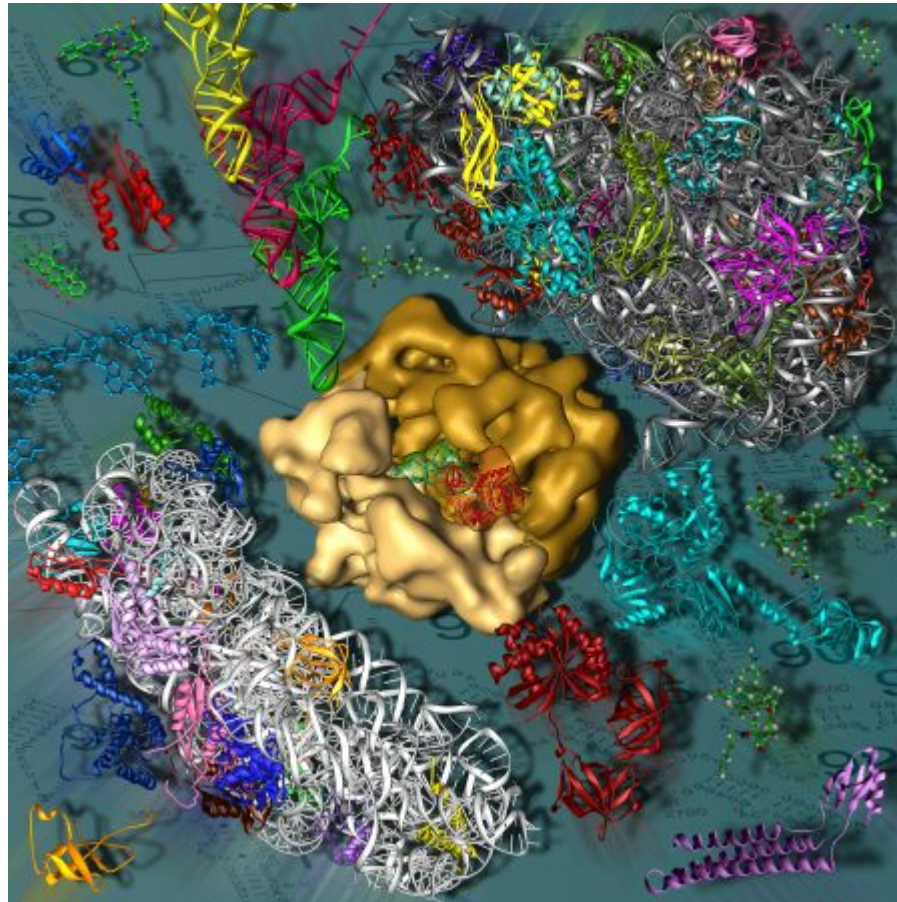


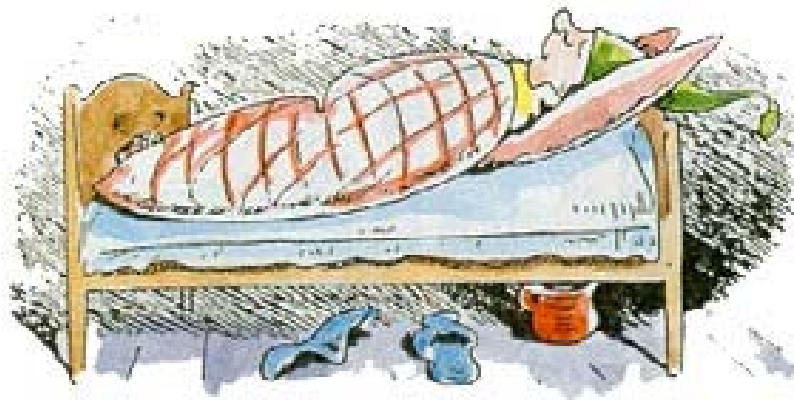
Protein Crystallography Needs.



ESRF UP WP11 Workshop
Exploiting the GRID for Synchrotron Radiation Data Analysis

Frank Schlünzen / DESY-IT

Protein Crystallography Needs.



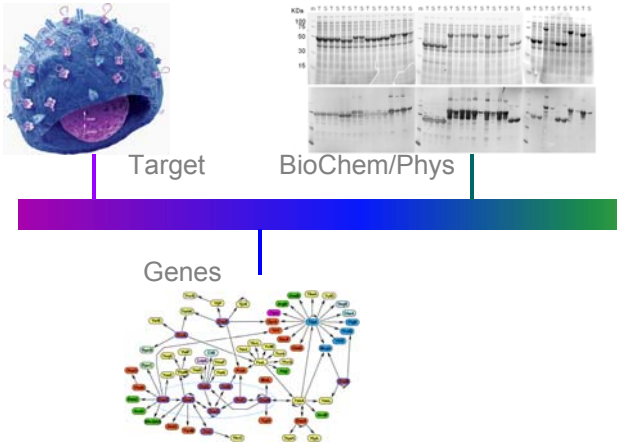
... the essentials

Outline.



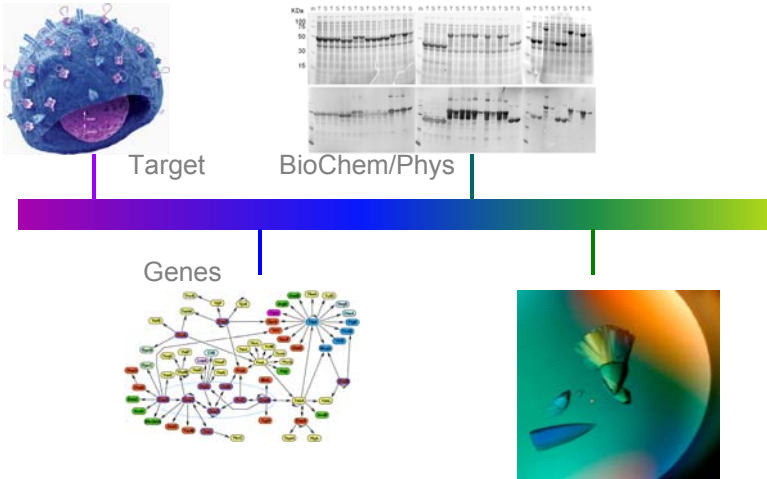
- Workflow (what's done)
- Resources (what's used)
- Data Rates / Access / Storage / Transfer
- Grid Tools
- Basic Needs

Workflow.



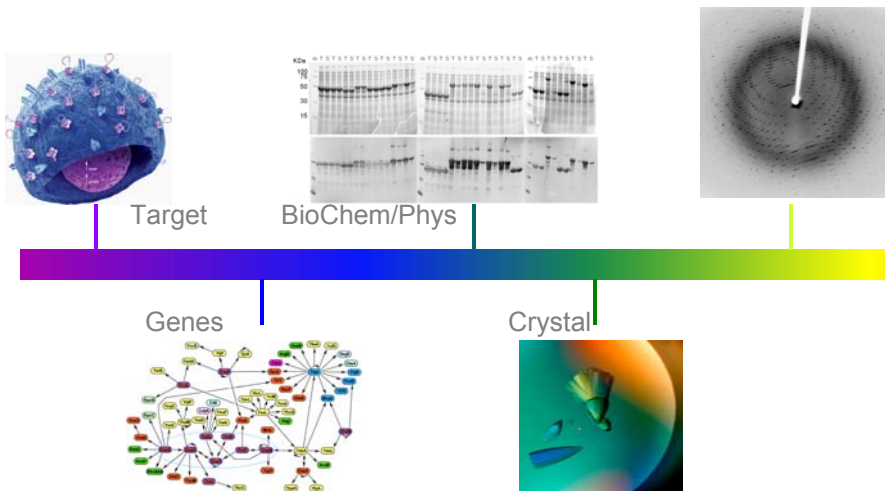
Preparation

Workflow.



