



Australian Government

Department of Climate Change, Energy, the Environment and Water
Australian Antarctic Division



AUSTRALIAN
ANTARCTIC
PROGRAM

Integrated Digital East Antarctica

Foundational software development and data practices

Michael Sumner, Research Software Engineer



Integrated Digital East Antarctica Program

Aleks Terauds
(Program Leader)

Data and Models

Lenneke Jong
(Model Development and Integration)

Mike Sumner
(Research Software Engineer)

Tania Raymond (0.4 FTE)
(Science Support - publications)

Synthesis and Visualisation

Ben Raymond
(Computational Ecologist)

Martin Cox (0.2 FTE)
(Krill Acoustics)

Position Vacant
(Research Data Synthesis)

Why Integrated Digital East Antarctica ?



Strategic



Science



Collaboration



Knowledge



IDEA will deliver....

A digital representation of Antarctica and the Southern Ocean that integrates data from multiple disciplines and provides unprecedented insights into connections and change trajectories across environments and biomes



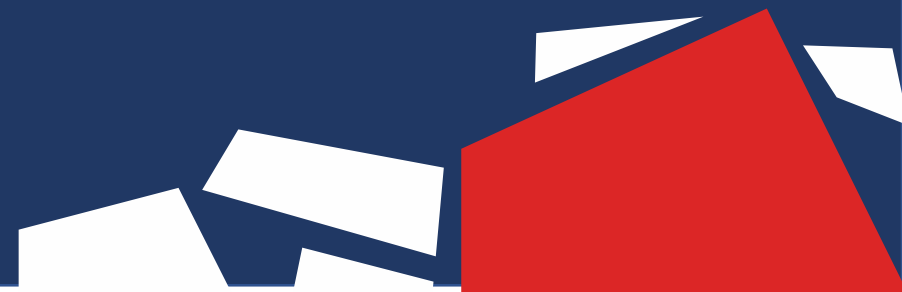
Australian Government

Department of Climate Change, Energy, the Environment and Water

Australian Antarctic Division



AUSTRALIAN
ANTARCTIC
PROGRAM





Australian Government

Department of Climate Change, Energy, the Environment and Water
Australian Antarctic Division



AUSTRALIAN
ANTARCTIC
PROGRAM

IDEA will deliver....

- A broad selection of relevant and actionable data
- Interfaces and methods of access to data to support different tasks and user communities
- Analytical tools and compute capacity.
- Data products, reports and policy ready summaries

IMPACT (End Use Cases)



2023-24

Applications and work packages

...and beyond

Data Processing and Visualisation Tools

Data Management Systems

Data Acquisition and Storage



Interdependencies & interconnectedness

Interdependencies

- Compute capacity, compute environment, software, upstream capability and agility
- Alignment of community needs and technology
- Australian Antarctic Data Centre
- People – data doesn't integrate itself; compute infrastructure requires people

Interconnectedness

- IDEA was developed to connect, collaborate and partner across the Australian Antarctic Science Program
- The scale and breadth of these connections and partnerships depends on future funding



