



the globus alliance

www.globus.org

Virtual Data Language exemplar: MRI atlas generation

Mike Wilde

University of Chicago /
Argonne National Lab



Virtual Data is a Large Team Effort

The GriPhyN Virtual Data System

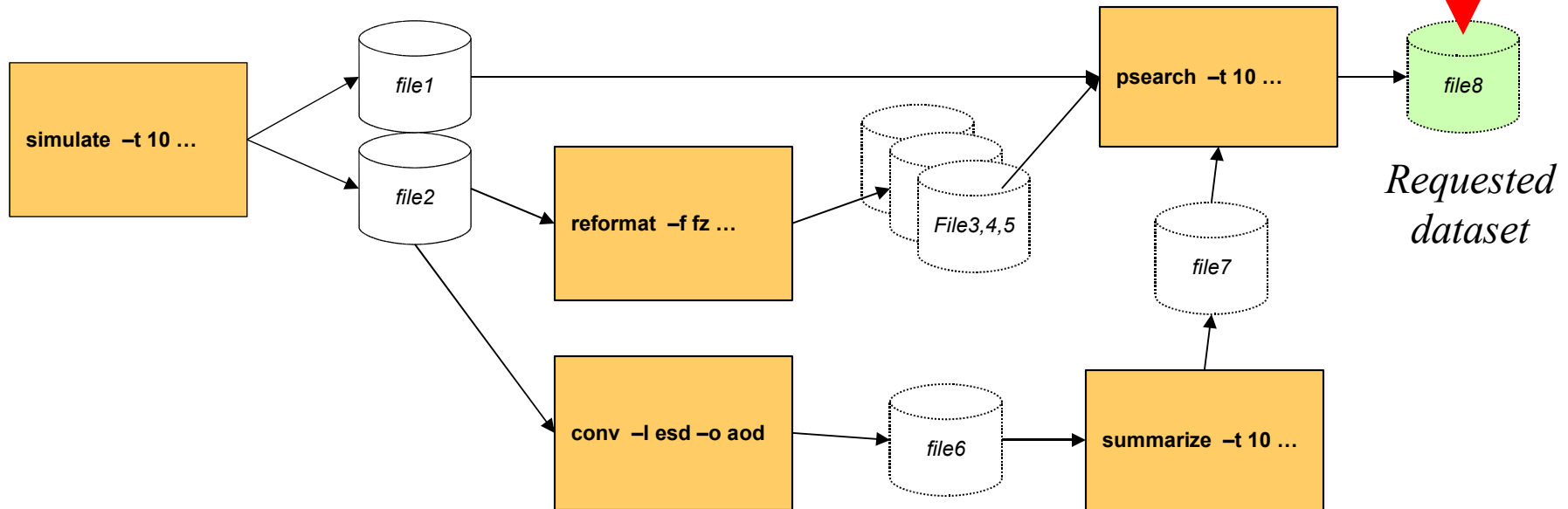
is the work of Ian Foster, Jens Voeckler, Mike Wilde and Yong Zhao, University of Chicago and Argonne National Laboratory, and Ewa Deelman, Gaurang Mehta, and Karan Vahi, USC Information Sciences Institute

MRI Applications and datasets were provided by Jack Van Horn and Jed Dobson, fMRI Data Center, Dartmouth College



Virtual Data

Describes analysis workflow



- The recorded virtual data “recipe” here is:
 - Files: $8 < (1,3,4,5,7)$, $7 < 6$, $(3,4,5,6) < 2$
 - Programs: $8 < \text{psearch}$, $7 < \text{summarize}$,
 $(3,4,5) < \text{reformat}$, $6 < \text{conv}$, $(1,2) < \text{simulate}$



VDL: Virtual Data Language Describes Data Transformations

- Transformation
 - Abstract template of program invocation
 - Similar to "function definition"
- Derivation
 - "Function call" to a Transformation
 - Store past and future:
 - > A record of how data products were generated
 - > A recipe of how data products can be generated
- Invocation
 - Record of a Derivation execution
- These XML documents reside in a "virtual data catalog" – VDC - a relational database