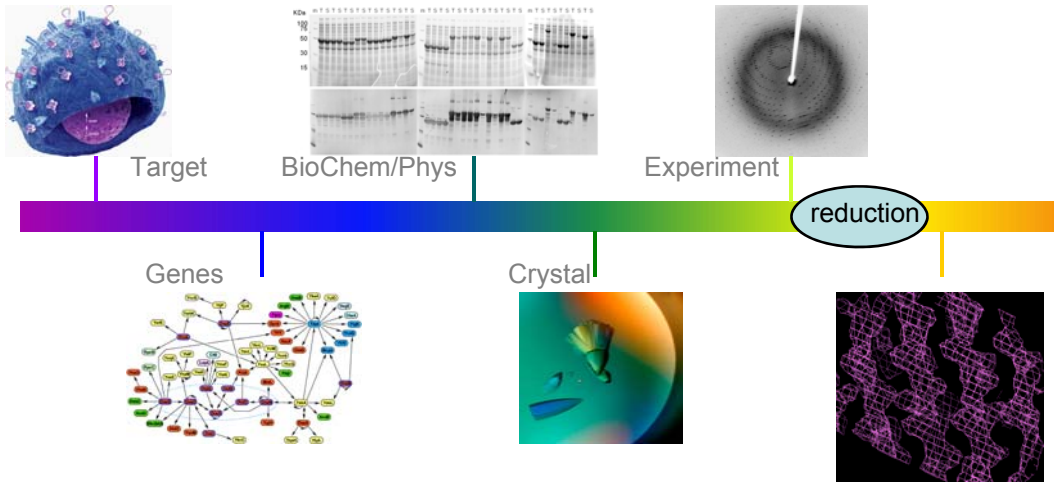
Crystallization

Workflow.



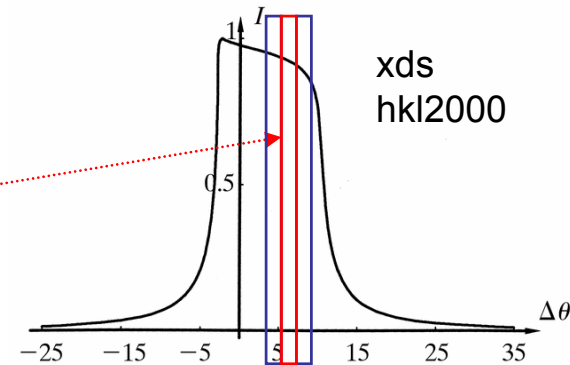
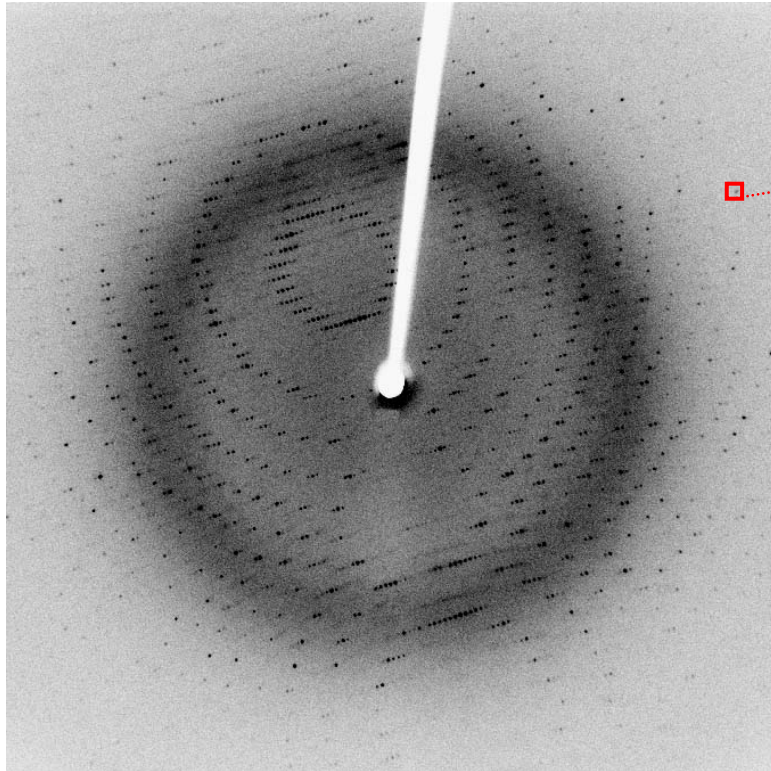
Experiment

Workflow.

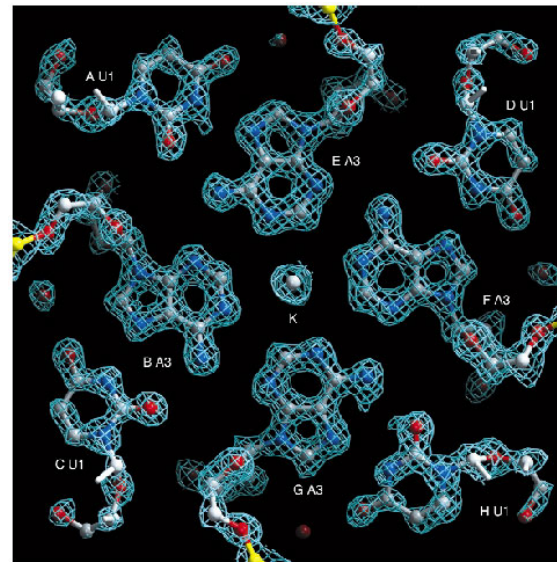
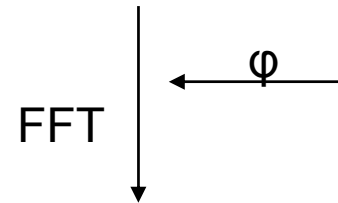


Analysis

Data reduction.

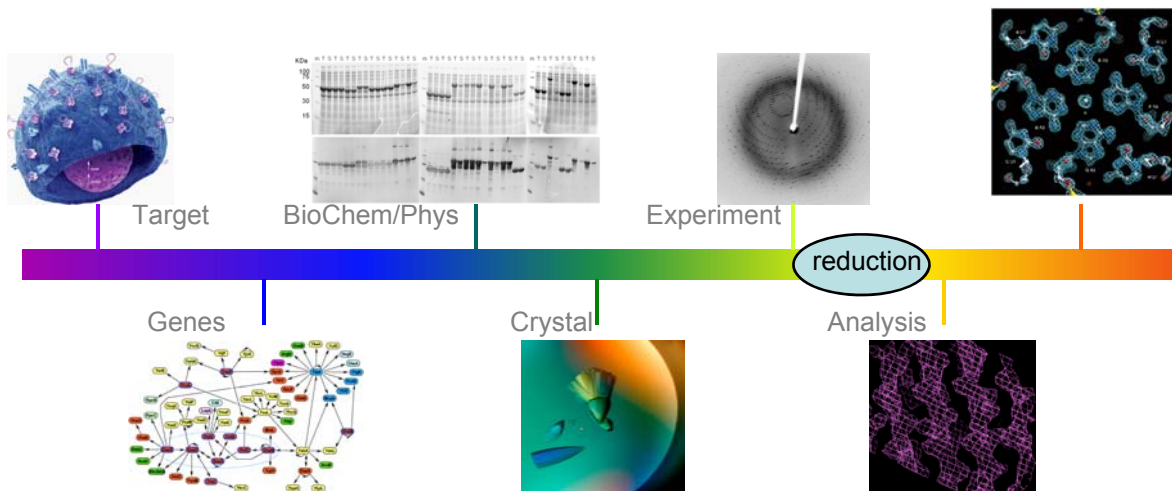


Refinement parameters
-orientation
-cell dimensions
-mosaicity
-decay
smoothly varying



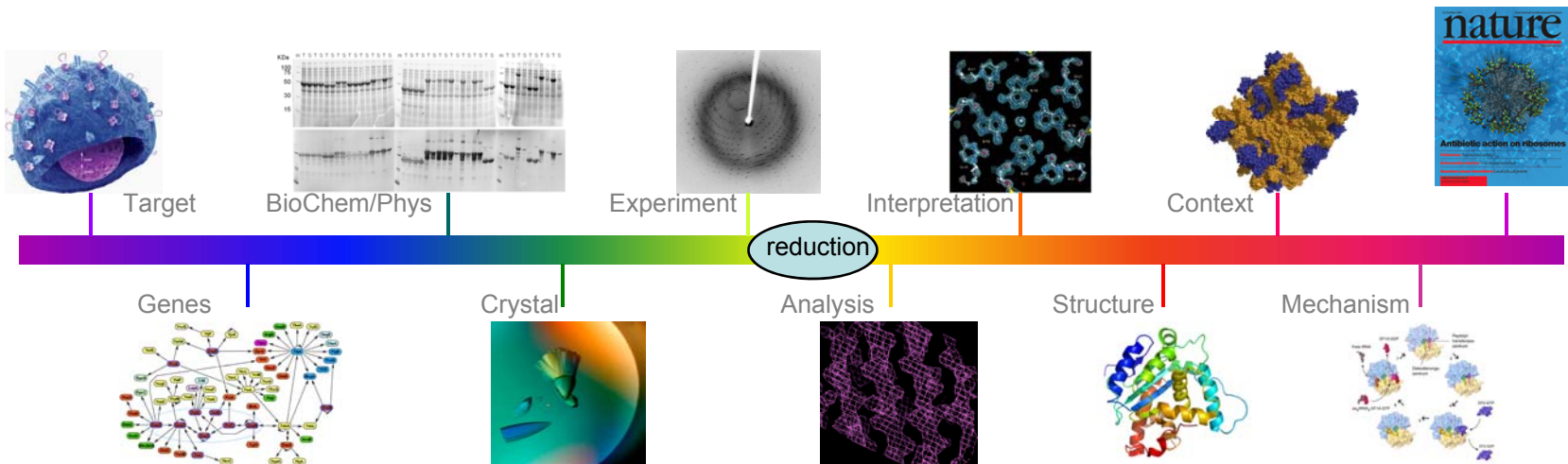
- dependencies between images
- multi-core - yes
- distributed/grid computing - no
- image corruption → ±200images

Workflow.



