Module 5: Data Management
EVA meeting, Bergen 8 May 2014

Ingo Bethke
Data Storage

Resources in EVA

- EVA depends on the national data storage facilities (NorStore)
- NorStore allocations are per calendar year
- requested storage in Tb (without backup):

<table>
<thead>
<tr>
<th>Year</th>
<th>DISK</th>
<th>TAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>159</td>
<td>334</td>
</tr>
<tr>
<td>2015</td>
<td>180</td>
<td>+278</td>
</tr>
<tr>
<td>2016</td>
<td>334</td>
<td>+798</td>
</tr>
<tr>
<td>2017</td>
<td>769</td>
<td>+1614</td>
</tr>
</tbody>
</table>

- uncertain that NorStore can meet EVA’s storage demands
Resources in EVA

- EVA depends on the national data storage facilities (NorStore)
- NorStore allocations are per calendar year
- requested storage in Tb (without backup):
  
<table>
<thead>
<tr>
<th>Year</th>
<th>DISK</th>
<th>TAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>159</td>
<td>334</td>
</tr>
<tr>
<td>2015</td>
<td>180</td>
<td>+278</td>
</tr>
<tr>
<td>2016</td>
<td>334</td>
<td>+798</td>
</tr>
<tr>
<td>2017</td>
<td>769</td>
<td>+1614</td>
</tr>
</tbody>
</table>

- uncertain that NorStore can meet EVA’s storage demands

Current availability with "dusage -p ns2345k"

<table>
<thead>
<tr>
<th>Project</th>
<th>Account</th>
<th>Resource</th>
<th>Type</th>
<th>Usage</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS2345K</td>
<td>PROJECT</td>
<td>norstore_osl</td>
<td>Disk</td>
<td>147TB</td>
<td>200TB</td>
</tr>
<tr>
<td>NS2345K</td>
<td>PROJECT</td>
<td>tapestore</td>
<td>Tape</td>
<td>130TB</td>
<td>334TB</td>
</tr>
<tr>
<td>NS2345K</td>
<td>PROJECT</td>
<td>norstore_osl</td>
<td>Files</td>
<td>3264780</td>
<td>None</td>
</tr>
<tr>
<td>NS2345K</td>
<td>PROJECT</td>
<td>tapestore</td>
<td>Files</td>
<td>8170</td>
<td>33400</td>
</tr>
</tbody>
</table>

→ plan to move Earthclim legacy data to new National Data Archive
How do we internally manage the resources?

- allocation per project requires discipline from users
- keep your WP leader updated on your storage needs
- communicate storage issues as follows:
  researchers → WP leaders → project leader → NorStore
- reduce **DISK** storage where possible:
  1) consciously select output parameters, 2) use compression,
  3) migrate data to tape, 4) delete obsolete data
Data Storage

How do we internally manage the resources?

- allocation per project requires discipline from users
- keep your WP leader updated on your storage needs
- communicate storage issues as follows: researchers → WP leaders → project leader → NorStore
- reduce DISK storage where possible:
  1) consciously select output parameters, 2) use compression, 3) migrate data to tape, 4) delete obsolete data

Tape and compression guides

http://wiki.met.no/noresm/norstoretape
http://wiki.met.no/noresm/noresm2nc4mpi
Data Services

Data Services

Data Portal http://noresg.norstore.no

To download data: add datasets to your Data Cart, then click on Expand or wget.

Current Selections
- remove all
- project:CMIP5
- model:NorESM1-M

Search Categories
- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm

Results
- Data Cart
Data Services

Data Overview (Alf G.) http://wiki.met.no/noresm/listofruns

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Simulation type</th>
<th>Contact</th>
<th>Disk (space)</th>
<th>Ensemble</th>
<th>Name (click for additional info)</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO2 time evolution</td>
<td>C, 20thC</td>
<td>OS</td>
<td>norstore</td>
<td>Y</td>
<td>1850SOx</td>
<td>PEGASOS</td>
</tr>
<tr>
<td>CMIP</td>
<td>C, 20thC</td>
<td>OS</td>
<td>norstore</td>
<td>Y</td>
<td>N20TRAERCN_X (3 simulations X)</td>
<td>CMIP5</td>
</tr>
<tr>
<td>Aerosols &amp; radiative forcing</td>
<td>AMIP, PD&amp;P/1 yrs</td>
<td>AK</td>
<td>norstore (289 G) AeroCom disk</td>
<td>N</td>
<td>aerocomA2noresm_r128_X (7 simulations X)</td>
<td>EarthClim/AeroCom</td>
</tr>
<tr>
<td>Arctic aerosol optics</td>
<td>AMIP, (3 yrs)</td>
<td>PD</td>
<td>norstore (9 G)</td>
<td>N</td>
<td>NF2006to2008r164</td>
<td>EarthClim/ACCESS</td>
</tr>
</tbody>
</table>

