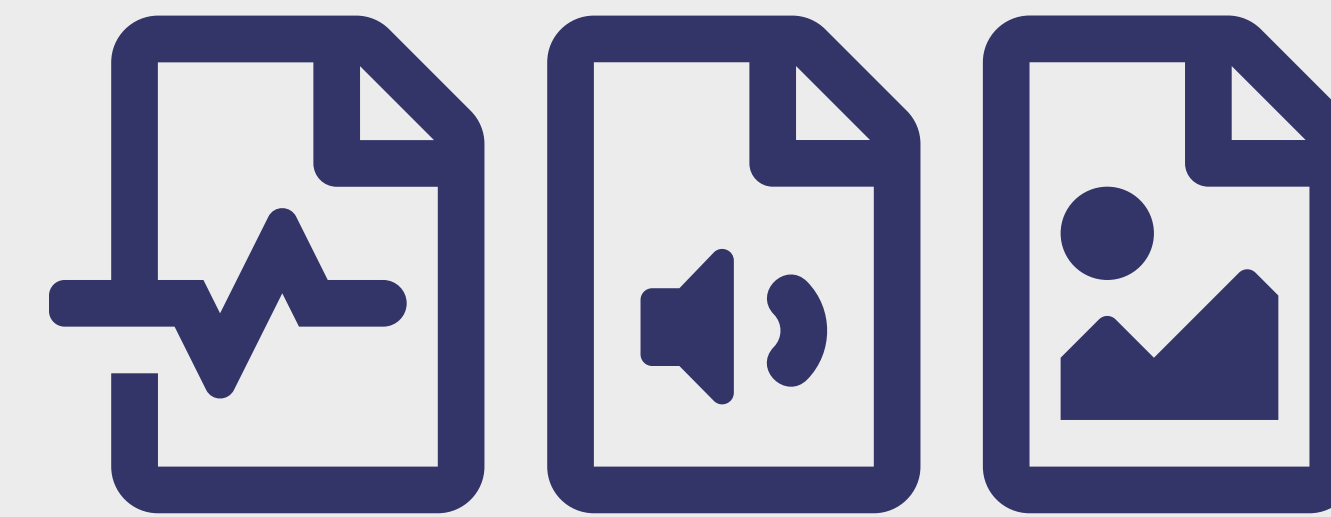


The Arctic Data Center is geographically focused rather than domain specific, which fosters cross-disciplinary data discovery and synthesis. However, developing guidelines and support for heterogeneous data and metadata across social, physical, and biological sciences can be challenging.

Open source tools and workflows help us collaborate as a team to address some of these issues.

