Gfdnavi: A tool to archive, share, distribute, analyze, and visualize geophysical fluid data and knowledge



Takeshi Horinouchi (Kyoto Univ → Hokkaido Univ(soon)),
Seiya Nishizawa (Kyoto Univ → Kobe Univ),
Chiemi Watanabe (Ochanomizu Univ),
T. Koshiro, A. Tomobayashi, S. Otsuka,
Y. Morikawa, Y.-Y. Hayashi, M. Shiotani, and
GFD Dennou Club (Davis project)

What is Gfdnavi

- = Geophysical fluid data navigator
- A suite of software to construct Webbased database of geophysical fluid data
- Functionality:
 - Search
 - Data analysis and visualization
 - Documentation of analysis results
- Available:

http://www.gfd-dennou.org/arch/davis/gfdnavi/

Background

Problems of Web-based database and analysis tools

- Limited analysis capability
 → We often end up with downloading data
- Not very suitable to desktop use
 Service are not available to local data

We would rather like to extend desktop tools (such as IDV) to cover persistent data services

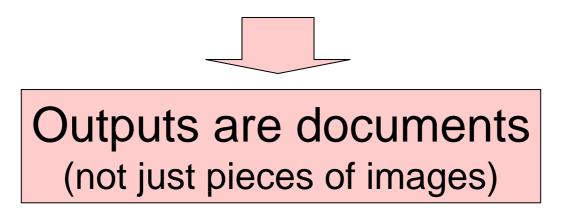
More on the analysis capability

- Impossible to predefine sufficient functionality (since we are scientists)

 Programmability is the key
- Programmability in two ways:
 - Programmable on web-browser
 - Web-service API (program locally)
 - \rightarrow Both are desirable

Visualization is not the goal

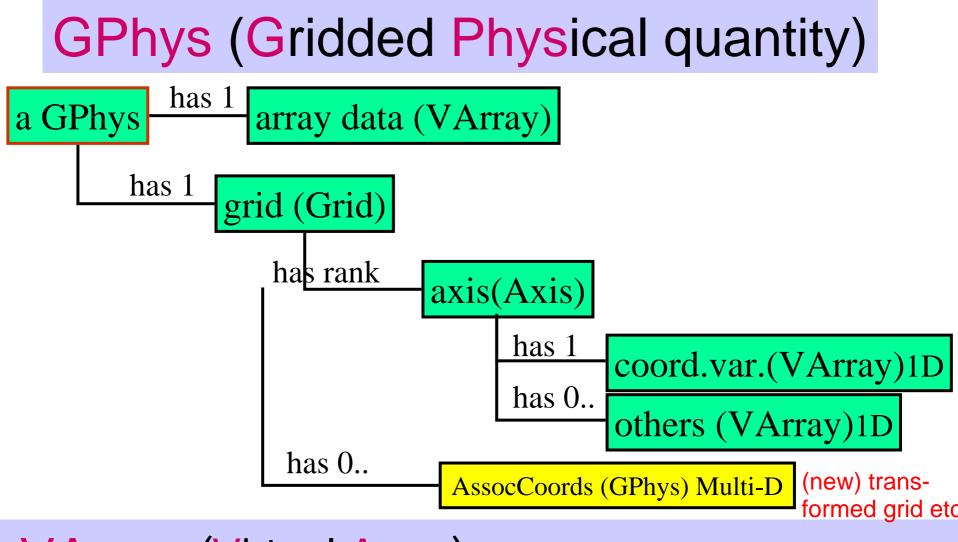
- To others (scientists / society): reports
- While working: memos / internal documents
- To collaborators: reports / know-how / discussion



Foundation of Gfdnavi

Two fundamental libraries used to build Gfdnavi (open-source)

- **GPhys** a Ruby library to analyze and visualize geophysical fluid data (*by Horinouchi etc since 2003*)
 - For consolidated access to data in files (NetCDF, GRIB, GrADS, NuSDAS, HDF5-EOS) or on runtime memory – A community infrastructure for data analysis [http://ruby.gfd-dennou.org/] (since 1999)
- Ruby on Rails Development framework for Web application (since 2005)
 - Made it drastically easy to develop Web applications with RDB
 - Written in/for Ruby \rightarrow We can use GPhys directly



VArray (Virtual Array) – with attributes (incl. units) Abstracts Data Storage (Entity can be in file(s) or multi-D Array on memory; can also be a mapped subset of another VArray or aggregation of VArrays)

Why do we use Ruby?

- Since we wanted a language for daily data analysis
 - Easy (fast) to write
 - Interactive use \rightarrow like GrADS

- Python is also fine (but we love Ruby)

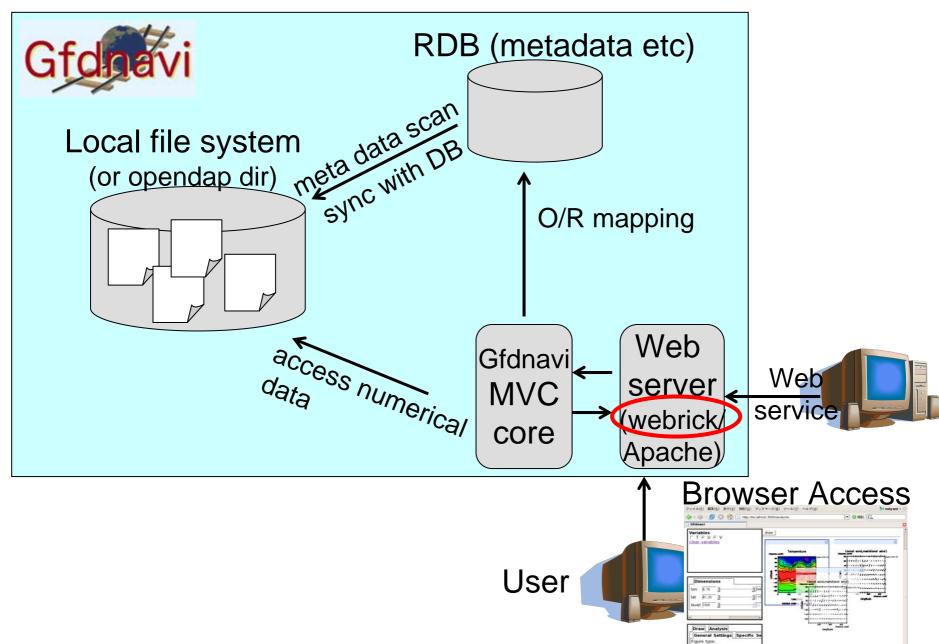
Introducing Gfdnavi

Since 2006

Early History (Aug 2006):

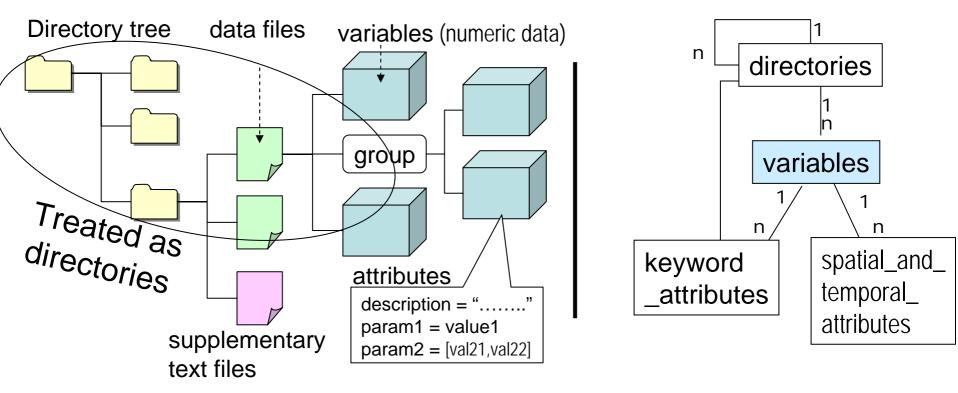
- Rough design by Horinouchi etc (at a meeting of the GFD-Dennou davis project)
- First implementation by S Nishizawa In two weeks (since then he is the most contributing to its development)

