-season data sets may have components that change over the data collection timespan.

References and Reading

- Berman, F. and V. Cerf. (2013, August 09). Who Will Pay for Public Access to Research Data? *Science* Vol. 341, Issue 6146, pp. 616-617. DOI: 10.1126/science.1241625

- Guo, P.J. (2012). Software Tools to Facilitate Research Programming (Doctoral dissertation). Retrieved from http://pgbovine.net/publications/Philip-Guo_PhD-dissertation_software-tools-for-research-programming.pdf

- Guo, P. (2013, October 30). Data Science Workflow: Overview and Challenges. Retrieved from http://cacm.acm.org/blogs/blog-cacm/169199-data-science-workflow-overview-and-challenges/fulltext

- Hey, T., S. Tansley, and K. Tolle, ed. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/