NcData

A proposed scheme for providing easy data interoperability between Iris and Xarray
(and netcdf files), as in Iris issue #4944 "Xarray bridge"

As has been suggested there, it seems possible that that a lower-level intermediate representation which offering exact netcdf file formatting control would be of value to Iris, but also possibly to Xarray users

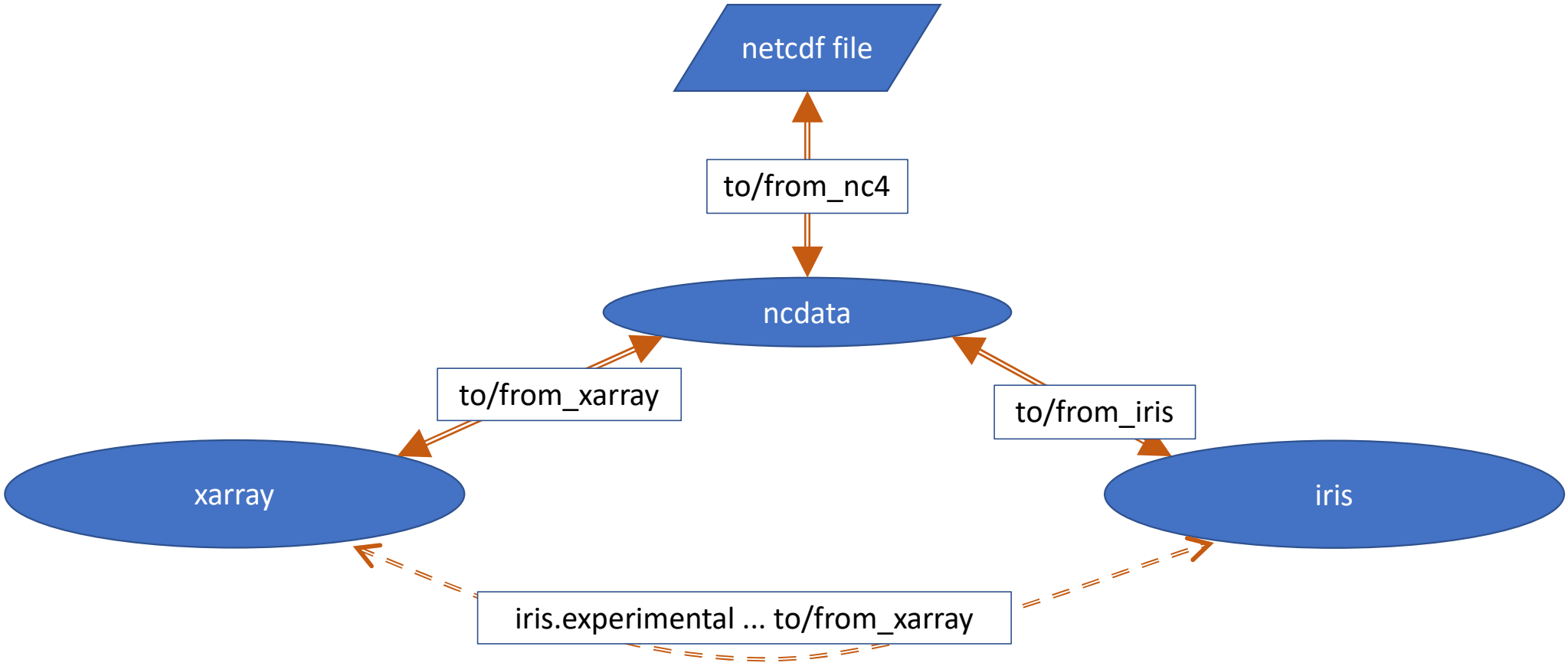
One suspects that the need for this may be much less from Xarray, than from Iris users.
Nevertheless, providing this as a separate repo, with no required dependency on Iris, seems attractive.

Current Outline Code :
most recent version
Notes as-of now (Jan 2023) :

- barely tested, but apparently ~working
- presented as a branch of Iris, but ripe for extraction into its own repo “**scitools/ncdata**”

Data Interoperability Operations

N.B.: NcData / Iris / Xarray exchanges are “copy-free and lazy-preserving”



NcData A new ***scitools*** package "scitools/ncdata".

A simple abstract representation of netcdf-style (CDM) data, with ...

- read/write interface to Xarray
- read/write interface to Iris
- read/write interface to netcdf files (via Python netCDF4)
- all "copy-free and lazy-preserving"

A possible package organisation :

- package `ncdata`
  - classes :
    `NcData` ; `NcVariable` ; `NcDimension` ; `NcAttribute`
  - subpackage `ncdata.xarray`
    - functions :
      - `to_xarray(ncdata) -> xarray.Dataset`
      - `from_xarray(xarray.Dataset)-> NcData`
  - subpackage `ncdata.netcdf4`
    - functions :
      - `to_nc4(netCDF4.Dataset)`
      - `from_nc4(netCDF4.Dataset) -> NcData`
  - subpackage `ncdata.dataset_like`
    - classes :
      `Nc4DatasetLike` ; `Nc4VariableLike`

Third-party Package dependencies

`ncdata`	→ **`numpy`**
`ncdata.xarray`	→ **`xarray`**
`ncdata.netcdf4`	→ **`netCDF4`**
`ncdata.dataset_like`	→ **`(? nil)`**

(netCDF4 probably not actually required)

Note: testing will require netCDF4, xarray and iris (→dask).

None of those is a required dependency for operation (depending on usage)

Notes : the less obvious

- ncdata contains no specific knowledge of Iris (*), and very little of Xarray
- ncdata may be used with any of netCDF4 / Xarray / Iris ***alone***
  - for low-level data wrangling
  - all of these are ***optional*** dependencies : unused packages (formats) are not required
- The Nc4DatasetLike ***could*** be used by other packages to interface to “ncdata”
  - however, compatibility is currently minimal == just enough to support Iris load+save
  - so, other uses could require it to become “***less minimal***”

(*) except *: **which*** bits of netCDF4.Dataset API the Nc4DatasetLike must support

Development To-Dos

- **ncdata :**
  - **write it ; provide repo ; publish on PyPI and conda-forge**
  - todo (Jan 2023) : docs, tests, repo configuration, package distribution
- **Xarray** : no code changes
  - update section on alternative packages and interoperation (Iris)
  - ? amend notes on status of existing to_iris/from_iris ?
- **Iris** :
  - support nc4-dataset(-like) for netcdf load + save
    - see https://github.com/SciTools/iris/pull/5024 "**Load+save things other than filepaths**"
  - support "shortcut" for direct transfers of array data during "netcdf load+save"
    - see PoC code solutions
  - add experimental utility routines : to_xarray / from_xarray
    - - these mean user ***does not need to import, or use 'ncdata' themselves***.