VDL Describes Workflow via Data Dependencies



```
TR tr1(in a1, out a2) {  
  argument stdin = ${a1};  
  argument stdout = ${a2}; }
```

```
TR tr2(in a1, out a2) {  
  argument stdin = ${a1};  
  argument stdout = ${a2}; }
```

```
DV x1->tr1(a1=@{in:file1}, a2=@{out:file2});
```

```
DV x2->tr2(a1=@{in:file2}, a2=@{out:file3});
```

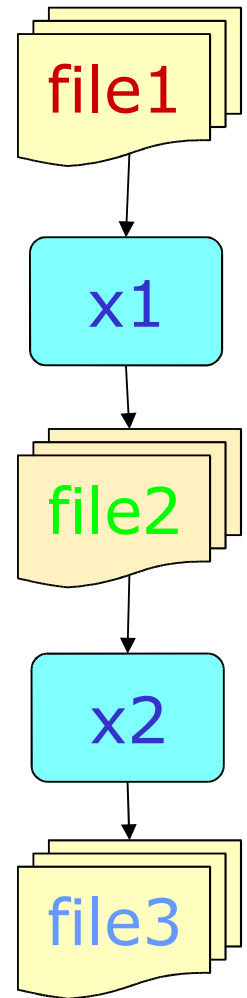




Image analysis algorithms used

- UCLA Automatic Image Registration (AIR)