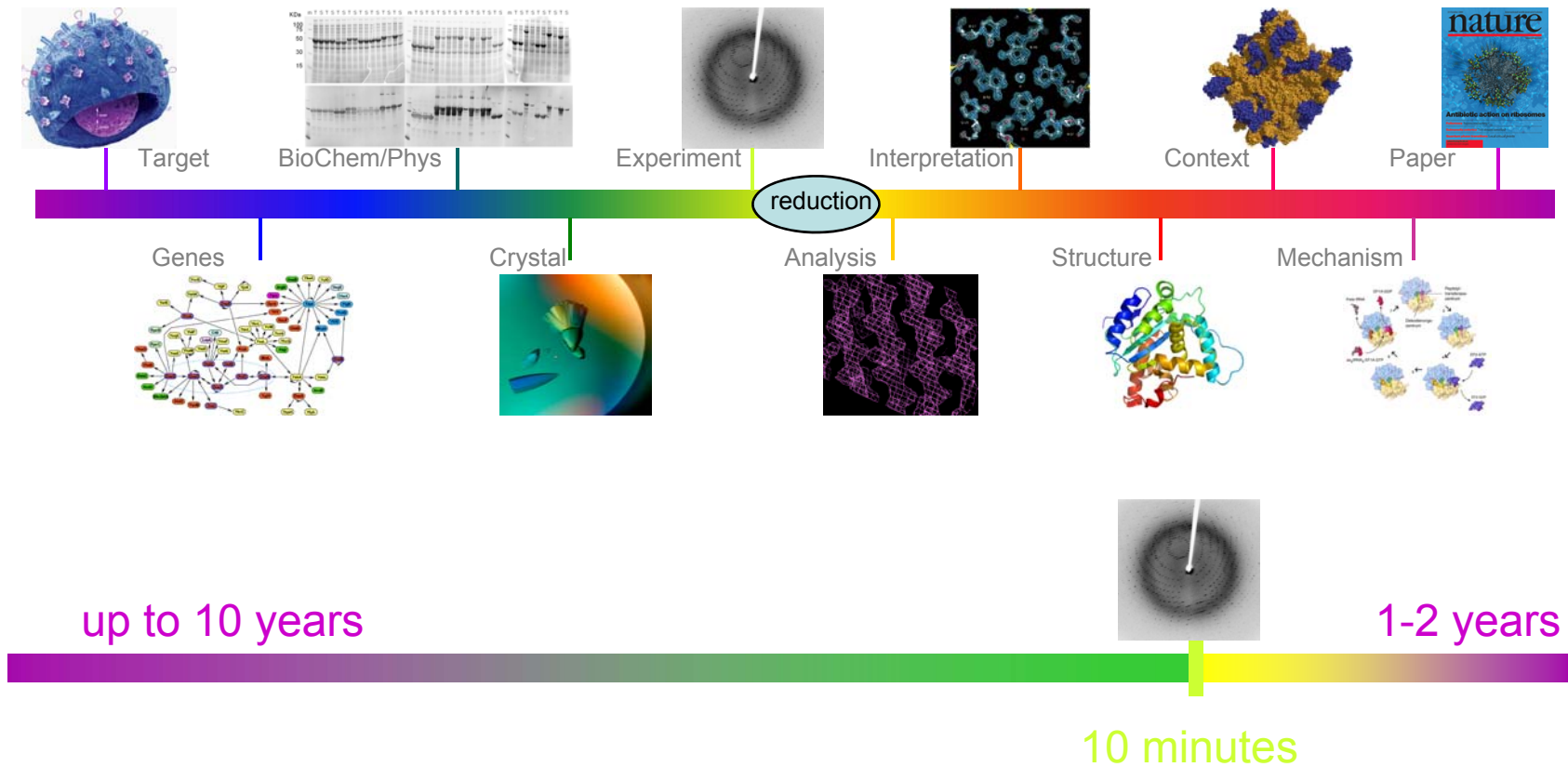
Refinement / Interpretation

Workflow.