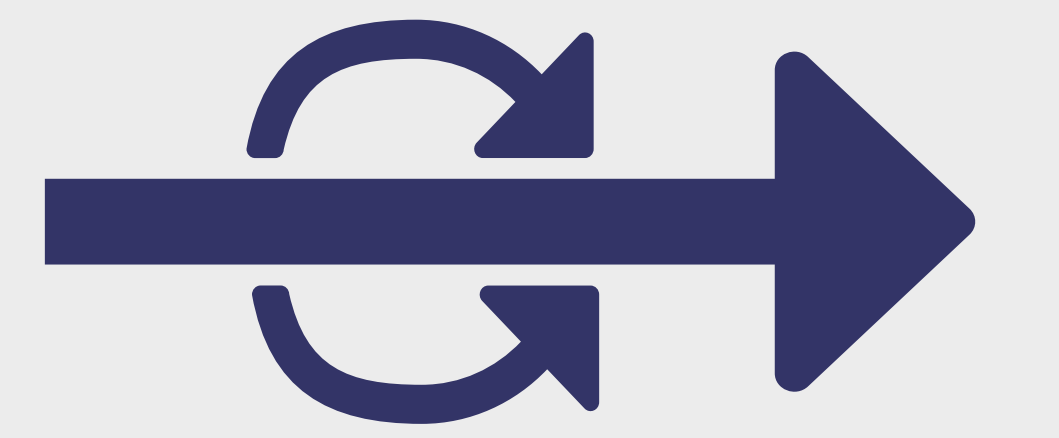
## Challenges



Heterogeneous data



Continually increasing data quality standards



Concurrent development of software and R tools

## Data processing

### Edit metadata



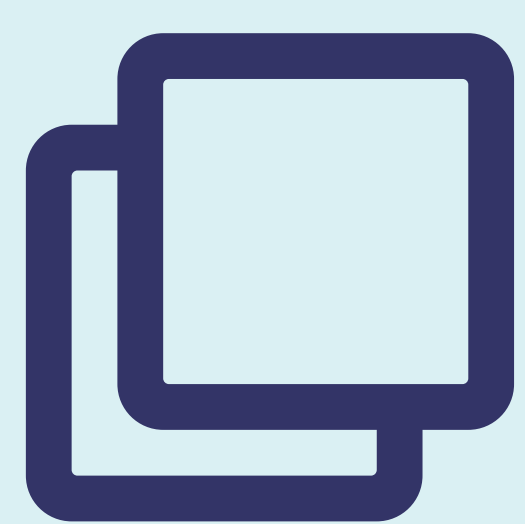
The Arctic Data Center stores scientific metadata using Ecological Metadata Language (EML). We make scripted edits to metadata to provide a reproducible record of the changes, as well as to add sections not yet supported by our web-based editor.

### Interact with repositories



We upload and download metadata, data objects, and data packages (groups of data objects and their associated metadata) through R. We also use R to copy data packages between repositories in the DataONE network.

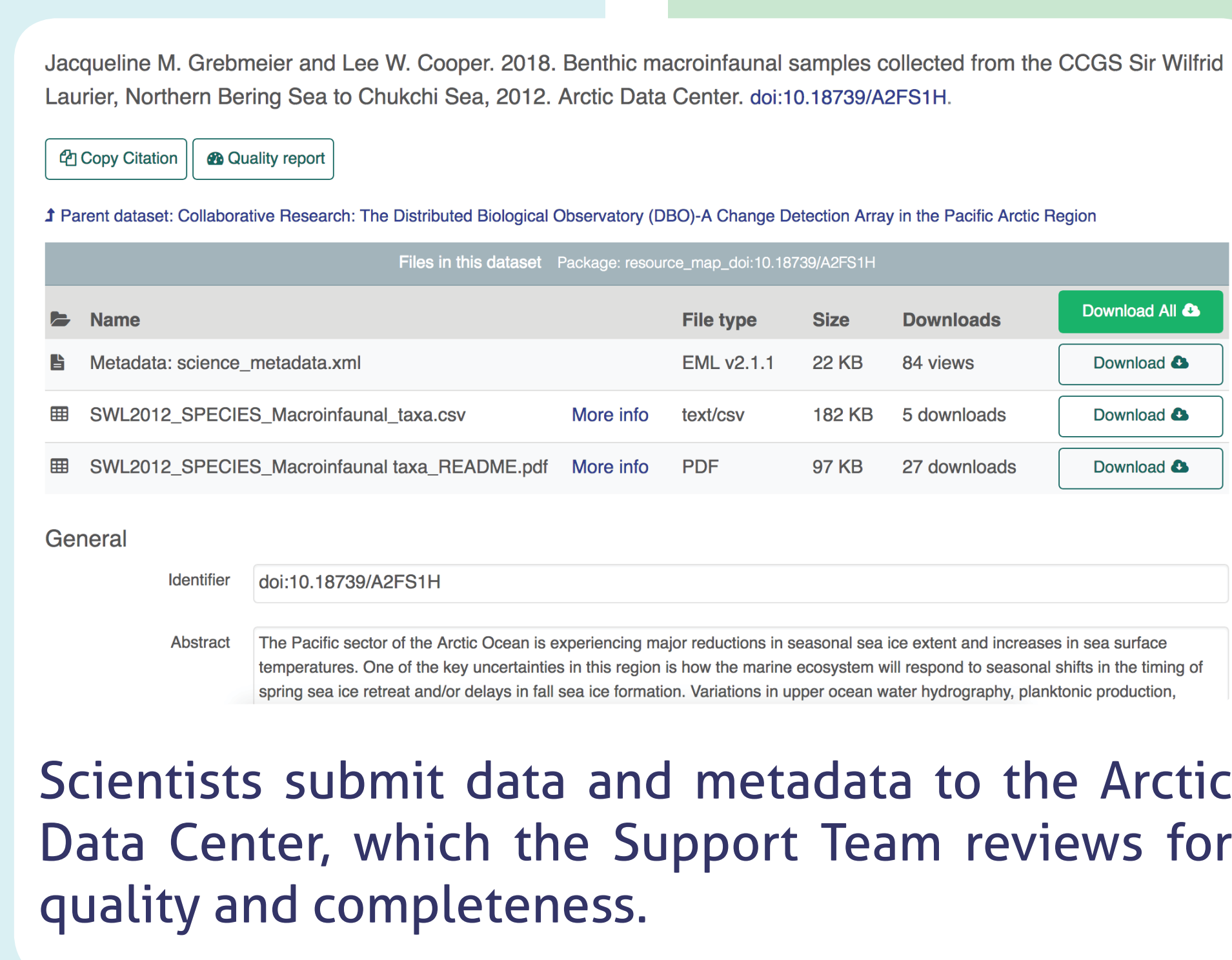
### Batch-process data packages



Scripted changes to EML allow us to programmatically process groups of related data packages. For example, measurements taken at different sites or across several years may have metadata records that vary only in geographic or temporal coverage.