Overview



Metadata DB

- Metadata
 - name-value attributes; with a few standard field names
 - geospatial- and time-coordinate info
 - size, user info etc
- Directory structure (inherit metadata from parent directories)
- Generated by automatic scan (with a command)
 - variables: reading attributes through GPhys
 - directories: directory name and "Readme"-type texts

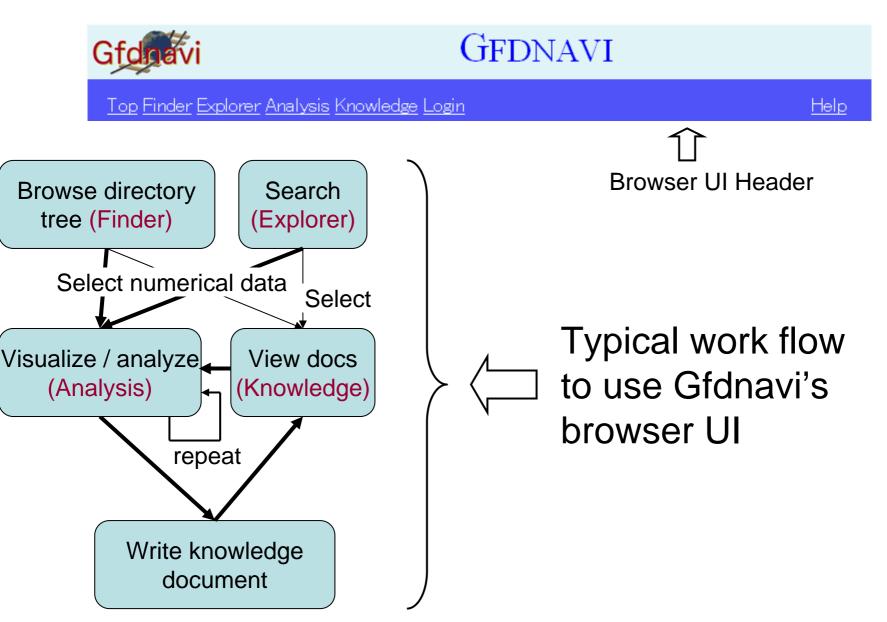


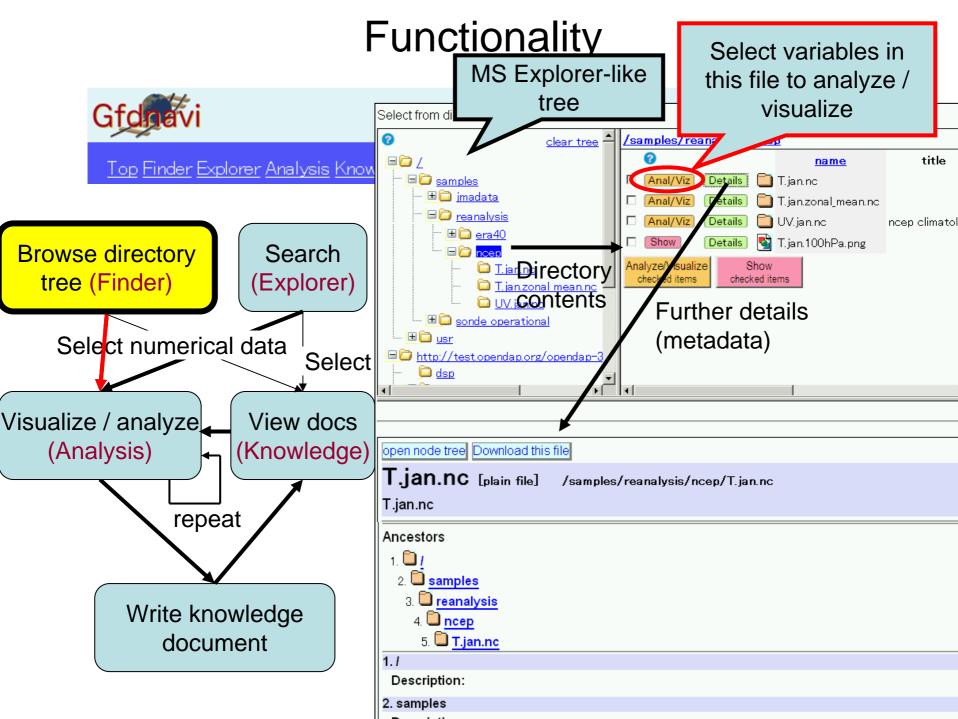
User Interface

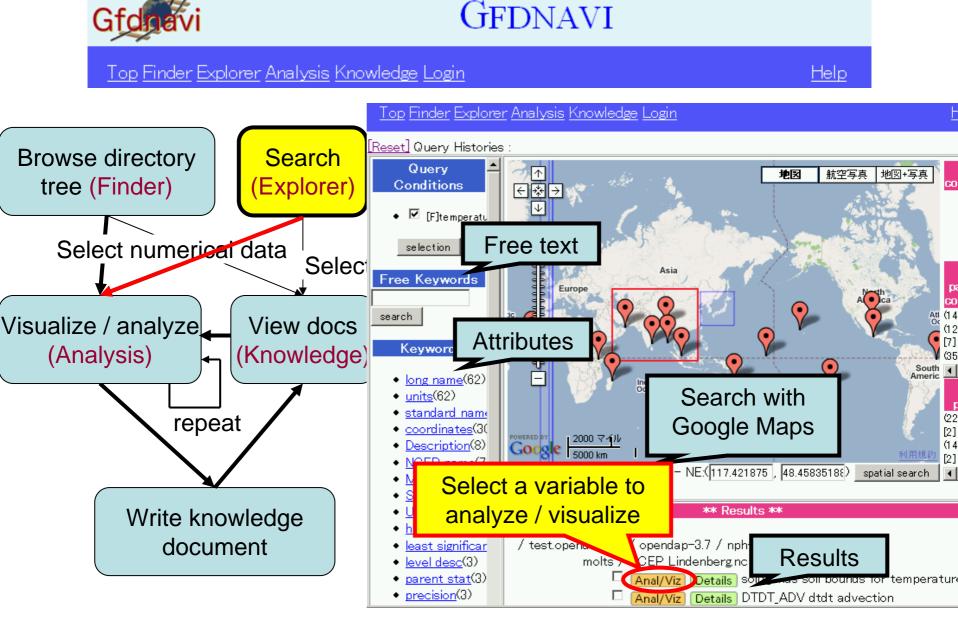
Home : Independent simple html \rightarrow replaceable