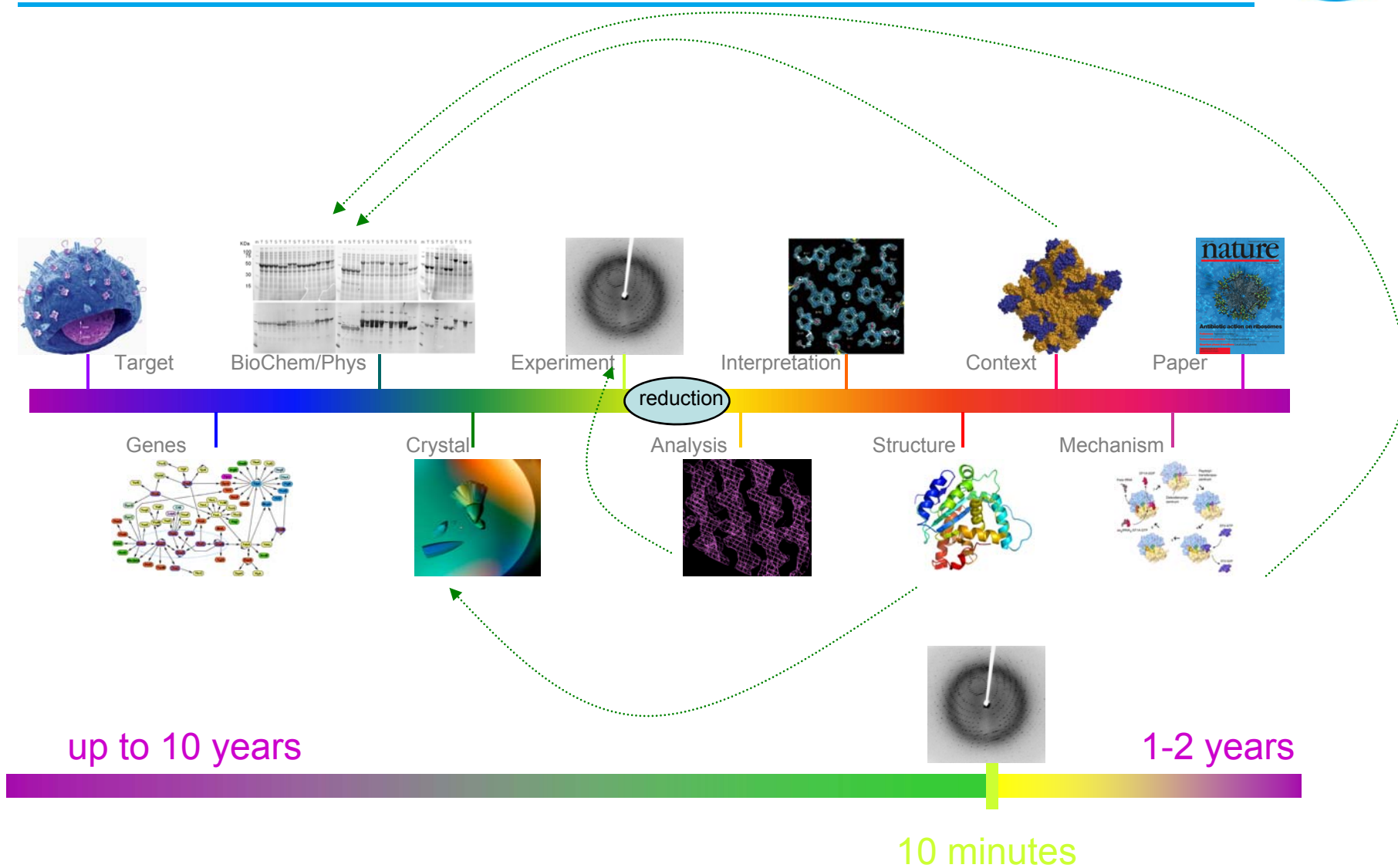


Publication

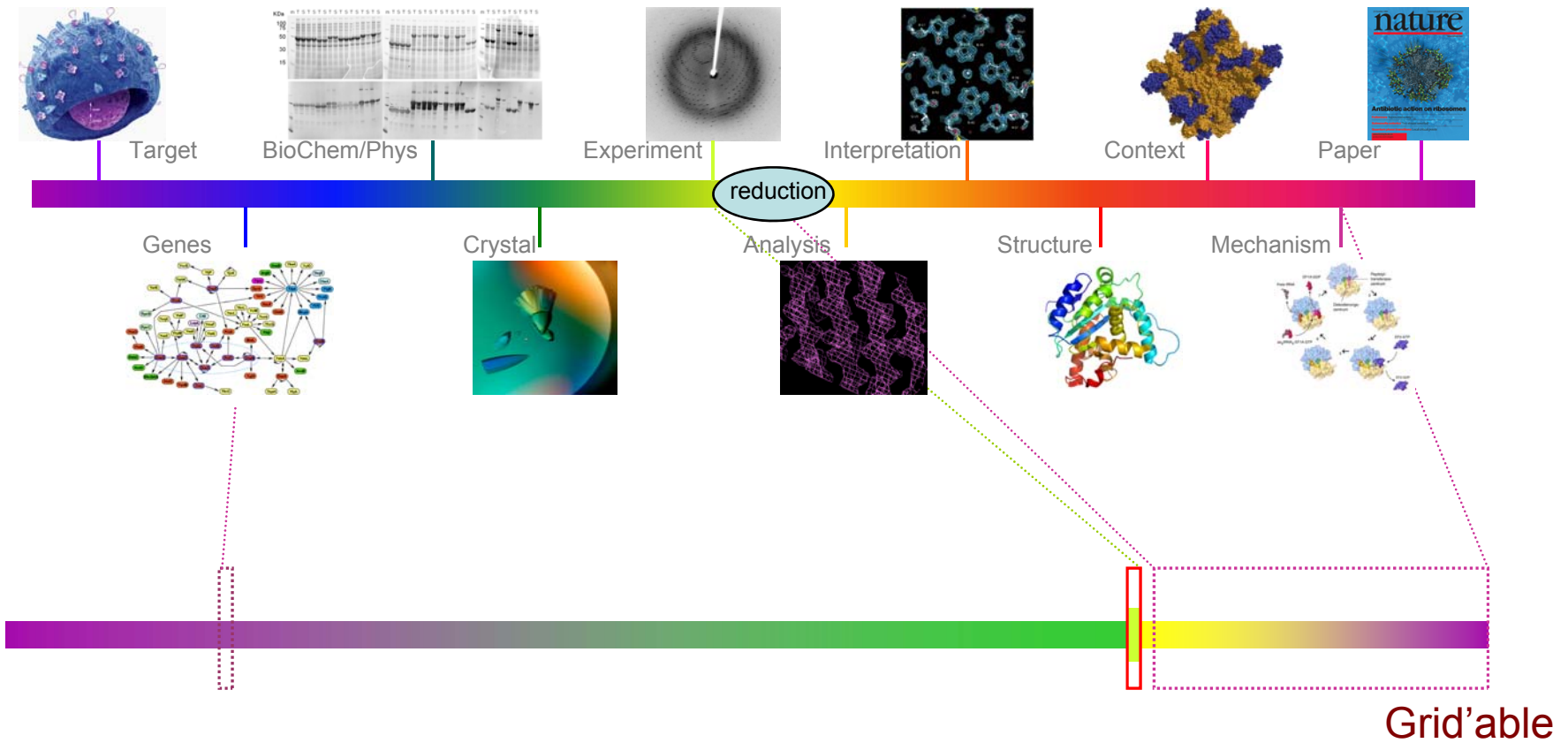
Timeframe.



Timeframe.



Timeframe.

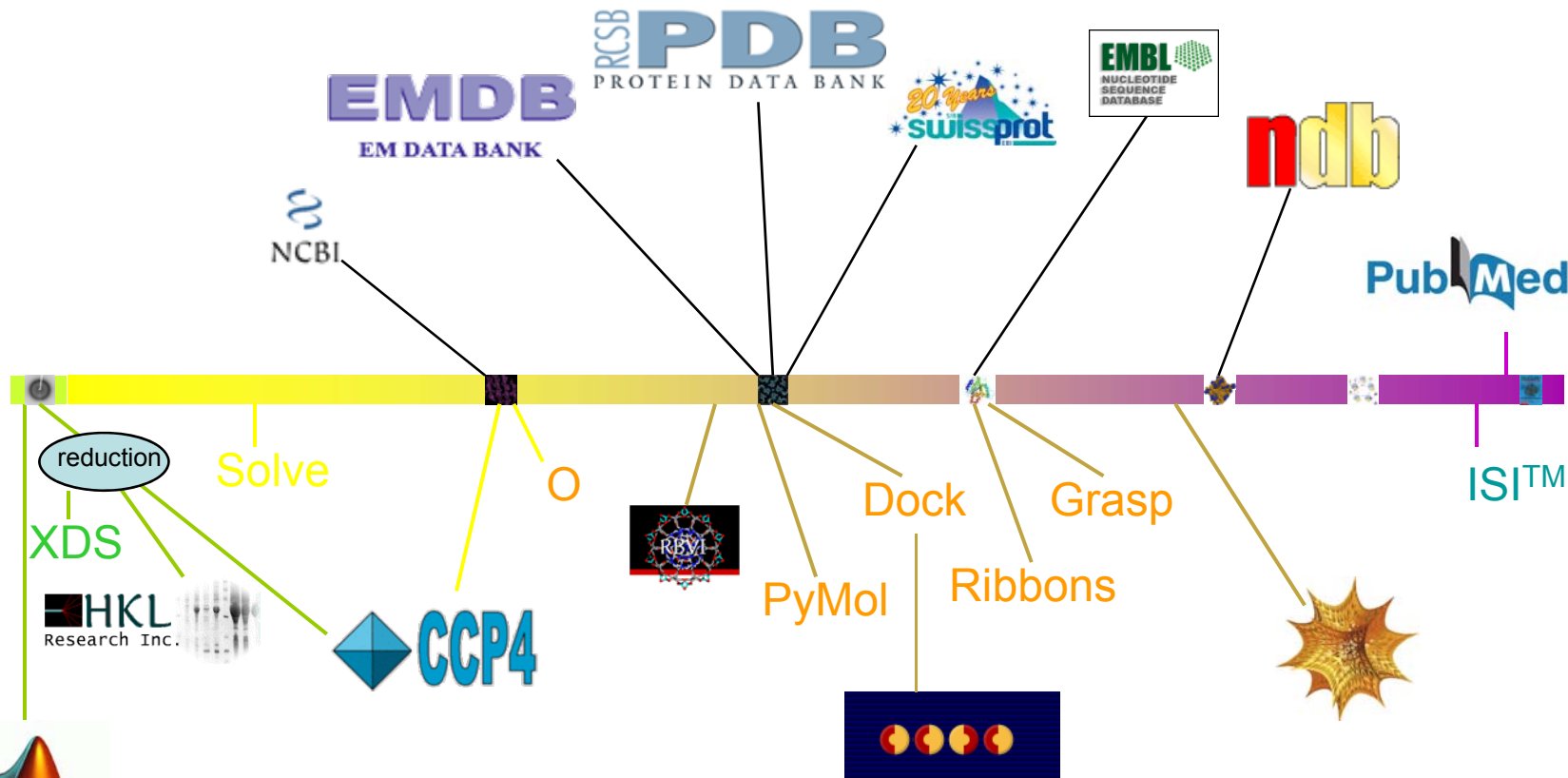


Resources.