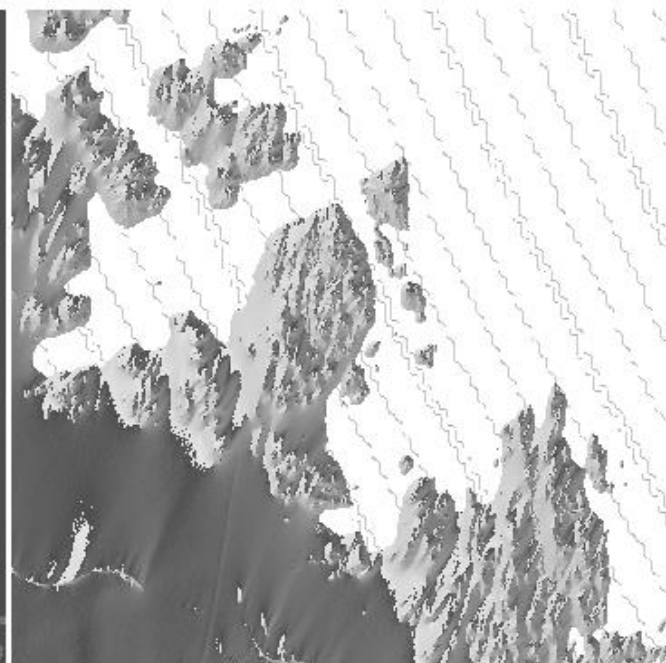
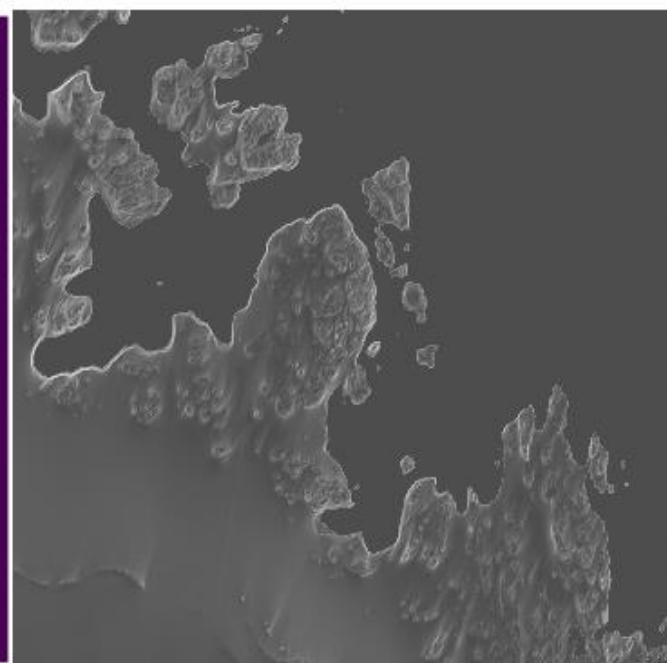
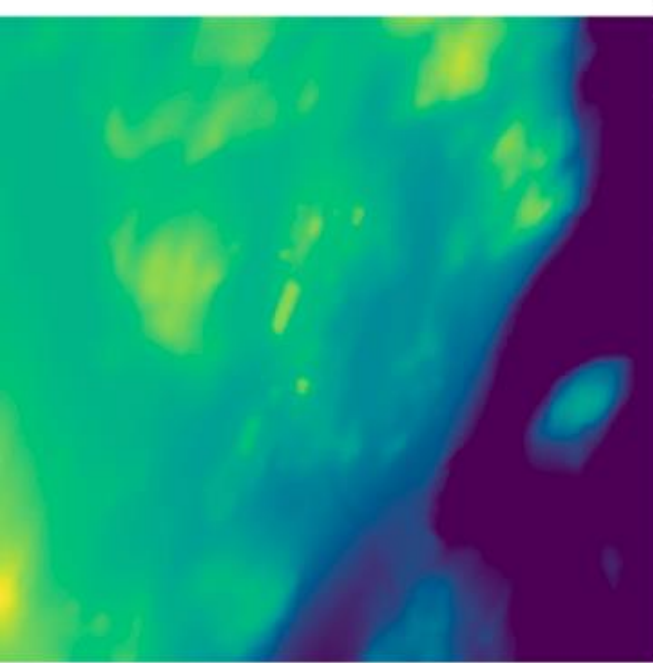