Diagnostics (Detelina I.) http://noresg.norstore.no/public

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
<th>Description (click for details)</th>
<th>Diagnostics</th>
<th>Contact</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1850_f19 tn11_01_default</td>
<td>NorESM</td>
<td>AMOC sensitivity to isopycnal mixing</td>
<td>N1850_f19 tn11_01_default-Obs, 1-200y: Attn Ocn Ice Lnd</td>
<td>Detelina Ivanova</td>
<td></td>
</tr>
<tr>
<td>N1850_f19 tn11_01_E12</td>
<td>NorESM</td>
<td>AMOC sensitivity to isopycnal mixing</td>
<td>N1850_f19 tn11_01_E12-Obs, 1-200y: Attn Ocn Ice Lnd</td>
<td>Detelina Ivanova</td>
<td></td>
</tr>
<tr>
<td>MyExp</td>
<td>MyModel</td>
<td>MyExp Short Description (ReadME with details linked)</td>
<td>Type, Period: Attn Ocn Ice Lnd</td>
<td>MyEmail</td>
<td></td>
</tr>
</tbody>
</table>
Data Sharing

Norstore project area
- access limited to members of Earthclim/Eva
Data Sharing

Norstore project area
- access limited to members of Earthclim/Eva

ESG data portal – standard publication
- default option for model intercomparison projects (e.g., CMIP)
- strict requirements on post-processing and publishing format
  → data duplication, requires time and human resources
- contact: alfg@met.no (Oslo), ingo.bethke@uni.no (Bergen)
Data Sharing

Norstore project area
- access limited to members of Earthclim/Eva

ESG data portal – standard publication
- default option for model intercomparison projects (e.g., CMIP)
- strict requirements on post-processing and publishing format
  → data duplication, requires time and human resources
- contact: alfg@met.no (Oslo), ingo.bethke@uni.no (Bergen)

ESG data portal – simplified publication
- data published "as is" with help of single command (~5 min)
- no need for post-processing → no data duplication
- no access control, i.e., all data freely downloadable
- "unlisted" option = data downloadable but not exposed on portal
**Step 1:** choose data folder or file that you want to share and find an acronym for the dataset

Example for data folder: /projects/NS2345K/noresm/thredds/CORE2/ARCTIC

Example for dataset id: CORE2–ARCTIC
Step 1: choose data folder or file that you want to share and find an acronym for the dataset

Example for data folder: /projects/NS2345K/noresm/thredds/CORE2/ARCTIC
Example for dataset id: CORE2-ARCTIC

Step 2: run publisher script

Syntax: esgpublish <path to data> <dataset id> [unlisted]
Example: esgpublish /projects/NS2345K/.../CORE2/ARCTIC CORE2-ARCTIC
ESG data portal – simplified publication

Access option 1: Use wget-script from http://noresg.norstore.no/thredds/fileServer/esg_dataroot/ns2345k/CORE2-AMOC/wget_CORE2-AMOC.sh
ESG data portal – simplified publication

**Access option 1:** Use wget-script from [http://noresg.norstore.no/thredds/fileServer/esg_dataroot/ns2345k/CORE2-AMOC/wget_CORE2-AMOC.sh](http://noresg.norstore.no/thredds/fileServer/esg_dataroot/ns2345k/CORE2-AMOC/wget_CORE2-AMOC.sh)

**Access option 2:** Visit portal page [http://noresg.norstore.no](http://noresg.norstore.no) and search for dataset with project-id=noresg-ns2345k and dataset-id=CORE2-ARCTIC
ESG data portal – simplified publication

**Access option 3:** Browse dataset catalogue on thredds server

http://noresg.norstore.no/thredds/esgcet/3/CORE2-AMOC.v20140505.html

guide at http://wiki.met.no/noresm/norstoreesg