Databases

Open Access



Software

not exactly open

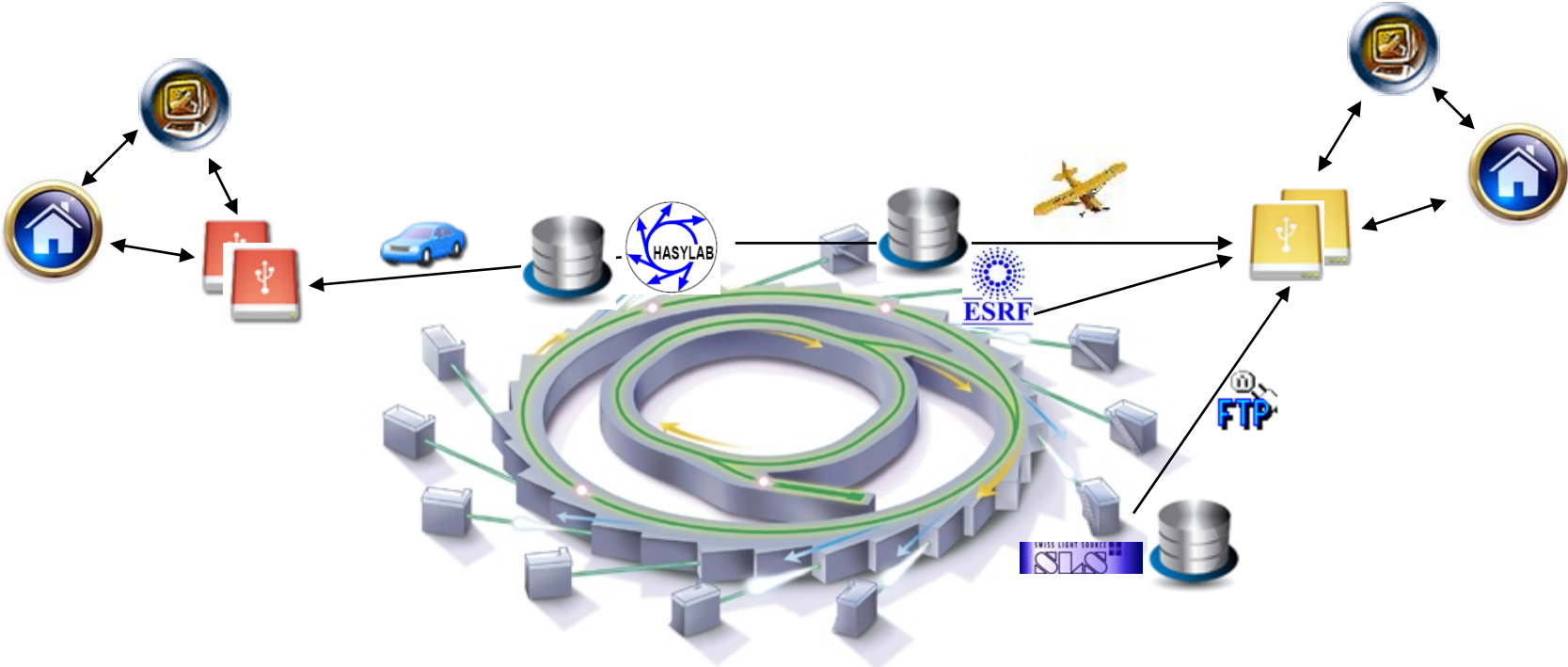
OS



Data Access.



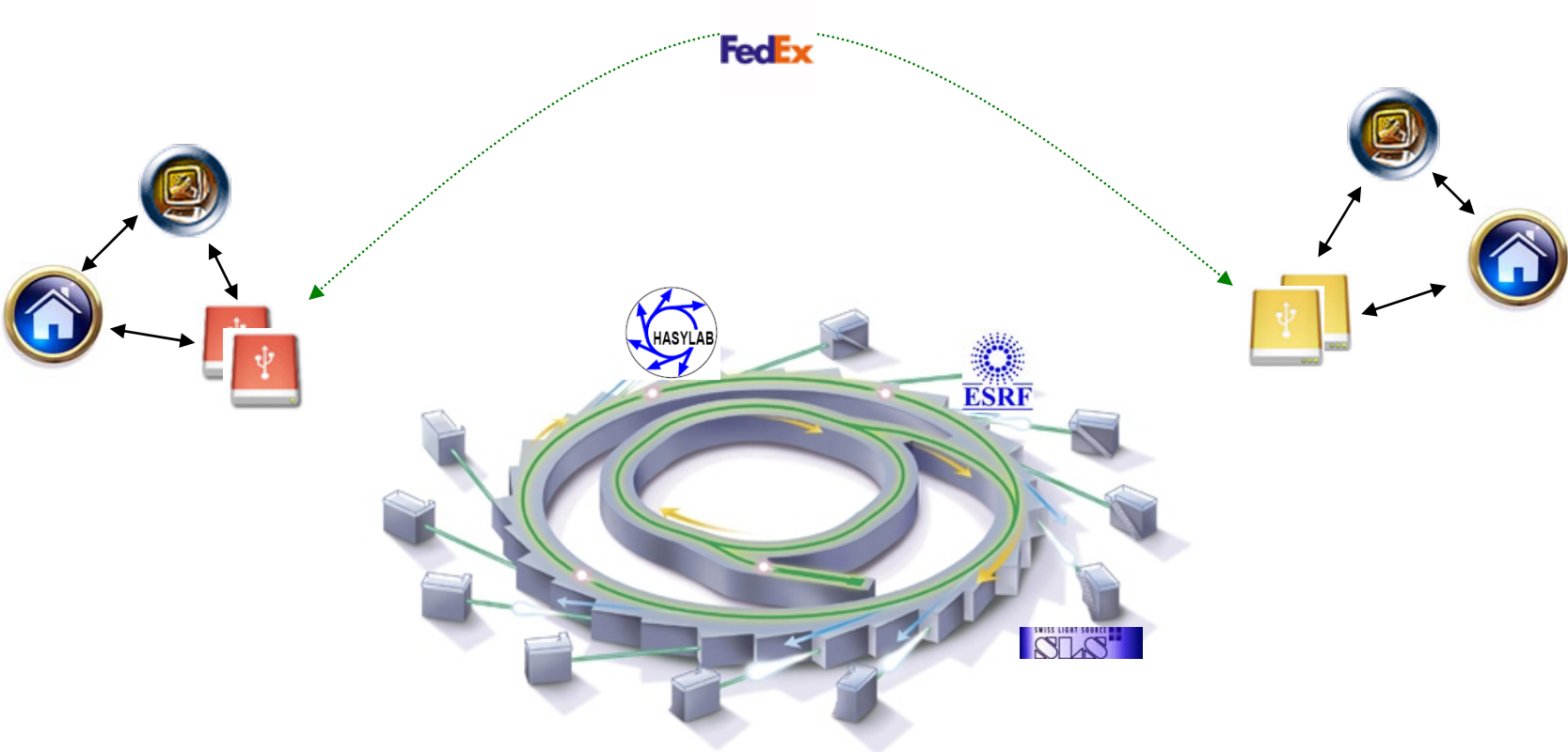
- Transfer data once
 - no validation ! no bookkeeping !



Data Access.



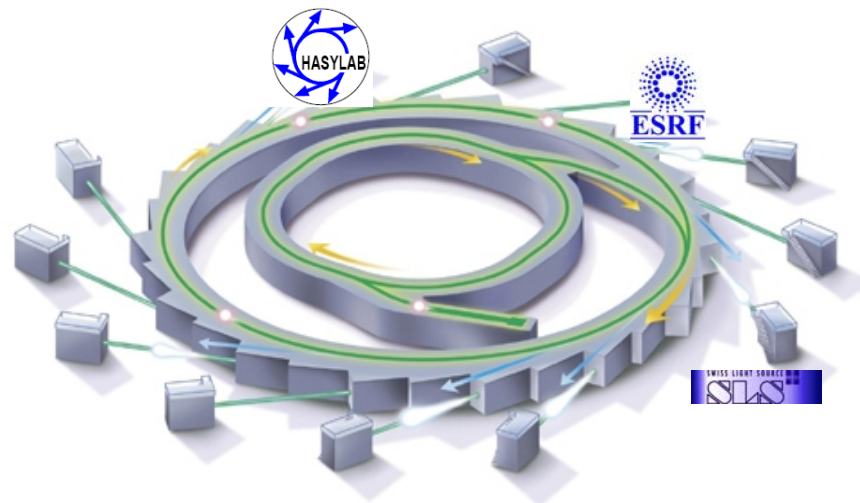
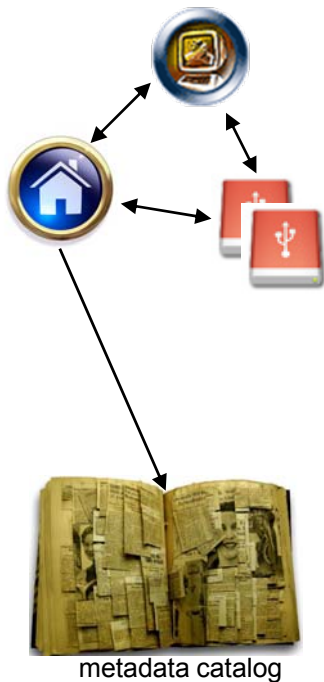
- Transfer data once
- Process data
- Don't touch it for a year or two
 - share data with colleagues abroad



Data Access.