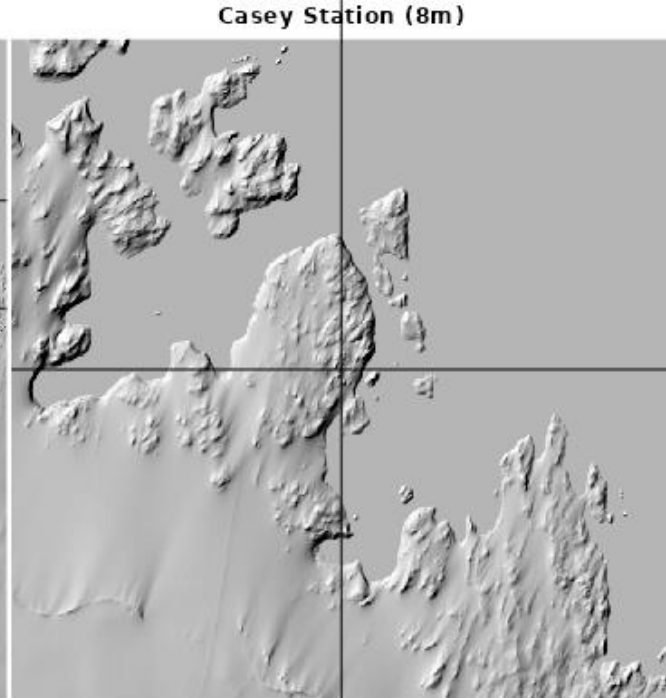
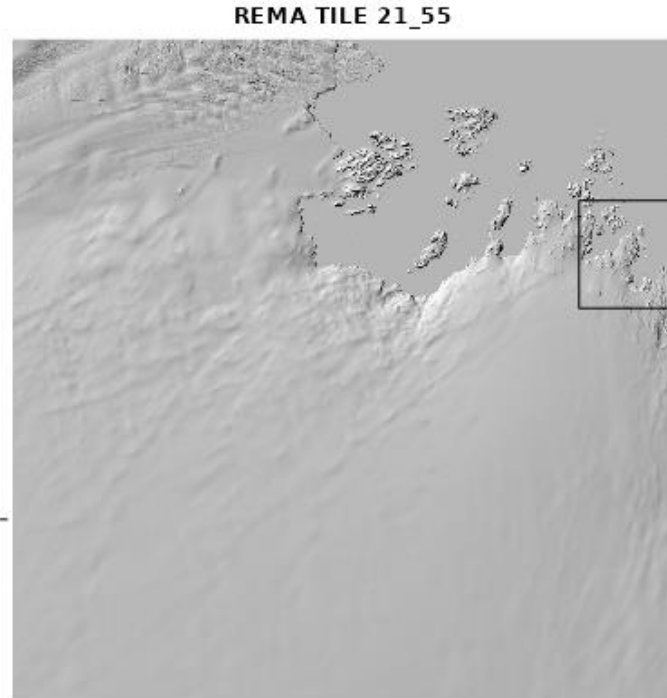
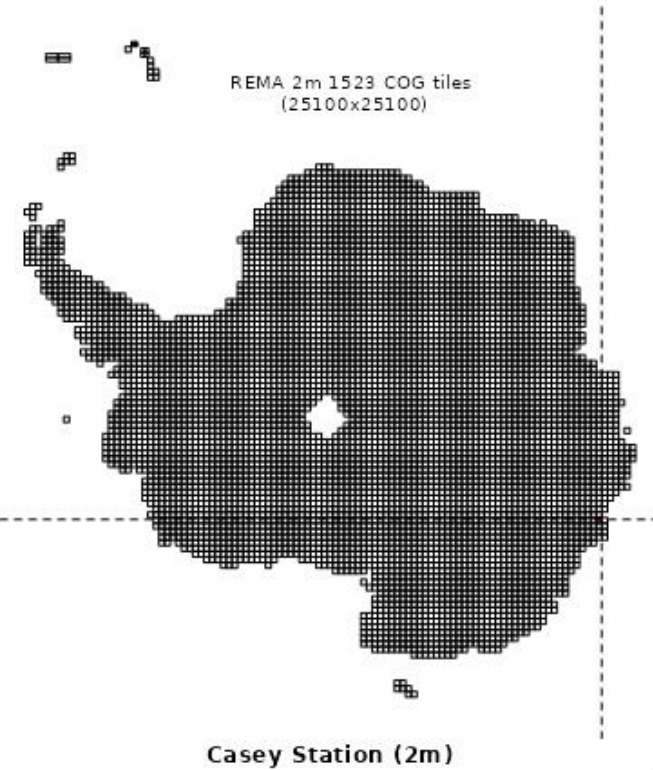
Key datasets and issues