Support Team  
1 project coordinator, 3 fellows, 4 interns



Jacqueline M. Grebmeier and Lee W. Cooper. 2018. Benthic macrofaunal samples collected from the CCGS Sir Wilfrid Laurier, Northern Bering Sea to Chukchi Sea, 2012. Arctic Data Center. doi:10.18739/A2FS1H.

Name	File type	Size	Downloads	Download All
Metadata: science_metadata.xml	EML v2.1.1	22 KB	84 views	Download
SWL2012_SPECIES_Macrofaunal_taxa.csv	text/csv	182 KB	5 downloads	Download
SWL2012_SPECIES_Macrofaunal_taxa_README.pdf	PDF	97 KB	27 downloads	Download

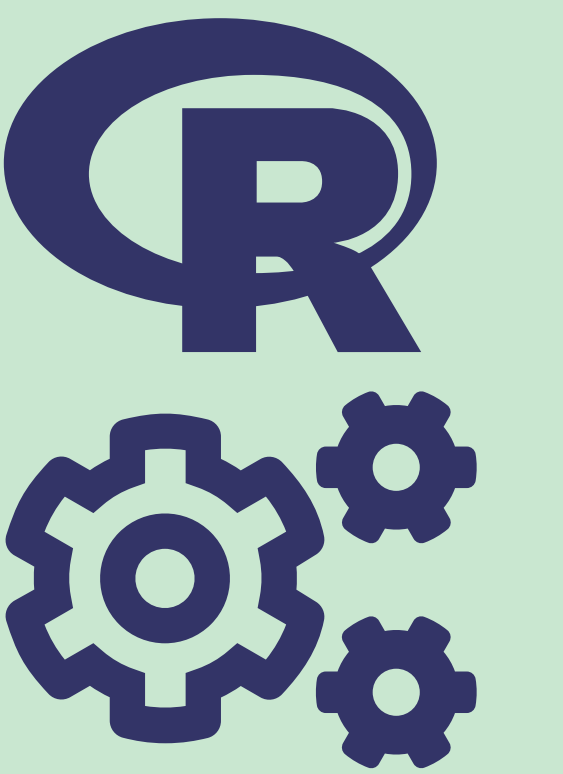
General  
 Identifier: doi:10.18739/A2FS1H  
 Abstract: The Pacific sector of the Arctic Ocean is experiencing major reductions in seasonal sea ice extent and increases in sea surface temperatures. One of the key uncertainties in this region is how the marine ecosystem will respond to seasonal shifts in the timing of spring sea ice retreat and/or delays in fall sea ice formation. Variations in upper ocean water hydrography, planktonic production,

Scientists submit data and metadata to the Arctic Data Center, which the Support Team reviews for quality and completeness.

## Resource development

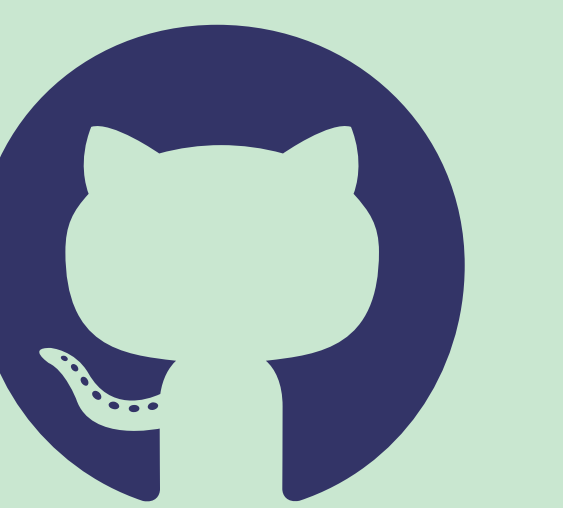
### Build tools

We have written R functions to (a) process data packages more efficiently, including a Shiny app for attribute metadata; (b) perform quality assurance on data packages; and (c) improve data use and synthesis.



### Share code

GitHub streamlines the process of sharing scripts and R functions within the Support Team, as well as with synthesis scientists using the repository. Versioned and transparent code enhances remote collaboration.



### Create training materials

For our training materials, we compile RMarkdown files into an online book format (bookdown), integrating code chunks with written explanations to develop compelling and reproducible examples.