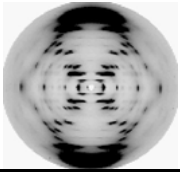

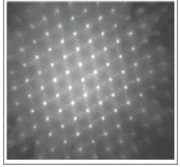

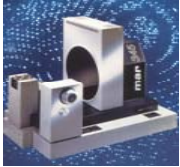



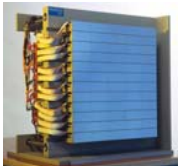


- Transfer data once
- Process data
- Don't touch it for a year or two
- Figure out that Interpretation is impossible
 - go back to raw data
 - raw data deleted and/or corrupted and/or non-locatable
- Doomed



Data Rates.



	BL	Detector	Exp. Time Readout	Time/D Frames/D	FrameVol DataVol	DataRate	Storage	Life/yr
'90	F1/CHESS X11/Doris		1h / 1h 2h / 1h	Days /100 Weeks/100	4 / 400MB	200MB/d 100MB/d		5-10
'93	F1/CHESS KEK		20' / 10' 30' / 60'	Days /100 Days /100	8 / 800MB 32 / 3.2GB	2GB/d 1GB/d		5-10
'96	BW6/Doris		20' / 3'	24h /200	16 / 1.6GB	5GB/d		4-6
'00	PX1/SLS ID29/ESRF SBC/APS		1' / 5"	1h /1000	16 / 16GB	500GB/d		0-3
'07	PX1/SLS Petra III XFEL		0.05" / 0 0.01" / 0 10 ⁻¹² s / 0	5' / 10000 2' / 10000 1" / 30000	16 / 160GB 16 / 160GB 16 / 500GB	30TB/d 60TB/d 2PB/d	3k€/d 6k€/d 200k€/d	

Long term archive at least doubles the costs!

Data Transfer.



Transfer a typical 18 GB dataset (995 images)

cp (usb)	40' = 0.6h
lcg-cp	567' = 9.4h
wget	363' = 6.0h
scp (afs)	33' = 0.6h

Processing a typical 18 GB dataset (XDS)

grid	$567' + 70' + 23' = 11.0h$
local	$40' + 70' + 0' = 1.9h$

Requirement

Transfer time \ll processing time

- Command line (glite)
- g-eglipse
- amga
- p-grade, ildg-browser

- many different approaches
- very difficult to identify suitable one
- ... or just to evaluate

Classical .

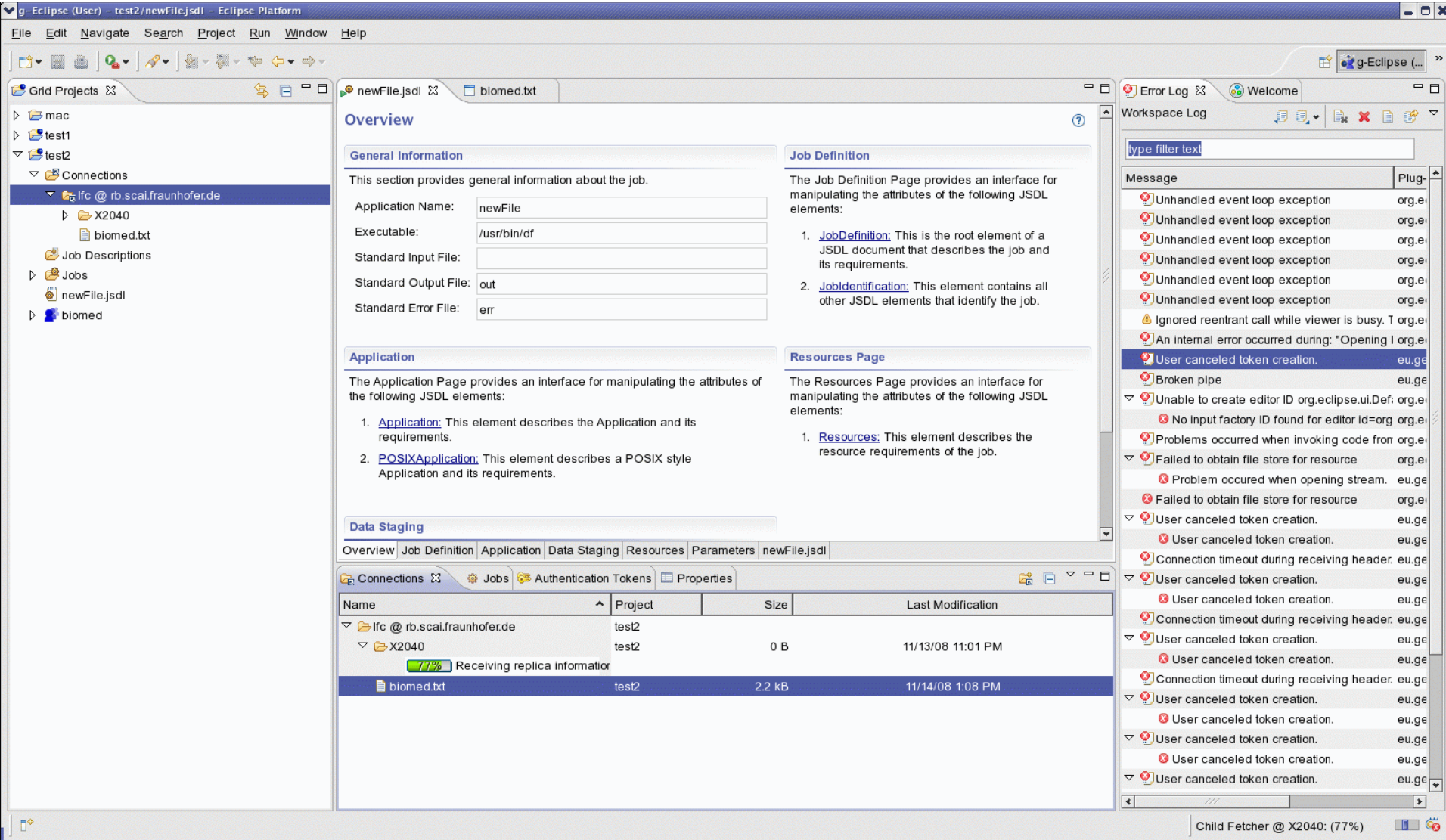


```
Executable = "myjob2.sh";
StdOutput  = "out";
StdError   = "err";
InputSandbox = {"myjob2.sh","XDS.INP"};
OutputSandbox = {"out","err","results.tgz"};
```

```
#!/bin/sh -x
# --- won't work without ...
export LFC_HOST=rb.scai.fraunhofer.de
# --- transfer software ( ~2GB)
lcp-cp -v --vo biomed lfn:/grid/biomed/schluenz/sw/XDS-linux_ifc_Intel+AMD.tar.gz file:$PWD/XDS.tar.gz
tar xvzf XDS.tar.gz
export PATH=$PWD/XDS-linux_ifc_Intel+AMD:$PATH
lcp-cp -v --vo biomed lfn:/grid/biomed/schluenz/sw/ccp4.tgz file:$PWD/ccp4.tgz
tar xvzf ccp4.tgz
# --- process data
mkdir DATA
i=5
while [ $i -lt 1000 ]; do
lcp-cp -v --vo biomed lfn:/grid/biomed/schluenz/X2040/x2040_30_${i}.img file:$PWD/DATA/x2040_30_${i}.img
i=$((i+1))
done
# --- run ccp4 ...
export CCP4BASE=$PWD
source $CCP4BASE/ccp4/6.0.2/setup-scripts/sh/ccp4.setup
source $CCP4BASE/ccp4/6.0.2/setup-scripts/sh/ccp4-others.setup

$CCP4BASE/ccp4/6.0.2/ccp4-6.0.2/bin/amore <<EOF
.... And so on ...
```