Large files, inconsistent metadata standards

- REMA yuge dataset, mix of indexing practices
- GHRSSST legacy format, incomplete metadata, entropy issues
- AMSR sea ice, legacy format, no metadata, mix geospatial/model standards
- Various model output, special grids, incomplete metadata
- Many, many others with variations on the above ...

Foundational software development

- Remove R-only code
- Contribute changes upstream (if GDAL reads it, every package can)
- Prefer on-the-fly cross-language virtual-file workflows to all-encompassing datacubes



REMA 2m,
8m, 30m
continental
scale DEM



Remove R-only code (**with light, cross-lang newness!**)

Datacube raadtools function ``readghrsst()``

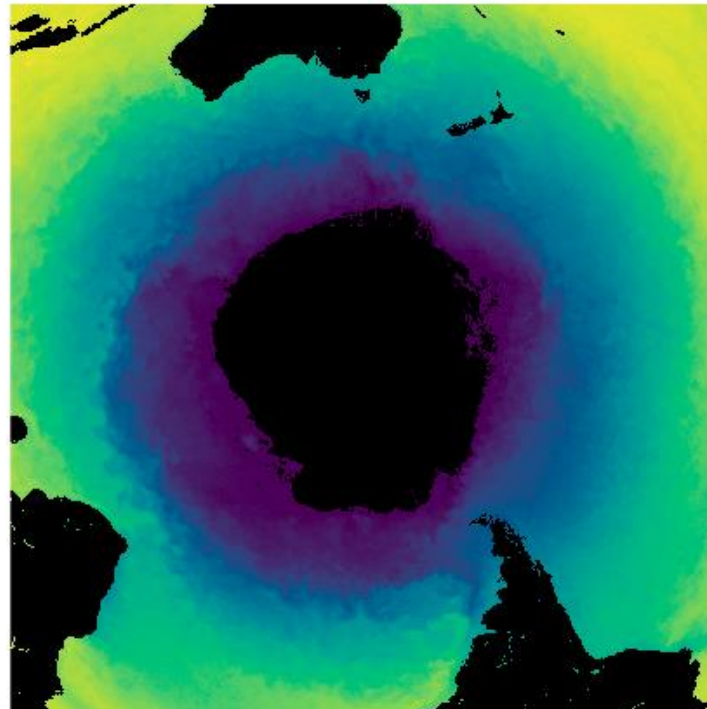
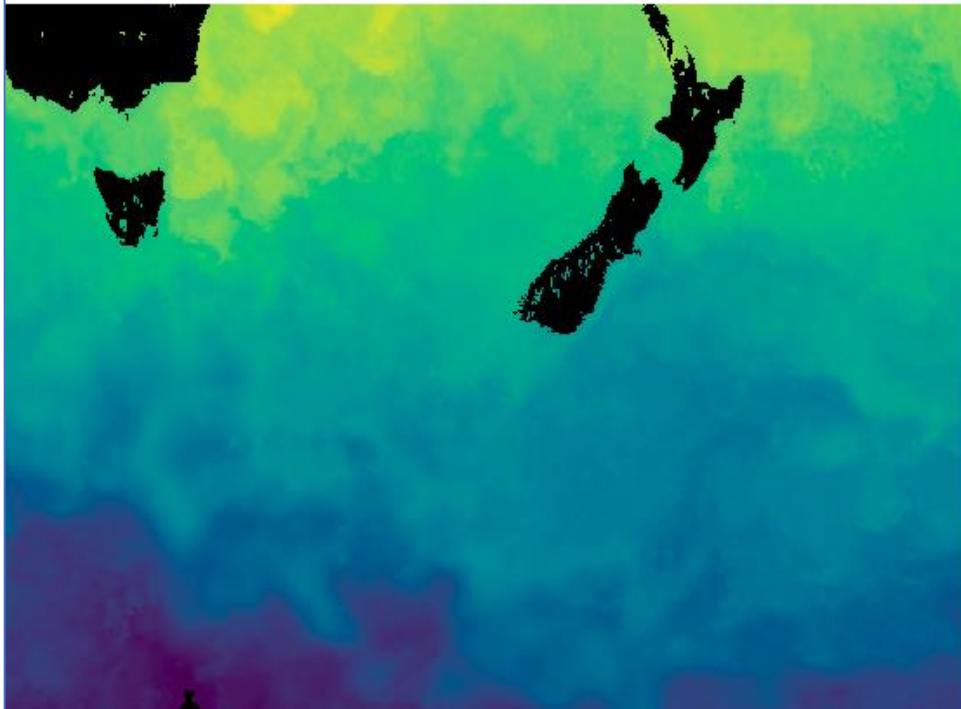
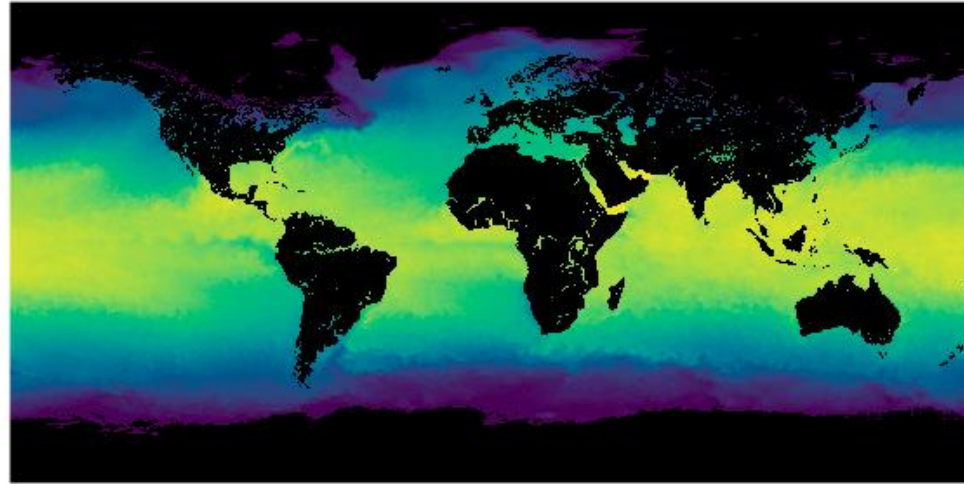
- Finds file by date-time query (**use STAC**)
- Identifies variable of interest in file (**use virtual-file descript**)
- Ensures correct extent and crs (**re-format and re-publish, OR virtual re-descriptors**)
- Memory manage query window, or write file (**use on-the-fly re-gridding to *any* grid extent/dimension/crs**)