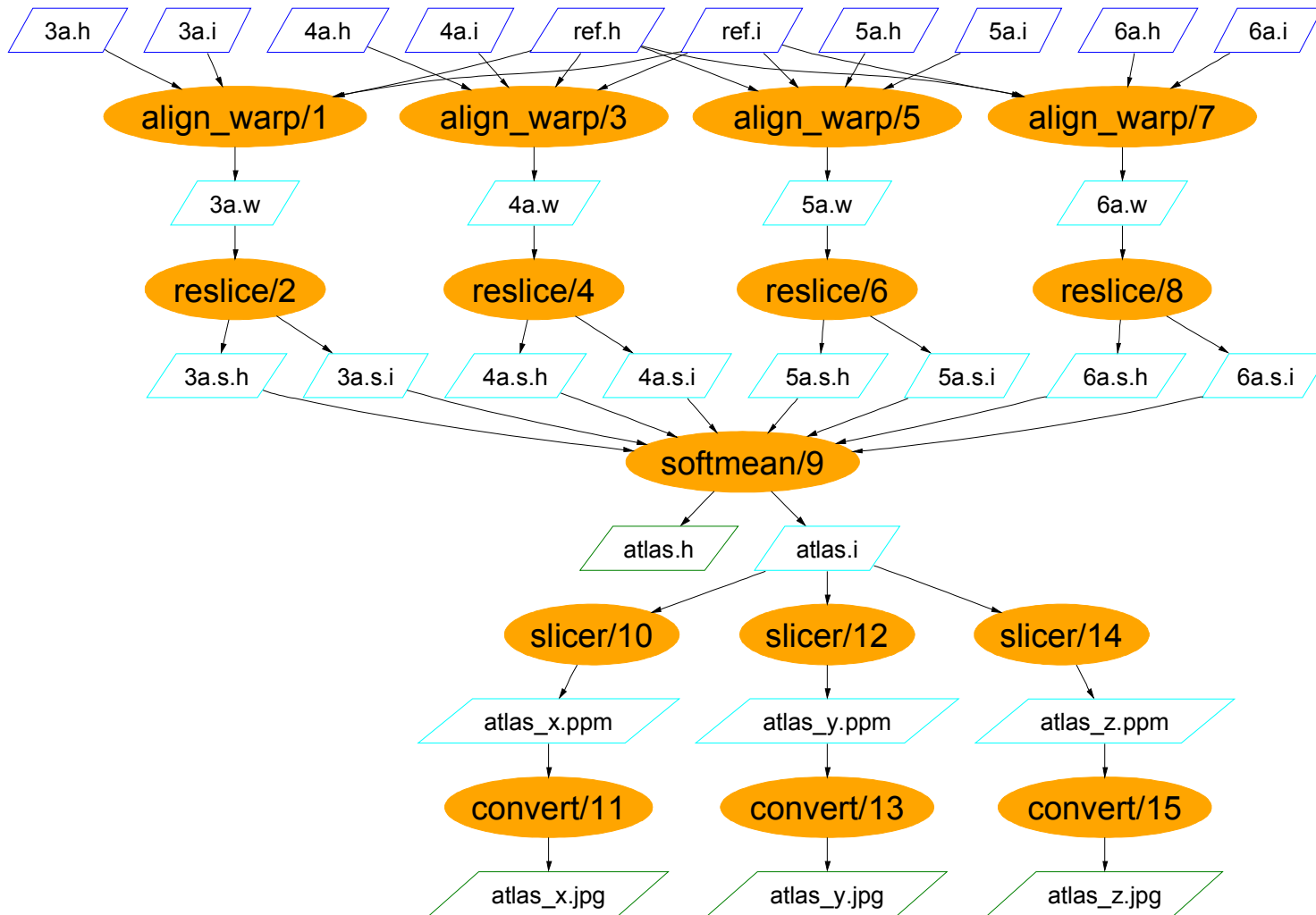


Automated
Image
Registration

- <http://bishopw.loni.ucla.edu/AIR5/index.html>