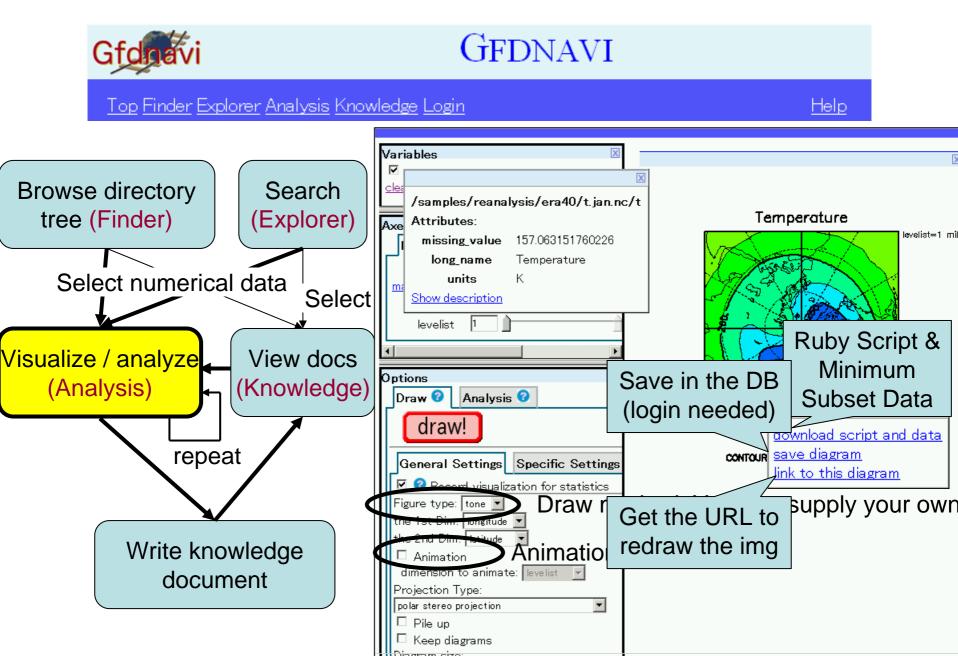


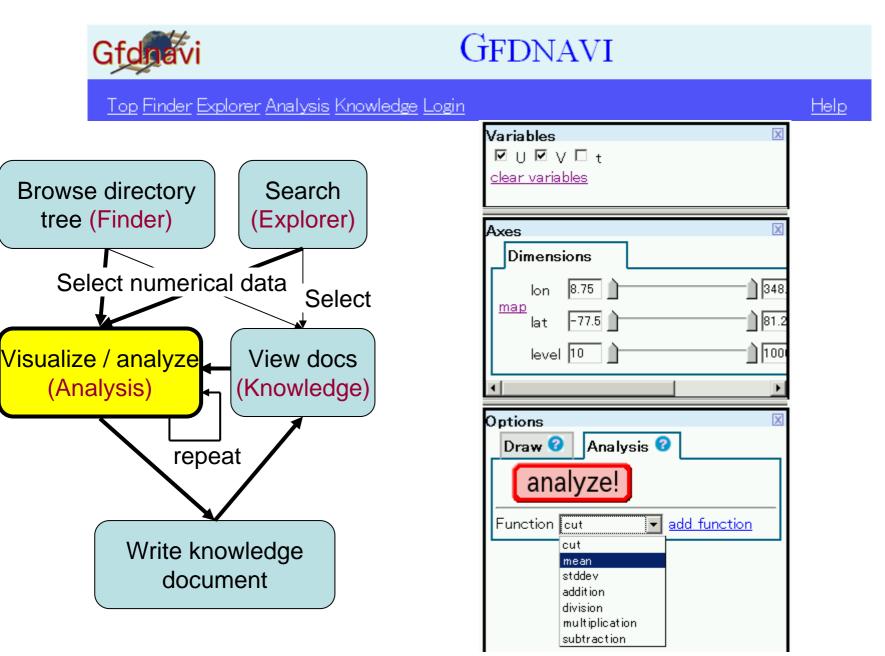
🥹Gfdnavi - Mozilla Firefox		
ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルブ(H)		\sim
	▼ ▶ Google	Q)
Gfdavi GFDNAVI		
<u>Top Finder Explorer Analysis Knowledge Login</u>		<u>Help</u>
<u>Search (Explorer)</u> <u>Search (old version)</u>		
Select from directory tree:		
Image: Clear tree localhost Image: Clear tree Image: Clear tree		