These changes sound subtle but together represent a new paradigm of lazy-data-access, virtualization, flexibility, and efficiency in downstream user languages: R, Python, Julia, Javascript, ... and, change is still coming and settling in many communities. STAC had **enormous** influence, and more is coming.



High resolution, daily data – huge files and lots of them



VRT 'a_coord_epoch' for vrt:// connection ✖

#8375 by mdsumner was merged last month 2 of 5 tasks

VRT 'scale' 'unscale' for vrt:// connection (2nd attempt) ✔

#8366 by mdsumner was merged on Sep 8 5 tasks done

vrt://NETCDF:20230423090000-JPL-L4_GHRSSST-SSTfnd-MUR-GLOB-v02.0-fv04.1.nc:analysed_sst?a_srs=OGC:CRS84&a_ullr=-180,90,180,90

This syntax takes a file description and virtualizes it in ways that we fix at the command line by writing a new file, specify the variable, set the extent and metadata correctly ... virtualization **without software** and can be done for many of the available options of `gdal_translate` and others.

VRT 'projwin' for vrt:// connection ✖

#8320 by mdsumner was merged on Sep 4 4 of 5 tasks

dev_environment.rst: fix docker label ✔

#8284 by mdsumner was merged on Aug 29

STACIT: apply vsis3 protocol to s3:// items ✔ backport release/3.7

#8168 by mdsumner was merged on Aug 6 4 tasks done

gpkg.rst stray backticks ✔

#7838 by mdsumner was merged on May 26 2 tasks

Update vrt.rst - fix tiny typo ✔

#7837 by mdsumner was merged on May 26

```
31 #' @export
32 readghrsst <- function (date, time.resolution = c("daily"),
33                       xylim = NULL, lon180 = TRUE,
34                       varname = c("analysed_sst", "analysis_error", "mask", "sea_ice_fraction"),
35                       setNA = TRUE,
36                       latest = TRUE,
37                       returnfiles = FALSE, ..., inputfiles = NULL) {
38   if (!lon180) stop("only lon180 is supported")
39   time.resolution <- match.arg(time.resolution)
40   varname <- match.arg(varname)
```

```
55   format = " extracting [:bar] :percent in :elapsed",
56   total = nfiles, clear = FALSE, width= 60)
57   pb$tick(0)
58   read_fun <- function(xfile, ext, varname = "", band = 1) {
59     pb$tick()
60     crop_if_needed(raster::raster(terra::rast(xfile, varname)), ext)
61   }
62   r0 <- stack(lapply(seq_len(nfiles), function(xi) read_fun(files$fullname[xi], ext = xylim, varname = varname)))
63   if (is.na(projection(r0))) projection(r0) <- "+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs +towgs84=0,0,0"
64   r0 <- setZ(r0, files$date)
65
66   ## need to determine if "filename" was passed in
67   dots <- list(...)
68   if ("filename" %in% names(dots)) {
69     r0 <- writeRaster(r0, ...)
70   }
71
72   if (nfiles == 1) {
73     r0 <- r0[[1L]]
74     r0 <- setZ(r0, files$date)
75   }
```