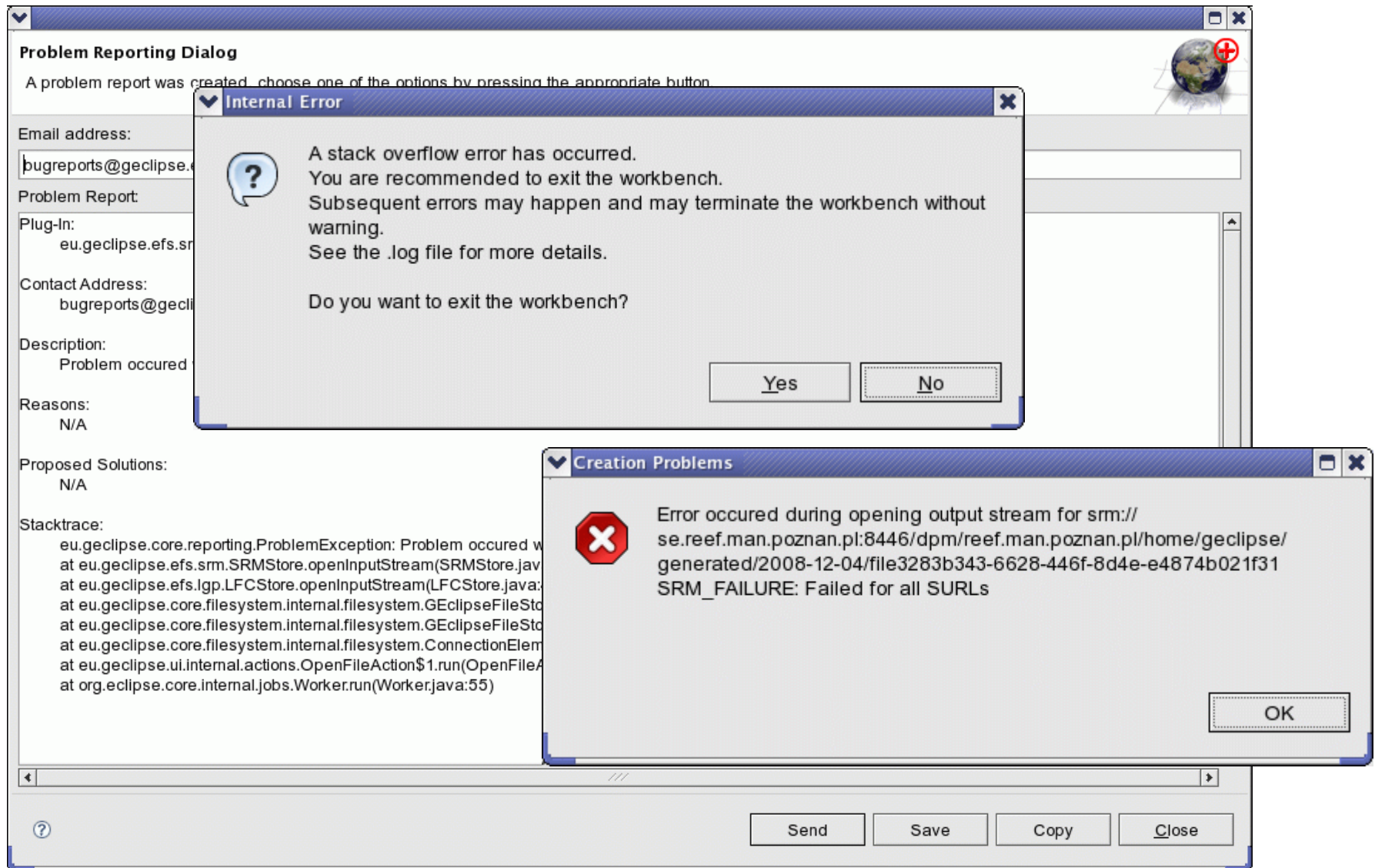
Basically works quite well (SL4)



The screenshot displays the G-eclipse IDE interface. The main window shows the 'Overview' page for a Job Definition named 'newFile.jsdl'. The 'General Information' section includes fields for Application Name (newFile), Executable (/usr/bin/df), Standard Input File, Standard Output File (out), and Standard Error File (err). The 'Job Definition' section explains the root element and lists 'JobDefinition' and 'JobIdentification'. The 'Application' section describes the Application and POSIXApplication elements. The 'Data Staging' section is currently empty. The 'Resources Page' section describes the resource requirements. The 'Connections' view at the bottom shows a table of connections:

Name	Project	Size	Last Modification
lfc @ rb.scai.fraunhofer.de	test2		
X2040	test2	0 B	11/13/08 11:01 PM
77% Receiving replica information			
biomed.txt	test2	2.2 kB	11/14/08 1:08 PM

The right-hand side of the interface shows the 'Error Log' and 'Workspace Log' panels. The 'Error Log' contains multiple entries for 'Unhandled event loop exception' and 'User canceled token creation'. The 'Workspace Log' shows a progress bar for 'Child Fetcher @ X2040: (77%)'.



The screenshot displays the G-eclipse software interface with two overlapping error dialog boxes. The background window is titled "Problem Reporting Dialog" and contains the following text:

A problem report was created. choose one of the options by pressing the appropriate button.

Email address:

Problem Report:

Plug-In: eu.geclipse.efs.sr

Contact Address: bugreports@gecli

Description: Problem occurred

Reasons: N/A

Proposed Solutions: N/A

Stacktrace:
eu.geclipse.core.reporting.ProblemException: Problem occurred w
at eu.geclipse.efs.srm.SRMStore.openInputStream(SRMStore.jav
at eu.geclipse.efs.lgp.LFCStore.openInputStream(LFCStore.java;
at eu.geclipse.core.filesystem.internal.filesystem.GEclipseFileSto
at eu.geclipse.core.filesystem.internal.filesystem.GEclipseFileSto
at eu.geclipse.core.filesystem.internal.filesystem.ConnectionElem
at eu.geclipse.ui.internal.actions.OpenFileAction\$1.run(OpenFileA
at org.eclipse.core.internal.jobs.Worker.run(Worker.java:55)

The "Internal Error" dialog box is overlaid on top of the "Problem Reporting Dialog" and contains the following text:

Internal Error

A stack overflow error has occurred.
You are recommended to exit the workbench.
Subsequent errors may happen and may terminate the workbench without warning.
See the .log file for more details.

Do you want to exit the workbench?

Buttons: Yes, No

The "Creation Problems" dialog box is overlaid on top of the "Internal Error" dialog box and contains the following text:

Creation Problems

Error occured during opening output stream for srm://
se.reef.man.poznan.pl:8446/dpm/reef.man.poznan.pl/home/geclipse/
generated/2008-12-04/file3283b343-6628-446f-8d4e-e4874b021f31
SRM_FAILURE: Failed for all SURLs

Button: OK

At the bottom of the main window, there are buttons for "Send", "Save", "Copy", and "Close".



- Great tool for data management & job handling
- Platform independent

- Difficult to setup (for a dumb user)
- Not very stable

- **Metadata catalog**

- command line tool
- web interface
- various API's
 - python, java, php

```
Query> addentry /grid/xray/ribo/species1 ID 1 Class 'Thermophilic' Organism 'Th. thermophilus'
Query> addentry /grid/xray/ribo/species2 ID 2 Class 'Thermophilic' Organism 'B. stearoht.'
Query> addentry /grid/xray/ribo/species3 ID 3 Class 'Archea' Organism 'H. marismortui'
Query> listentries /grid/xray/ribo
    >> /grid/xray/ribo/species1
    >> /grid/xray/ribo/species2
    >> /grid/xray/ribo/species3
Query> ls
    >> species1
    >> species2
    >> species3
```

- Command line works well
- Hardly acceptable for normal user
 - rather use sql-wizards ...