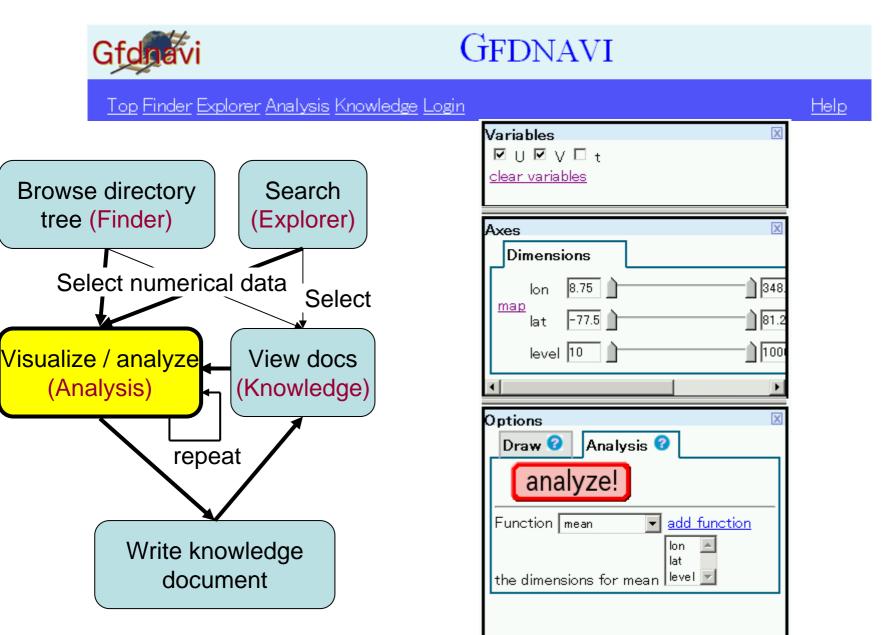








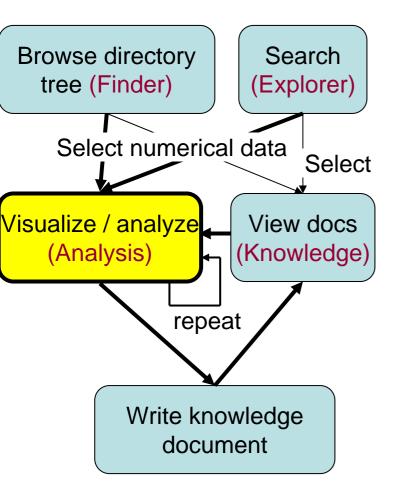






GFDNAVI

Top Finder Explorer Analysis Knowledge Login



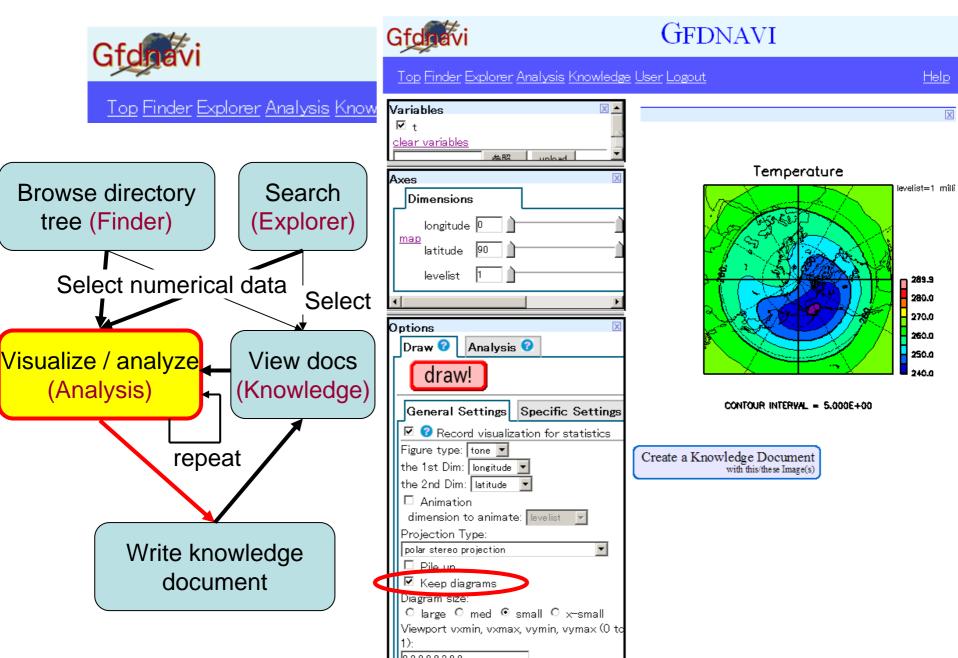
Top Finder Explorer Analys

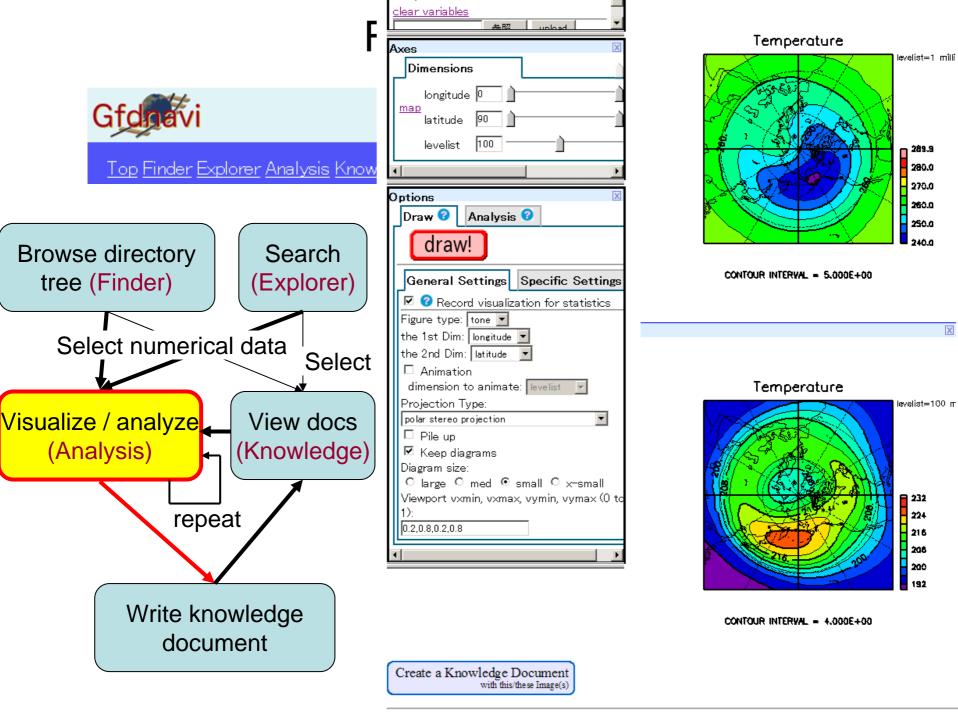
<u>Help</u>

Please login

Login: Password: login <u>signup</u>

Gfdazyi	Create Function	
	Function	
<u> </u>		spectrum
Browse directory tree (Finder) Search (Explorer)	save directory description	/usr/root/functions FFT ^2 along a specified dimension
Select numerical data Select Visualize / analyze	group number of input variables number of arguments	only me groups Create group 1 Create group
(Analysis) (Analysis) (Knowledge) Write knowledge document	script	{ arg0, gphys0 [gphys0.fft(arg0).abs ** 2]
	number of output variables	, ₃ 1 ▼
	Function Arguments	

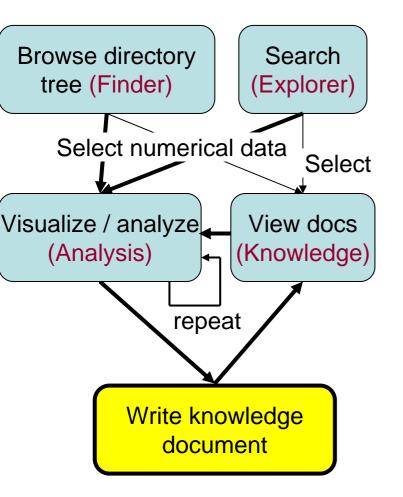






GFDNAVI

Top Finder Explorer Analysis Knowledge Login



Top Finder Explorer Analys

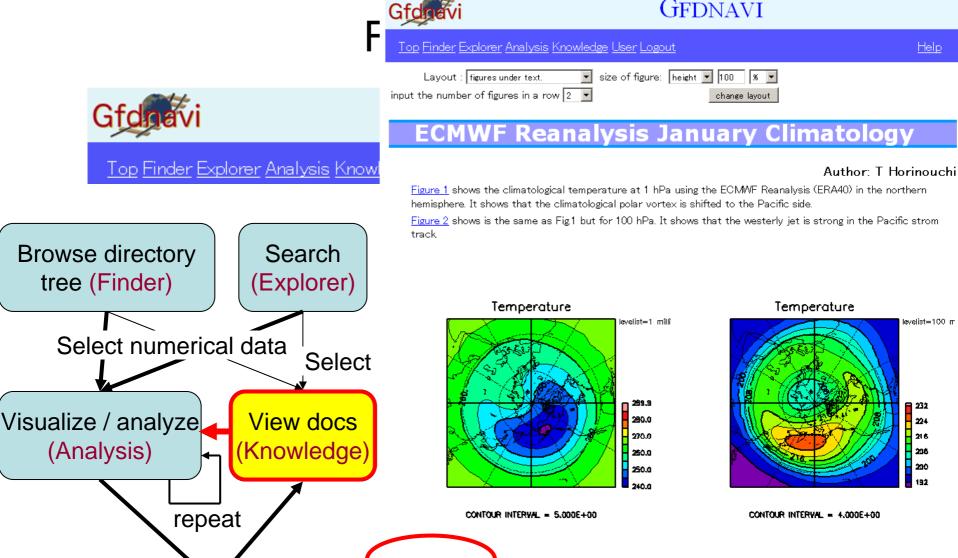
<u>Help</u>

Please login

Login: Password: login <u>signup</u>

FL <u>Save Images and Create a New Knowledge</u>

Gfdravi Top Finder Explorer Analysis Knowle	Title: ECMWF Reanalysis January Climatology Author: T Horinouchi Textbody: ((<figure 1="">)) shows the climatological temperature at 1 hPa using the ECMWF Reanalysis (ERA40) in the northern hemisphere. It shows that the climatological polar vortex is shifted to the Pacific side. ((<figure 2="">)) shows is the same as Fig.1 but for 100 hPa. It shows that the westerly jet is strong in the Pacific strom track.</figure></figure>
Browse directory tree (Finder) Search (Explorer)	Path: /usr/root/knowledge/ <u>tmp/eraT</u> knlge ex) /usr/root/knowledge/ <u>folder1/folder2/writing</u> knlge visible to:
Select numerical data Select Visualize / analyze (Analysis) View docs (Knowledge)	Caption: ERA Jan T at 1 hPa File Name: image001 Caption: Capt
repeat Write knowledge document	Figure 2 Caption: ERA Jan T at 100 hPa File Name: image002 Compression View this image in the original size



<u><redraw this image></u> <u><Get the URL></u> Fig. 2. ERA Jan T at 100 hPa

There are no comment on this document.