Collaboration = success



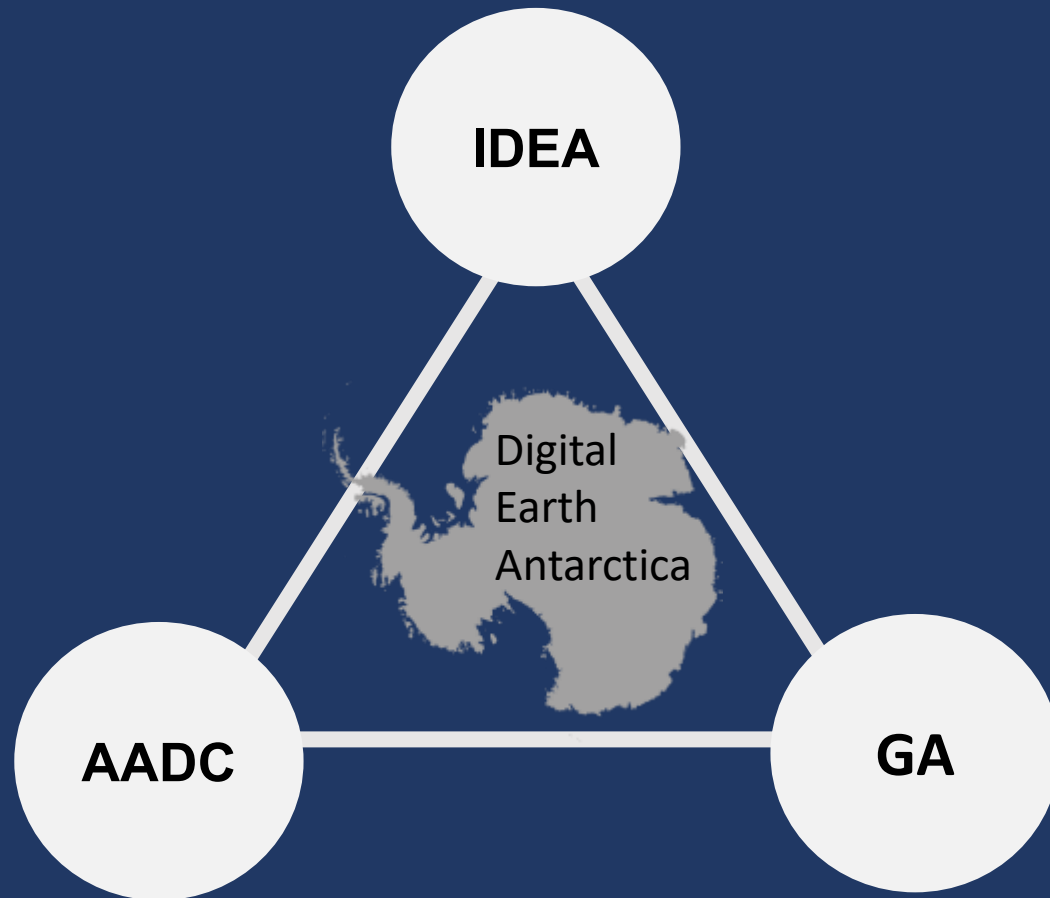
Australian Government

Department of Climate Change, Energy, the Environment and Water

Australian Antarctic Division

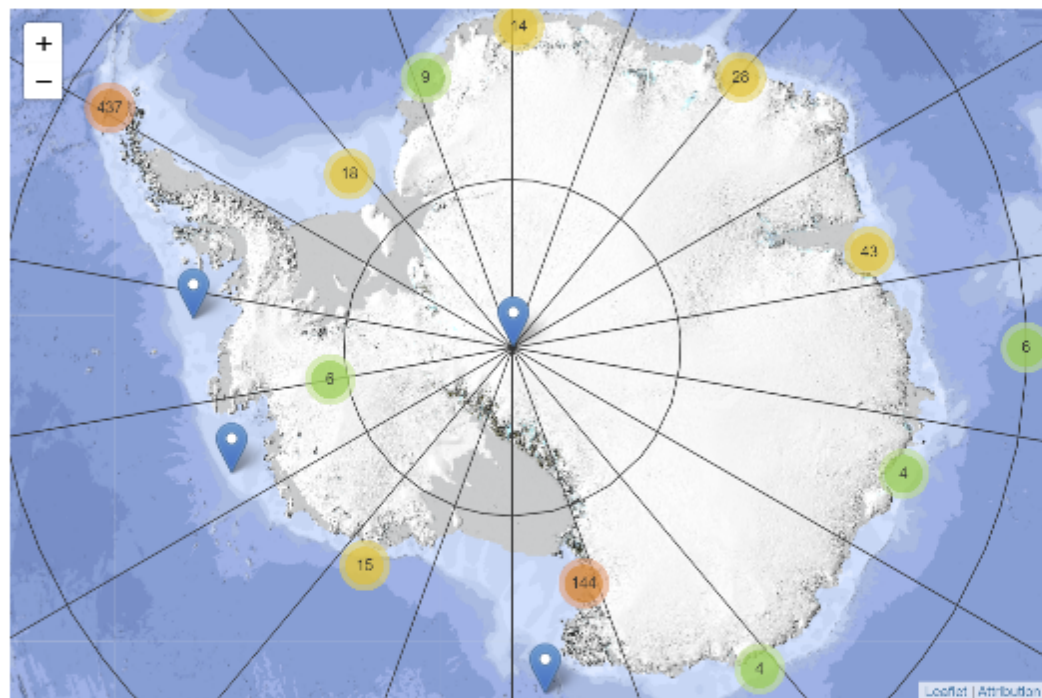


AUSTRALIAN
ANTARCTIC
PROGRAM



- Government Agencies
- Science Branch Programs
- Bureau of Meteorology
- Australian Antarctic Program Partnership
- CSIRO
- ARC Special Research Initiatives
- QUT Centre for Data Science
- AWS, Eratos, NVIDIA, Posit
- NCI Programs
 - TERN
 - ACCESS NRI
 - etc

Explorer [About this app](#)