MRI ATLAS Generation Workflow



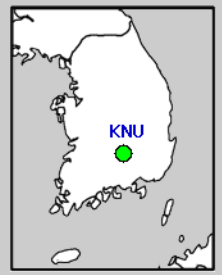
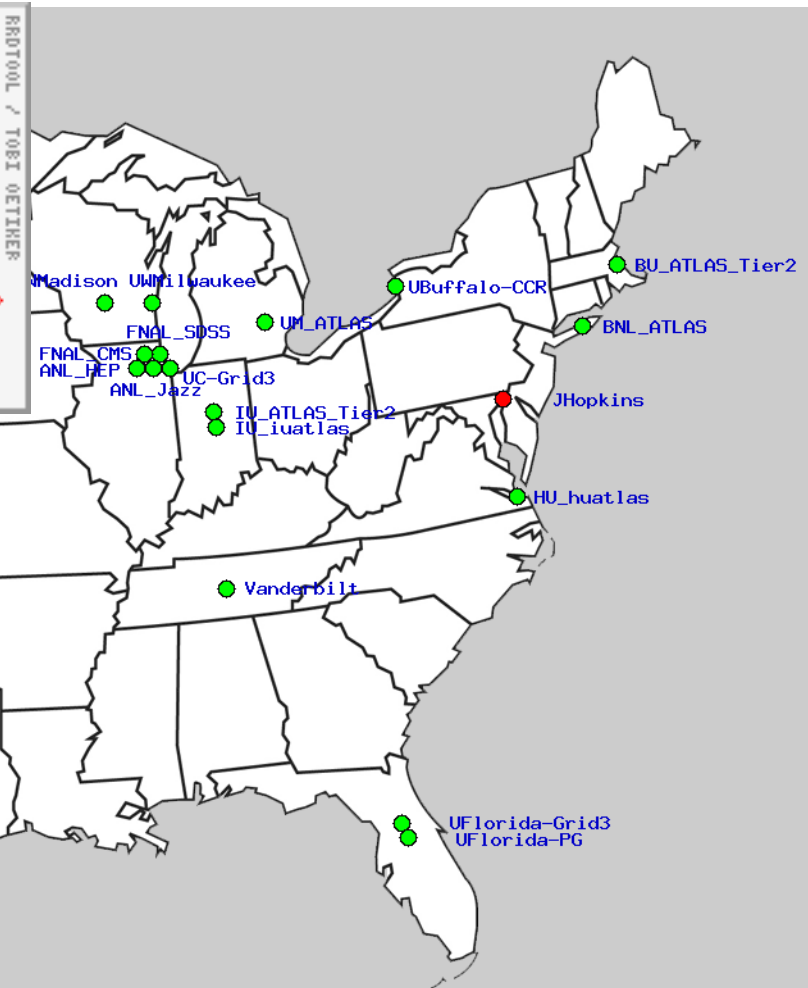
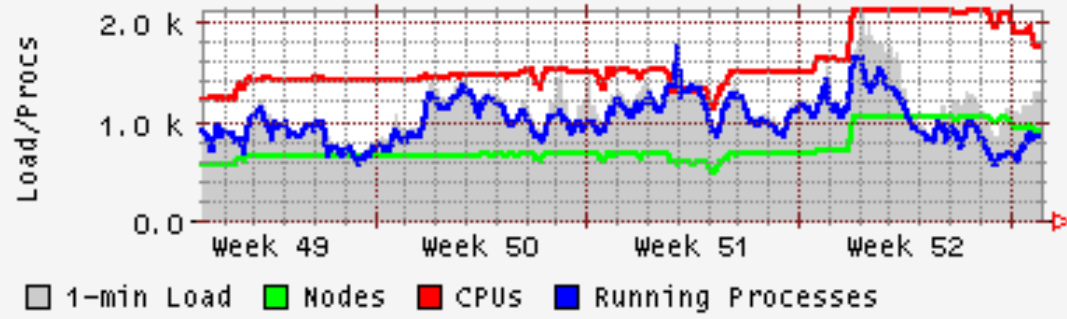