Get the URL>

at 1 hPa

Path: /usr/root/knowledge/tmp/eraT.knlge

Write a Coment on this document. Back to List

<u> <redraw this image></u>

in 1. ERA ام

Edit Back to List

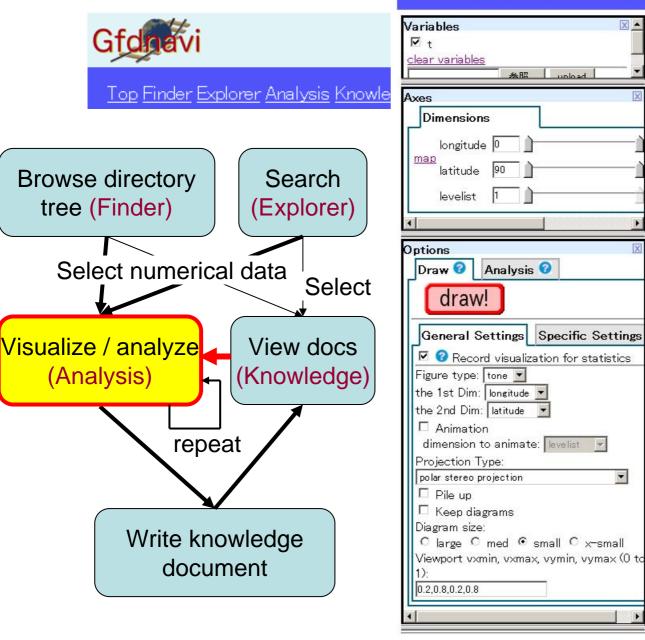
Write knowledge

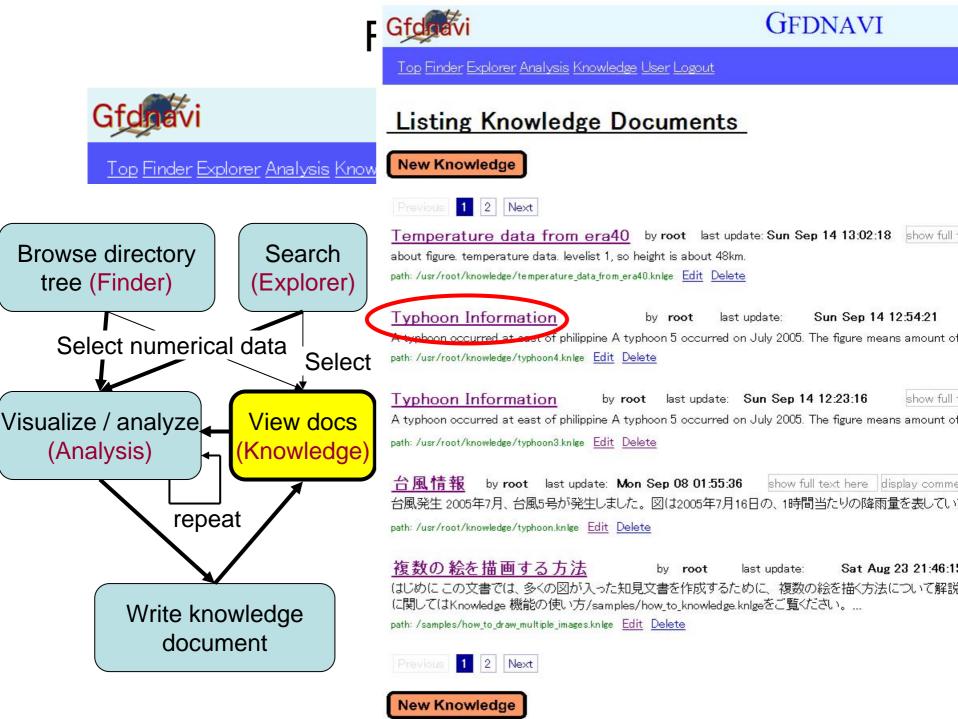
document



GFDNAVI

Top Finder Explorer Analysis Knowledge User Logout



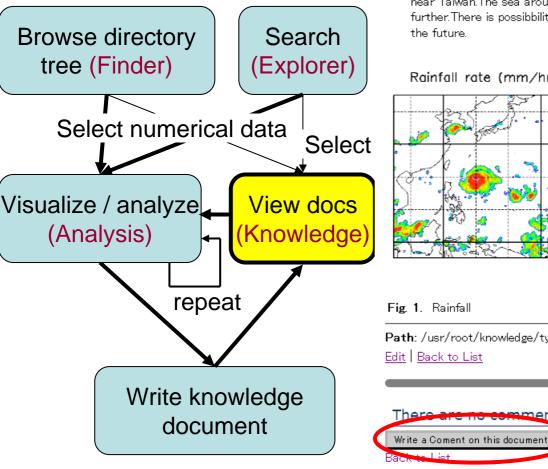


Typhoon Information

Author: Akinor



Top Finder Explorer Analysis Know



A typhoon occurred at east of philippine

A typhoon 5 occurred on July 2005.

The figure means amount of rainfall per hour at July 16, 2005.We can see a typhoon east of philippine.

Forecast of Course of typhoon

According to the forecast of Japan Meteorological Agancy, typhoon 5 will change direction of moveme near Taiwan.The sea around Okinawa is warmed by the intense heat of days, so it seems that the typ further. There is possibbility of comming off from the expectation and landing on West Japan. Please no the future.

Rainfall rate (mm/hr)

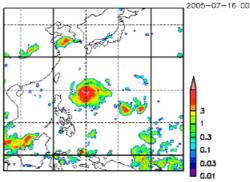
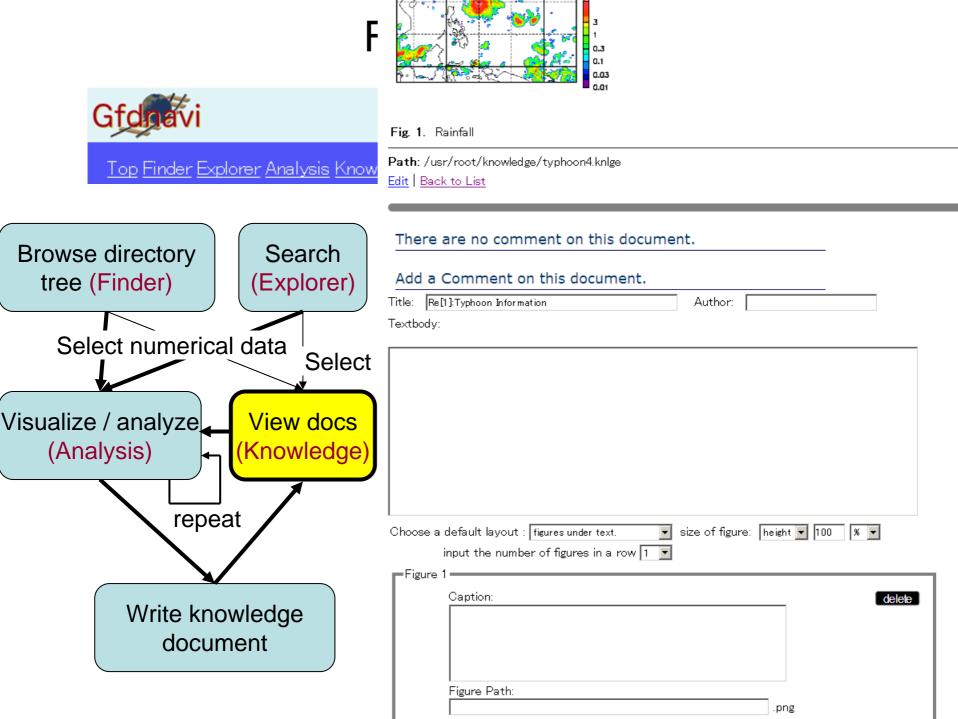


Fig. 1. Rainfall

Path: /usr/root/knowledge/typhoon3.knlge Edit | Back to List

mment on this document.



A typhoon occurred at east of philippine

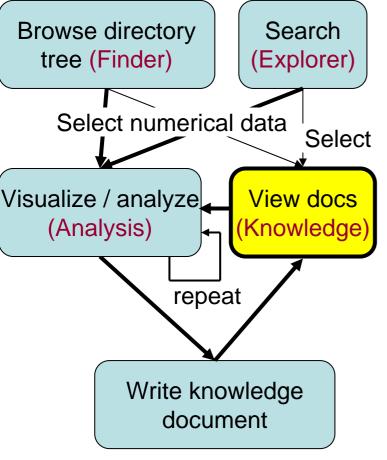
A typhoon 5 occurred on July 2005.