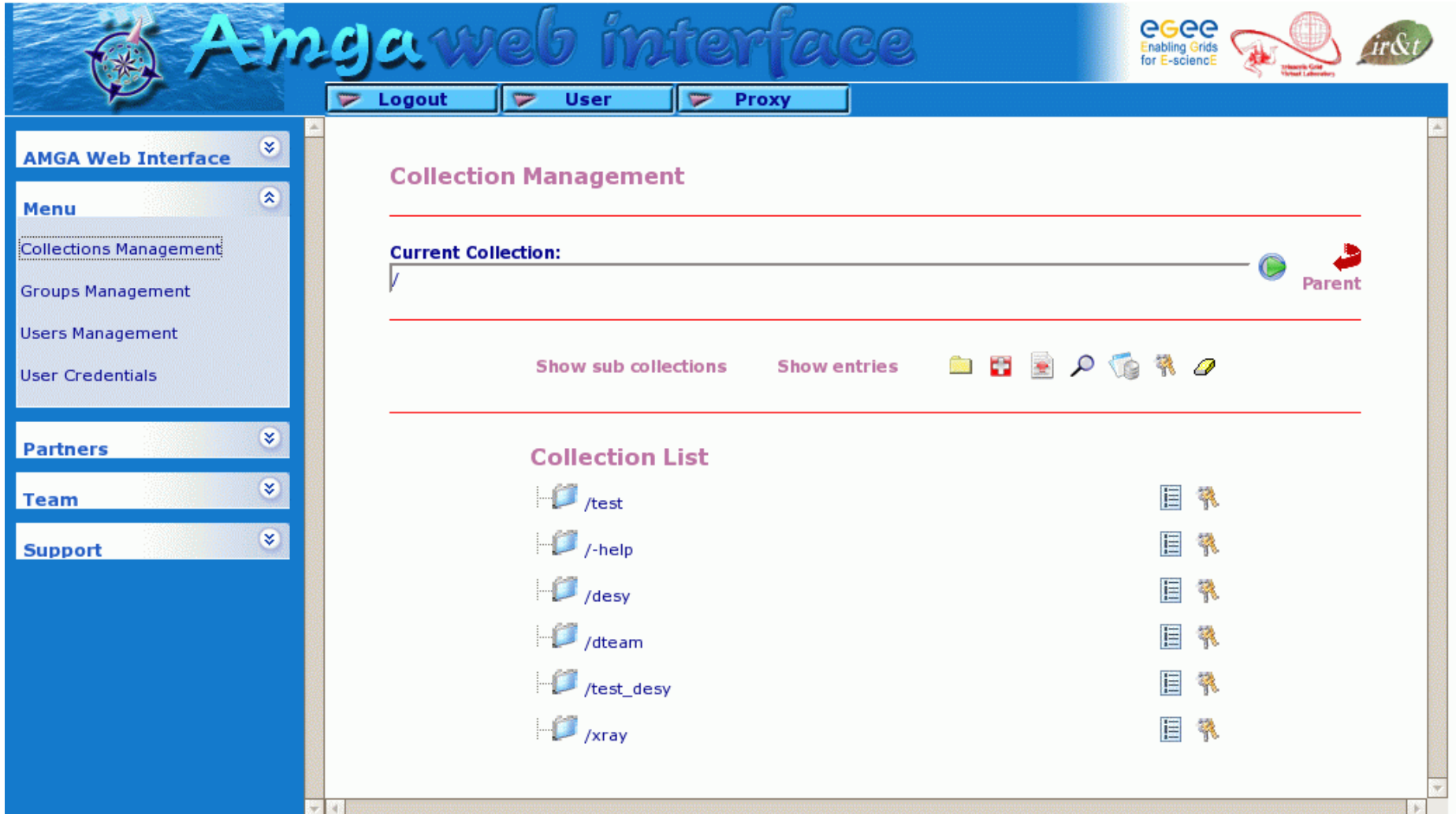
Provide your user ID to log on AMGA Server

User ID:

/afs/desy.de/

Browse...

Upload



AMGA web interface

Logout User Proxy

AMGA Web Interface

Menu


- Collections Management
- Groups Management
- Users Management
- User Credentials








Partners

Team













Support

Collection Management

Current Collection:  **Parent**

Show sub collections Show entries       

Collection List

/test	 
/-help	 
/desy	 
/dteam	 
/test_desy	 
/xray	 

Looks promising ...

HTTP Status 500 -

type Exception report

message

description The server encountered an internal error () that prevented it from fulfilling this request.

exception

```
java.lang.NullPointerException
    com.irt.amgawi.servlet.LoginService.doPost(LoginService.java:99)
    com.irt.amgawi.servlet.LoginService.doGet(LoginService.java:30)
    javax.servlet.http.HttpServlet.service(HttpServlet.java:627)
    javax.servlet.http.HttpServlet.service(HttpServlet.java:729)
```

note The full stack trace of the root cause is available in the Apache Tomcat/5.5.27 logs.

Apache Tomcat/5.5.27

... but not with my proxy ...

```
java amgajava
Could not read logging configuration: logging.properties (No such file or directory)
Proceeding...
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnectionContext <init>
INFO: Loading configuration: /afs/desy.de/user/s/schluenz/.mdclient.config
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnectionContext <init>
INFO: useSSL: true
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnectionContext <init>
INFO: certFile: null
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnectionContext <init>
INFO: keyFile: null
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnection connect
INFO: Connecting to grid-amga0.desy.de:8822
Dec 5, 2008 12:19:02 PM arda.md.javaclient.MDServerConnection initSSLSocketFactory
INFO: Using grid proxy authentication
Dec 5, 2008 12:19:03 PM arda.md.javaclient.MDServerConnection initSSLSocketFactory
INFO: Enabling SSL encryption
Exception in thread "main" javax.net.ssl.SSLHandshakeException: Received fatal alert: unknown_ca
    at com.sun.net.ssl.internal.ssl.Alerts.getSSLException(Alerts.java:150)
    at com.sun.net.ssl.internal.ssl.Alerts.getSSLException(Alerts.java:117)
    at com.sun.net.ssl.internal.ssl.SSLSocketImpl.recvAlert(SSLSocketImpl.java:1650)
    at com.sun.net.ssl.internal.ssl.SSLSocketImpl.readRecord(SSLSocketImpl.java:925)
    at com.sun.net.ssl.internal.ssl.SSLSocketImpl.waitForClose(SSLSocketImpl.java:1428)
    at com.sun.net.ssl.internal.ssl.HandshakeOutputStream.flush(HandshakeOutputStream.java:103)
    at com.sun.net.ssl.internal.ssl.Handshaker.sendChangeCipherSpec(Handshaker.java:591)
    at com.sun.net.ssl.internal.ssl.ClientHandshaker.sendChangeCipherAndFinish(ClientHandshaker.java:698)
    at com.sun.net.ssl.internal.ssl.ClientHandshaker.serverHelloDone(ClientHandshaker.java:624)
    at com.sun.net.ssl.internal.ssl.ClientHandshaker.processMessage(ClientHandshaker.java:160)
    at com.sun.net.ssl.internal.ssl.Handshaker.processLoop(Handshaker.java:495)
    at com.sun.net.ssl.internal.ssl.SSLSocketImpl.startHandshake(SSLSocketImpl.java:1116)
    at com.sun.net.ssl.internal.ssl.SSLSocketImpl.startHandshake(SSLSocketImpl.java:1100)
    at arda.md.javaclient.MDServerConnection.establishSSLConnection(MDServerConnection.java:329)
    at arda.md.javaclient.MDServerConnection.connect(MDServerConnection.java:271)
    at arda.md.javaclient.MDServerConnection.<init>(MDServerConnection.java:195)
    at amgajava.main(amgajava.java:10)
```

java-api quite tedious to get it working

- Nice metadata tool
- Platform independent
- Easy to setup

- Troubles with web interface
- Troubles with java api
 - Any real applications implementing java API ?
- API's a year behind client/server development (python 2.2)
- ignores X509 envs. proxy in /tmp/ ?

Crystallographers.



... tend to be impatient

... are extremely well connected.

→ give one them a problematic tool now, many won't touch it again.

(rumours spread fast)

Basic Needs.



- **Archive**
 - Long term, secure, reliable, private
 - Fast access (even if it's just a 'last resort')
 - **easy to use** (~explorer)
- **Grid tools**
 - stable
 - platform independent
 - **easy to use** (~drag'n drop)
- **Metadata catalogue**
 - tightly integrated
 - High throughput capable
 - **easy to use** (~google)
- **Online Computing**
 - Job debugging
- **Software**
 - central repository (nfs ? afs ?)
 - (xds, ccp4, solve, cns, (auto)dock, sharp,buster, arp/warp, ...)