Deployment on Grid3

- international Data Grid with dozens of sites and thousands of processors.
- operated jointly by the U.S. Grid projects iVDGL, GriPhyN and PPDG, and the U.S. participants in the LHC experiments ATLAS and CMS.
- Project highlights include:
 - Participation by more than 25 sites across the US and Korea which collectively provide more than 2000 CPUs
 - Resources used by 7 different scientific applications, including 3 high energy physics simulations and 4 data analyses in high energy physics, bio-chemistry, astrophysics and astronomy
 - More than 100 individuals are currently registered with access to the Grid
 - A peak throughput of 500-900 jobs running concurrently with a completion efficiency of approximately 75%

Virtual Data Language Enables Grid-wide distributed computation

Grid3 Grid Load last month



South Korea



Grid 3 this morning...

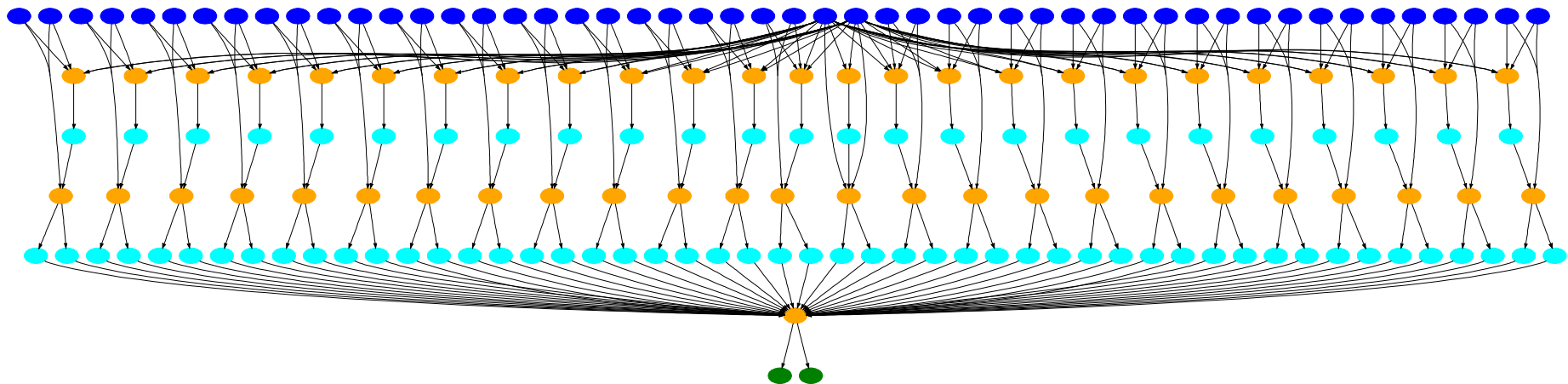


South Korea



Virtual Data Language Enables Grid-wide distributed computation

- Atlases from 25, 100, and 590 image datasets processed on distributed grid resources
- Test virtual data workflow locally, run same VDL code on wide-area Grid





```
TR air::align_warp( in reg_img, in reg_hdr, in sub_img, in sub_hdr, m, out warp )
{
    argument = ${reg_img};
    argument = ${sub_img};
    argument = ${warp};
    argument = "-m " ${m};
    argument = "-q";
}

TR air::reslice( in warp, sliced, out sliced_img, out sliced_hdr )
{
    argument = ${warp};
    argument = ${sliced};
}

TR air::warp_n_slice( in reg_img, in reg_hdr, in sub_img, in sub_hdr, m = "12",
    io warp, sliced, out sliced_img, out sliced_hdr )
{
    call air::align_warp( reg_img=${reg_img}, reg_hdr=${reg_hdr},
        sub_img=${sub_img}, sub_hdr=${sub_hdr},
        m=${m},
        warp = ${out:warp} );
    call air::reslice( warp=${in:warp}, sliced=${sliced},
        sliced_img=${sliced_img}, sliced_hdr=${sliced_hdr} );
}

TR air::softmean( in sliced_img[], in sliced_hdr[], arg1 = "y", arg2 = "null",
    atlas, out atlas_img, out atlas_hdr )
{
    argument = ${atlas};
    argument = ${arg1} " " ${arg2};
    argument = ${sliced_img};
}
```



```
TR air::align_warp( in reg_img, in reg_hdr, in sub_img, in sub_hdr, m, out warp )
{
    argument = ${reg_img};
    argument = ${sub_img};
    argument = ${warp};
    argument = "-m " ${m};
    argument = "-q";
}

TR air::reslice( in warp, sliced, out sliced_img, out sliced_hdr )
{
    argument = ${warp};
    argument = ${sliced};
}

TR air::warp_n_slice( in reg_img, in reg_hdr, in sub_img, in sub_hdr, m = "12",
    io warp, sliced, out sliced_img, out sliced_hdr )
{
    call air::align_warp( reg_img=${reg_img}, reg_hdr=${reg_hdr},
        sub_img=${sub_img}, sub_hdr=${sub_hdr},
        m=${m},
        warp = ${out:warp} );
    call air::reslice( warp=${in:warp}, sliced=${sliced},
        sliced_img=${sliced_img}, sliced_hdr=${sliced_hdr} );
}

TR air::softmean( in sliced_img[], in sliced_hdr[], arg1 = "y", arg2 = "null",
    atlas, out atlas_img, out atlas_hdr )
{
    argument = ${atlas};
    argument = ${arg1} " " ${arg2};
    argument = ${sliced_img};
}
```



```
DV air::i3472_3->air::warp_n_slice(  
  reg_hdr = @{"3472-3_anonymized.hdr"},  
  reg_img = @{"3472-3_anonymized.img"},  
  sub_hdr = @{"3472-3_anonymized.hdr"},  
  sub_img = @{"3472-3_anonymized.img"},  
  warp = @{"3472-3_anonymized.warp"},  
  sliced = "3472-3_anonymized.sliced",  
  sliced_hdr = @{"3472-3_anonymized.sliced.hdr"},  
  sliced_img = @{"3472-3_anonymized.sliced.img"} );  
  
DV air::i3472_4->air::warp_n_slice(  
  reg_hdr = @{"3472-3_anonymized.hdr"},  
  reg_img = @{"3472-3_anonymized.img"},  
  sub_hdr = @{"3472-4_anonymized.hdr"},  
  sub_img = @{"3472-4_anonymized.img"},  
  warp = @{"3472-4_anonymized.warp"},  
  sliced = "3472-4_anonymized.sliced",  
  sliced_hdr = @{"3472-4_anonymized.sliced.hdr"},  
  sliced_img = @{"3472-4_anonymized.sliced.img"} );  
  
...  
DV air::i3472_6->air::warp_n_slice(  
  reg_hdr = @{"3472-3_anonymized.hdr"},  
  reg_img = @{"3472-3_anonymized.img"},  
  sub_hdr = @{"3472-6_anonymized.hdr"},  
  sub_img = @{"3472-6_anonymized.img"},  
  warp = @{"3472-6_anonymized.warp"},  
  sliced = "3472-6_anonymized.sliced",  
  sliced_hdr = @{"3472-6_anonymized.sliced.hdr"},  
  sliced_img = @{"3472-6_anonymized.sliced.img"} );
```