The figure means amount of rainfall per hour at July 16, 2005. We can see a typhoon east of philippine

Forecast of Course of typhoon

According to the forecast of Japan Meteorological Agancy, typhoon 5 will change direction of movement near Taiwan. The sea around Okinawa is warmed by the intense heat of days, so it seems that the typ further. There is possibbility of comming off from the expectation and landing on West Japan. Please in the future.

Rainfall rate (mm/hr)



Top Finder Explorer Analysis Know

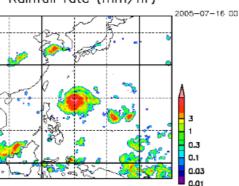


Fig. 1. Rainfall

Path: / usr/root/knowledge/typhoon3.knlge

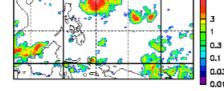
Edit Back to List

1 comment exists.

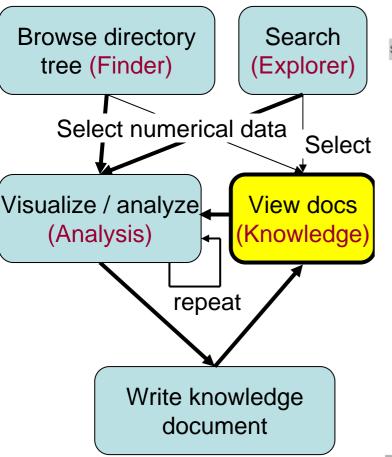
Show Summary of Comments Hide Comments Show full text of Comments

Re[1]:Typhoon Information author: Akinori Tomobayashi by root last update: Sun Sep After all, typhoon 5 went for Taiwan and landed China.

_Write a Coment on this document. Back to List







Path: /usr/root/knowledge/typhoon3.knlge Edit | Back to List

1 comment exists.

Show Summary of Comments Hide Comments Show full text of Comments

Re[1]:Typhoon Information author: Akinori Tomobayashi by root last update: Sun Sep After all, typhoon 5 went for Taiwan and landed China.

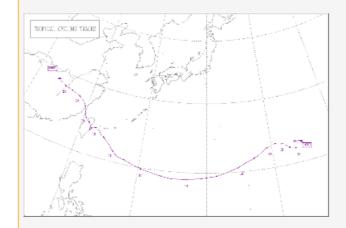


Fig. 1. course of typhoon 5 in 2005.

Path: /usr/root/knowledge/typhoon3_comment_1.knlge

Write a Coment on this document.

<u>Back to List</u>

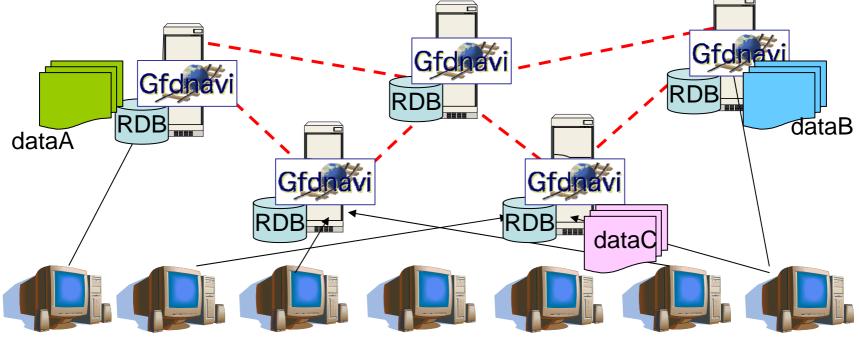
Web service

→ Tomorrow by Seiya Nishizawa

Network of Gfdnavi

Under development by C Watanabe (Ochanomizu Univ)

- ✓ To create peer-to-peer network for cross search and cross use **among Gfdnavi servers**
- Then one can access local data and remote data together



Summary

- Novel features of Gfdnavi
 - Seamless coverage from desktop use to public data service (by having custom web server)
 - Programmability (on browser & by web service)
 - Documentation of analysis results (dynamically reproducible/extendible) (→ memos / reports / PR / Blog for scientific collaboration)
- Good implementation
 - Extendibility (by using GPhys)
 - Swift development (by using RonR)

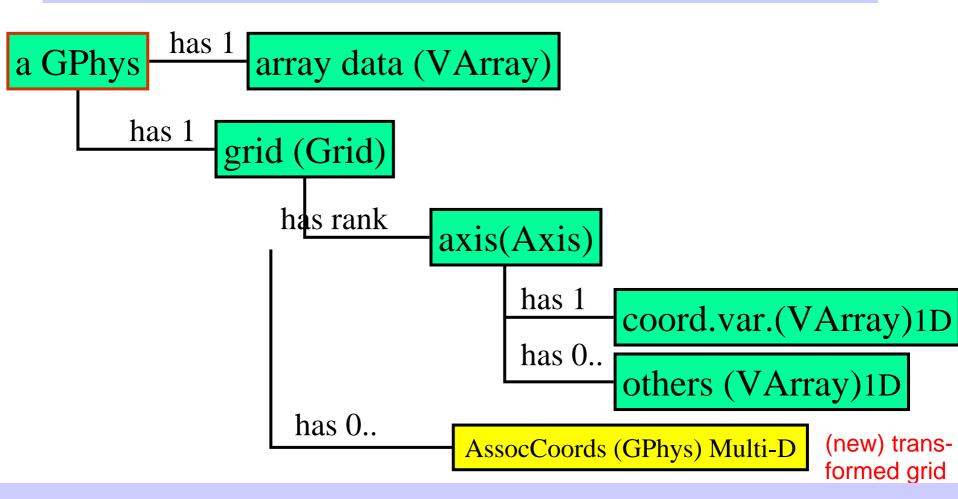
Tomorrow by S Nishizawa

Future Outlook

- Support Networking → Create a Web of scientific data & knowledge
- Increase analysis & visualization functionality (many needed)
- Improve remote API accesses (tomorrow's topic)

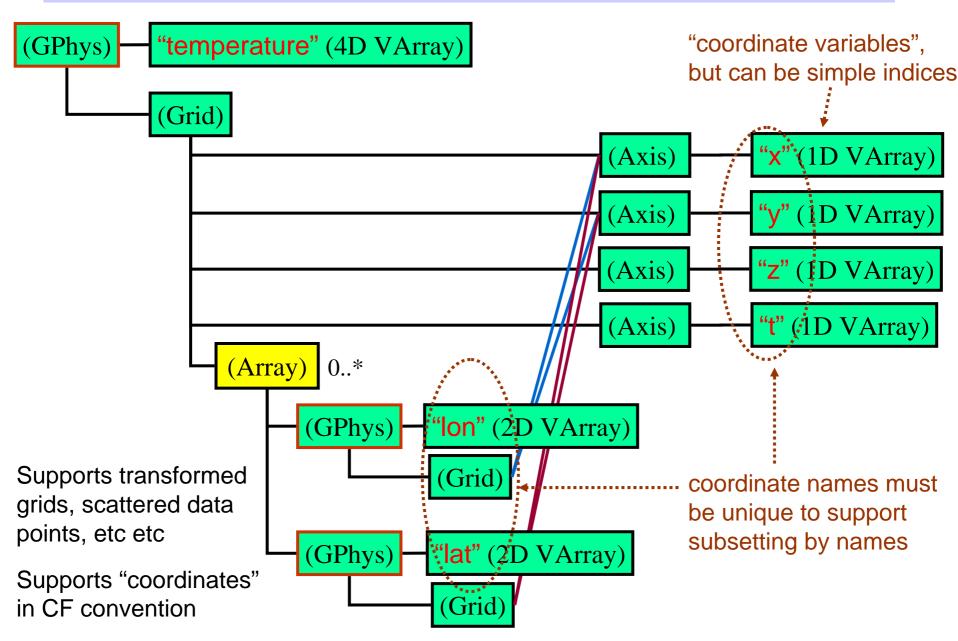
fin

GPhys (Gridded Physical quantity)