Sample filtering

Total number of samples: 23539. After filtering: 16291

Sample type

- Biological
- Biological
- Non biological

Substrate group

- All groups

Taxonomic filters (applied to biological samples only) ⓘ

Kingdom

- All

Class

- All

Free text search

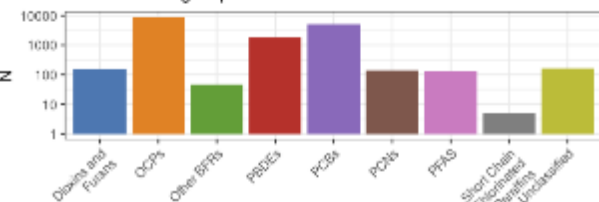
Contaminant name

Substrate

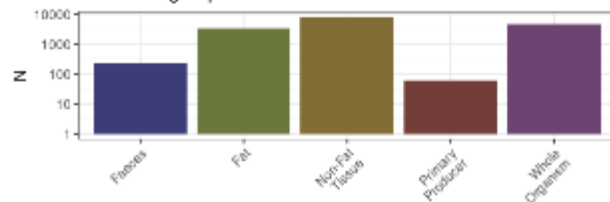
Phylum

Order

Contaminant group



Substrate group



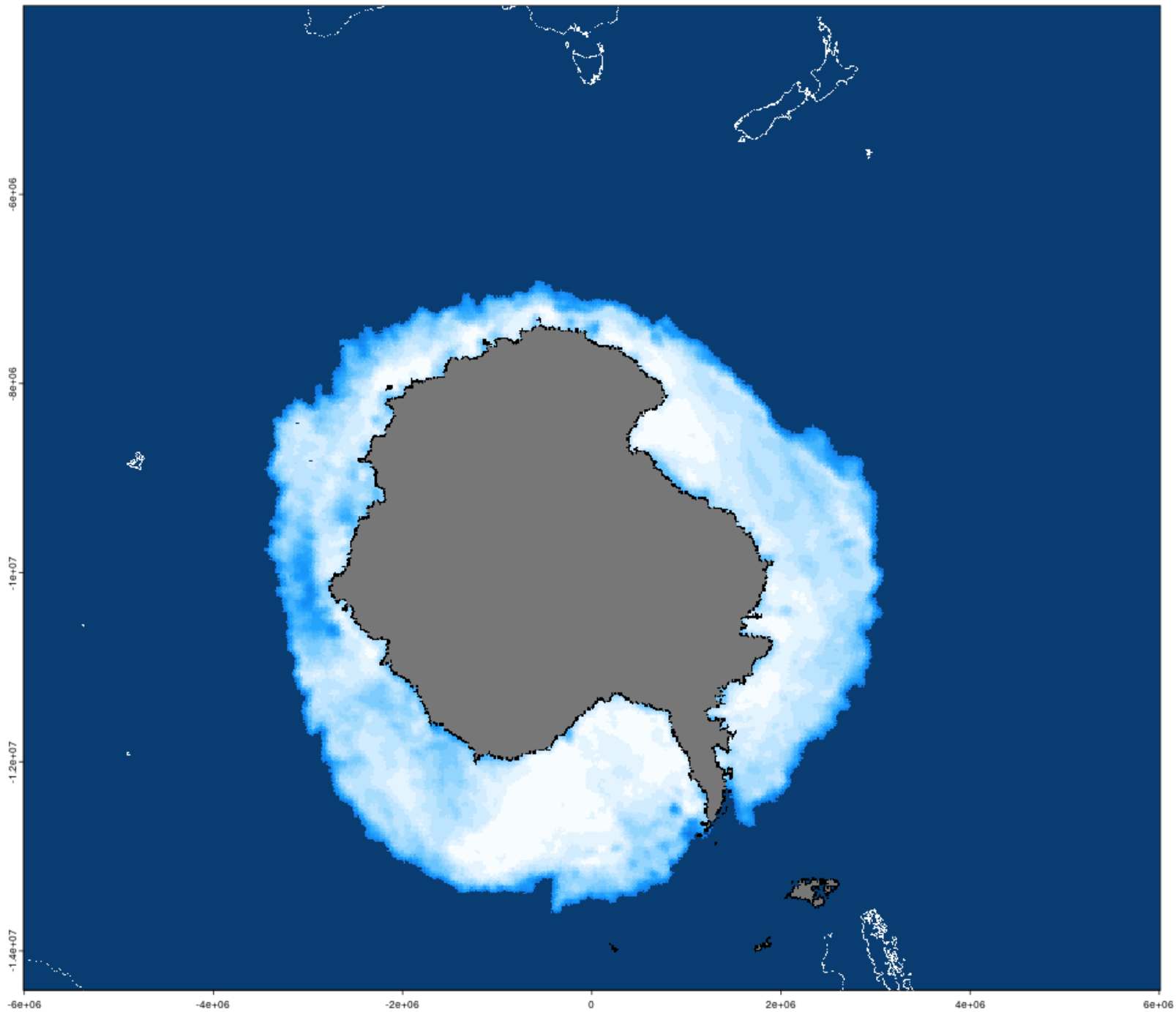
Measurements

Study

Details

DOI

Study	Details	DOI
All	All	All
4	Wild, S., Eulaers, I., Covaci, A., Bosal, R., Hawker, D., Cropp, R., Southwell, C., Emmerson, L., Lepoint, G., Eisenmann, P., Nash, S.B.(2021) South polar skua (<i>Catharacta macconnickii</i>) as biovectors for long-range transport of persistent organic pollutants to Antarctica. <i>Environmental Pollution</i> 292	https://doi.org/10.1016/j.envpol.2021.118358
5	Morales, P., Roscales, J.L., Muñoz-Arnanz, J., Barbosa, A., Jiménez, B. (2021) Evaluation of PCDD/Fs, PCBs and PBDEs in two penguin species from Antarctica. <i>Chemosphere</i> 286	https://doi.org/10.1016/j.chemosphere.2021.131871



Nuyina location

+

sea ice extent

Collaboration opportunities



Australian Government

Department of Climate Change, Energy, the Environment and Water

Australian Antarctic Division



AUSTRALIAN
ANTARCTIC
PROGRAM

- If you need Antarctic data
- Using or interest in downstream packages
- Do you want to contribute to software? R, Python, C++, Rust, Javascript, ... You can do it!!
- Foundational software contributions:
 - GDAL, Arrow, stream-reading in R and Python, re-gridding problems, reporting problem files