MRI Example: AIR Tools



```
DV air::i3472_3->air::warp_n_slice(  
  reg_hdr = @{"3472-3_anonymized.hdr"},  
  reg_img = @{"3472-3_anonymized.img"},  
  sub_hdr = @{"3472-3_anonymized.hdr"},  
  sub_img = @{"3472-3_anonymized.img"},  
  warp = @{"3472-3_anonymized.warp"},  
  sliced = "3472-3_anonymized.sliced",  
  sliced_hdr = @{"3472-3_anonymized.sliced.hdr"},  
  sliced_img = @{"3472-3_anonymized.sliced.img"} );  
DV air::i3472_4->air::warp_n_slice(  
  reg_hdr = @{"3472-3_anonymized.hdr"},  
  reg_img = @{"3472-3_anonymized.img"},  
  sub_hdr = @{"3472-4_anonymized.hdr"},  
  sub_img = @{"3472-4_anonymized.img"},  
  warp = @{"3472-4_anonymized.warp"},  
  sliced = "3472-4_anonymized.sliced",  
  sliced_hdr = @{"3472-4_anonymized.sliced.hdr"},  
  sliced_img = @{"3472-4_anonymized.sliced.img"} );  
...  
DV air::a3472_3->air::softmean(  
  sliced_img = [  
    @{"3472-3_anonymized.sliced.img"},  
    @{"3472-4_anonymized.sliced.img"},  
    @{"3472-5_anonymized.sliced.img"},  
    @{"3472-6_anonymized.sliced.img"} ],  
  sliced_hdr = [  
    @{"3472-3_anonymized.sliced.hdr"},  
    @{"3472-4_anonymized.sliced.hdr"},  
    @{"3472-5_anonymized.sliced.hdr"},  
    @{"3472-6_anonymized.sliced.hdr"} ],  
  atlas = "atlas",  
  atlas_img = @{"atlas.img"},  
  atlas_hdr = @{"atlas.hdr"}  
)
```



For Information and Software

- Virtual Data System
 - www.griphyn.org/chimera: Overview, papers, software
- Grids and Grid Software
 - www.globus.org – The Globus Toolkit
 - www.cs.wisc.edu/condor - The Condor Project
 - www.griphyn.org/vdt - Virtual Data Toolkit
 - www.ivdgl.org/grid2003 - Using Grid3



Acknowledgements: *Virtual Data is a Large Team Effort*

The GriPhyN Virtual Data System is the work of Ian Foster, Jens Voeckler, Mike Wilde and Yong Zhao, University of Chicago and Argonne National Laboratory, and Ewa Deelman, Gaurang Mehta, and Karan Vahi, USC Information Sciences Institute

fMRI Applications and datasets were provided by Jack Van Horn and Jed Dobson, fMRI Data Center, Dartmouth College



Acknowledgements

GriPhyN and iVDG are supported by the
National Science Foundation



The Globus Alliance and PPDG are supported in part by the US Department of Energy, Office of Science; by the NASA Information Power Grid program; and by IBM