VArray (Virtual Array) – Abstracts Data Storage (Can be in file(s) or multi-D Array on memory; can also be a subset or aggregation of (an)other VArray(s))

Example of GPhys's associated coordinates





What is Ruby on Rails

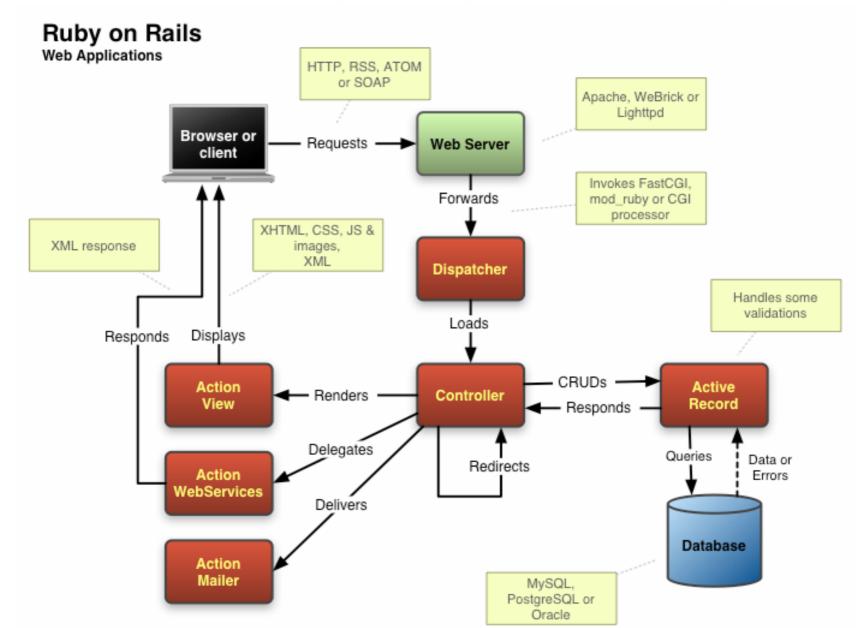
http://www.rubyonrails.org/



- Web development framework in Ruby
- With RDBMS (Mysql, Postgres, SQL Server, SQLite etc)
- Strong prototyping (e.g. Model-View-Controller (MVC) stucture)
- Comprehensive library (covering Ajax and Web service)
- Ruby-embedded html
 → suitable to use our Ruby library
- Has a private web server (Webrick); also runs on Apache, lighttpd etc
 - → One can personally run a web server anywhere with arbitrary port

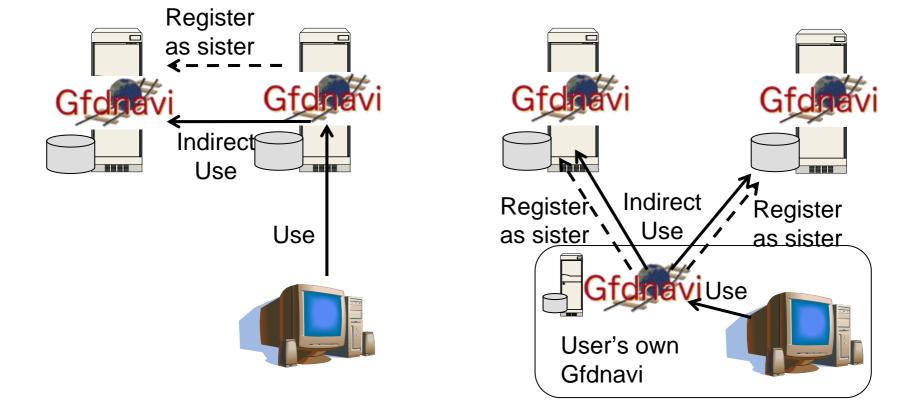
From *"Understanding Rails MVC"*:

http://wiki.rubyonrails.org/rails/pages/UnderstandingRailsMVC

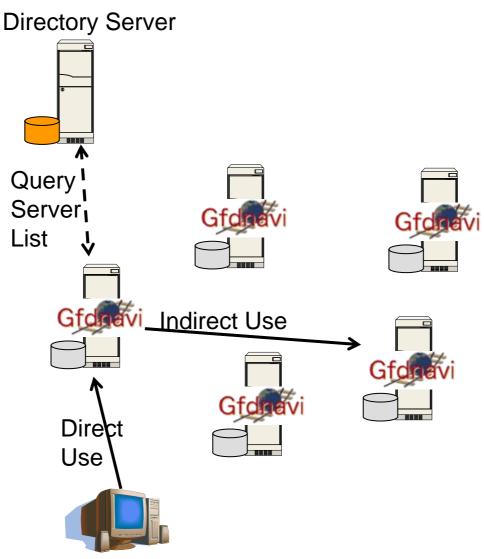


Sister-server method

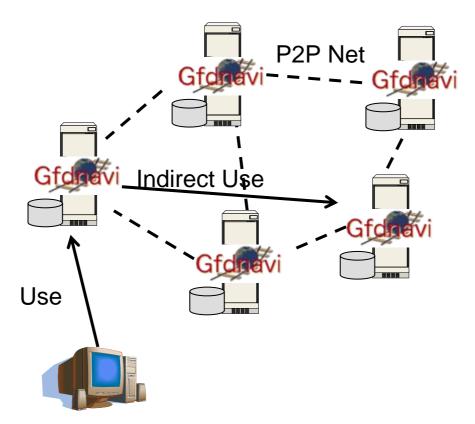
(a) Basic case: available in LAS. User can't choose peers (b) Gfdnavi: one can register any peer by running a Gfdnavi



P2P with directory server



Overlay network by P2P



Currently tested by C. Watanabe by using Overlay Weaver (Java-based p2p library) and Rails' Action Web Service – Decentralized p2p with distributed hash tables (DHT)

copy from old slides

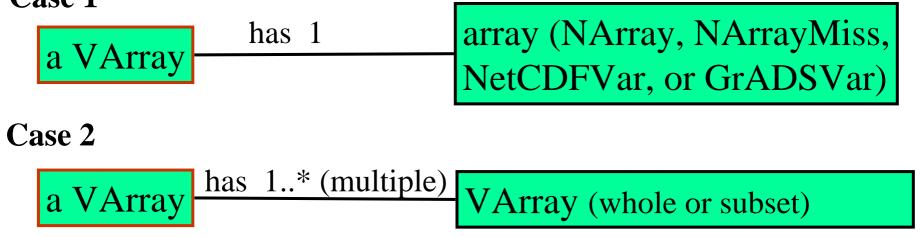
last revised: 2004/06/08

GPhys A class of gridded physical quantities

Takeshi Horinouchi (RISH, Kyoto Univ.)

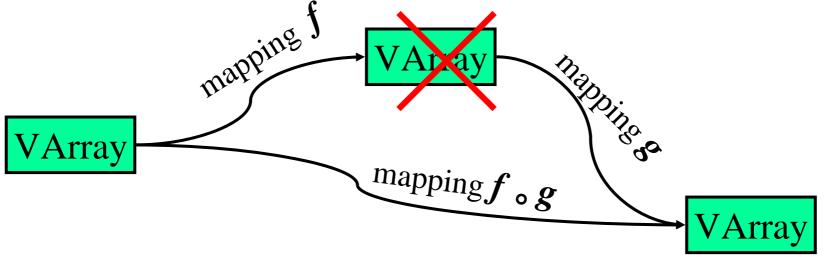


- Virtual Array. A class of Ruby (written in pure ruby), which represents array data in GPhys
- A VArray object behaves as an array, but its contents can be on various media: (case 1) simply a multi dimensional array on memory (NArray), or data in a NetCDF file (in this case, a file pointer is stored), or GrADS data; (case 2) It can also represent a subset of another VArray or multiple VArrays tiled.
- Can have attributes as variables in NetCDF datasets
- In reality, NetCDF are handled by a subclass VArrayNetCDF etc.etc. Case 1



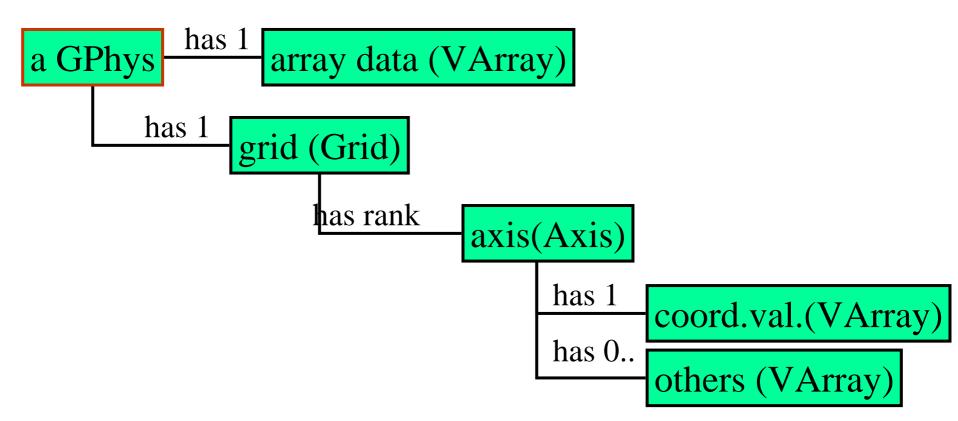
subset mapping of VArray

- Always kept direct by compositing mappings, in order to prevent long chains (see the figure below).
- Subset slicing (by such as va[0..10,3]) is done by subset mapping, not by making actual data extraction, if not explicitly specified otherwise. Therefore,
 - Computationally efficient
 - Suitable for writing in subsets of data in files.
- In other words, actual data cutting is deferred until needed to defer operations until needed is a policy of GPhys construction



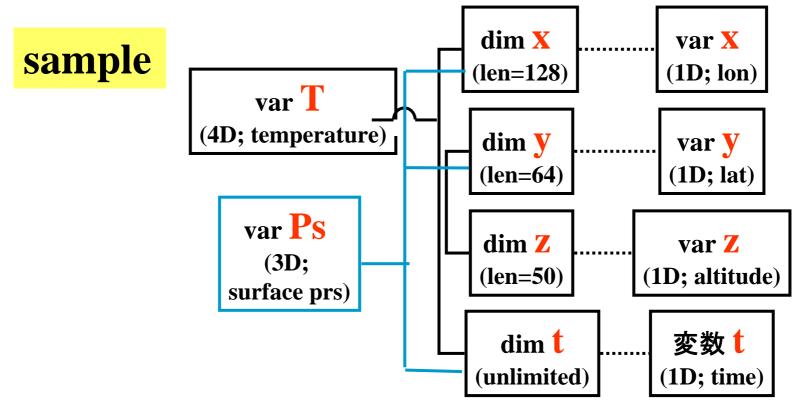
Structure of GPhys

- Consists of a grid (coordinates) and multi dimensional array data
- Can conduct mathematical operations (a GPhys behaves like an numeric array)



For your reference: Coordinates in NetCDF dataset

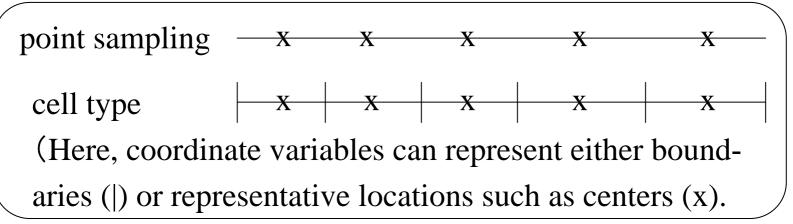
- Variables that have same names as dimensions hold coordinate values (locations)
- Weak point: this rule can be violated



Can construct GPhys objects along the trees

More on cooridnates

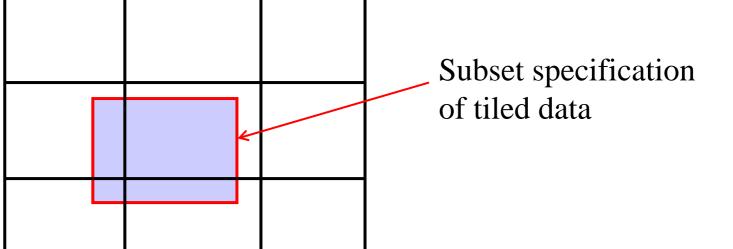
- 3 cases are prepared
 - point sampling
 - cell type
 - simple sequence (though it's not physical)



- For instance, how to integrate along an axis is known by the axis. GPhys simply requests the integration to its Grid, and the Grid ask it to the corresponding Axis. By default, trapezoidal formula is used if point samples or cell boundaries. (can be changed by users)
- If NetCDF data are read, those types are configured if the convention used supports such discrimination. (so far, convention support is weak, though)

Tiling

- Data divided into "tiles" can be treated as one consolidated GPhys object. Convenient to handle long time sequence divided by periods (such as by years) or outputs from parallel simulations on distributed-memory machines. Tiling is done by VArrayComposite.
- Subsets can be handled (see the figure below)
- May be applicable to parallel simulations in future?
- So far, automatic configuration is available only for NetCDF, by using an Array or Regexp (e.g., /data_x(¥d)y(¥d).nc/ for data_x0y0.nc, data_x0y1.nc, data_x1y0.nc, data_x1y1.nc)



Big data handling

- Iterator to handle data too big to read on memory at once.
 - GPhys::IO.each_along_dims_write the result also written in file (since the result of operations is often big too.) Another type of iterator is planned but yet to be implemented.
- Example:
 - Without the iterator:

```
in = GPhys::IO.open(infile, varname)
ofile = NetCDF.create(ofilename)
out = in.mean(0)  # now, the entire result is on memory
GPhys::IO.write( ofile, out )
ofile.close
```

```
— With the iterator, taking the last dimension to make a loop:
in = GPhys::IO.open(infile, varname)
ofile = NetCDF.create(ofilename)
out = GPhys::IO.each_along_dims_write(in, ofile, -1){ |in_sub|
[in_sub.mean(0)] # written in ofile each time
}
ofile.close
```

Units of physical quantities

- Handled by NumRu::Units (by E Toyoda)
- mlt,div,etc.: handled as should be
- add,sub: the units of the first term is inherited
 - e.g., addition of [m] and [km] is done after multiplying the second term by 1000. Warning is made if the units are incompatible (in that case, no conversion is made).
- Introduced a scalar numeric class with units UNumeric
 - GPhys, VArray, and UNumeric recognize one another (stronger to weaker in this order)
 - Example: to multiply the Coriolis parameter with a GPhys object u representing winds [m/s]:

f = UNumeric[1e-4,"s-1"] coriolis_frc = f * u # then the units will be in m.s-2

Distributed objects using dRuby

- Data service to remote clients
 - gphys-remote: a simple directory service (like the anonymous ftp, directories and data (in which GPhys objects can be defined) under a top directory is made accessible to remote hosts.
 - gave (GUI): can connect to gphys-remote server