

XSIL

Extensible Scientific Interchange Language

<http://www.cacr.caltech.edu/XSIL>

Roy Williams

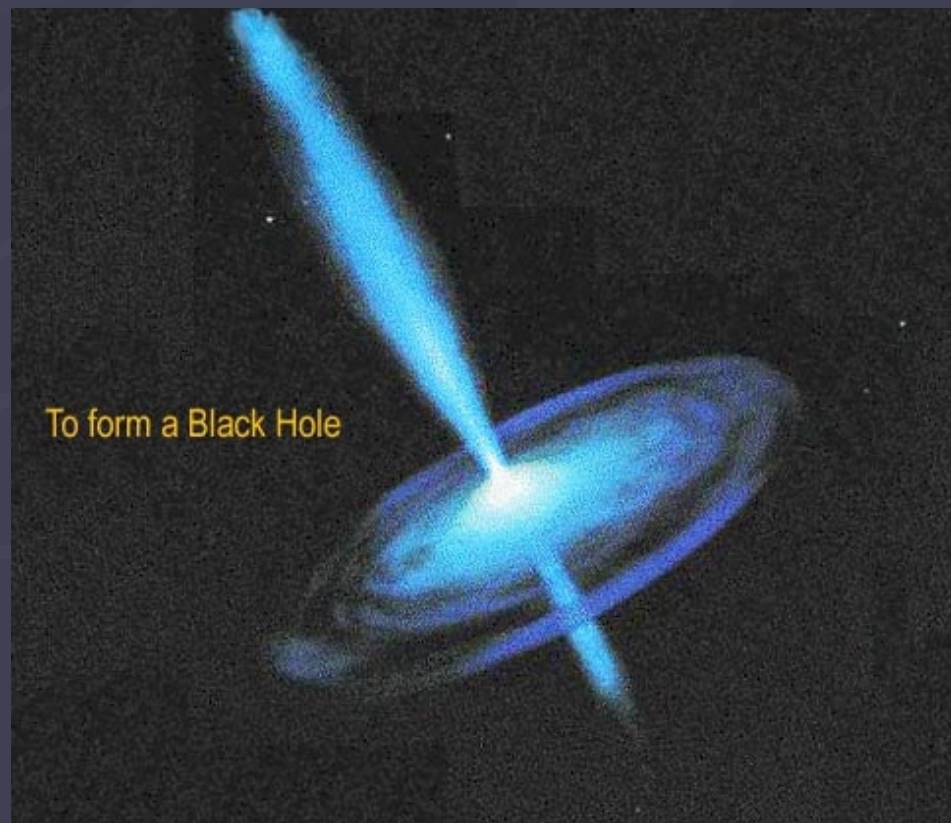
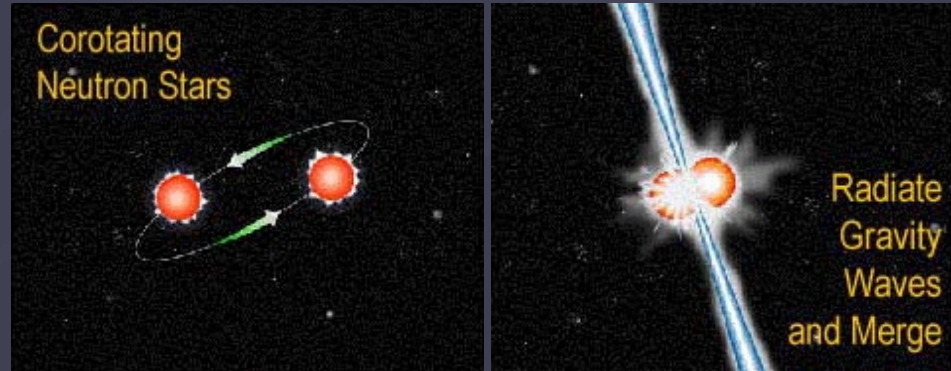
California Institute of Technology

- Notebook for Scientific Data and Metadata
 - Data and metadata exchange
 - Creating and editing metadata
 - Catalogue applications
 - Electronic Log applications

LIGO

Laser
Interferometric
Gravitational wave
Observatory

Listening to
Collisions of
Black Holes and
Neutron Stars



LIGO

Detection of cosmic
gravitational waves



LIGO Data Objects

Multichannel time-series

- Interferometer
- Seismometer
- Microphone
- Magnetometer
- Feedback systems ...

Frequency series

- Power spectra
- Noise spectra ...

Events

- Astrophysics, Instrument glitch, ...

Images

- beam shape

Tables

Parameters

What is that Binary Data?

```
lugh%  
lugh% ls -l *.bin  
-rw-r--r--  1 roy      staff    1049376 Oct  3 16:40 fold32127+11A.bin  
lugh%  
lugh% more fold32127+11A.bin  
?èóK-ã>p5--ìu¿Z9o?+ ?R{ÿ8?PóylÝ^L?ÝI$T;¿IH×BZ¿!q-~¿Z`>á%óÃ*Ü?  ^~ó%ó  
lugh%
```

Question: What does this binary file mean?

Answer: Look at the metadata!

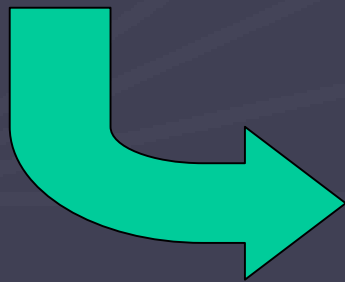
Where is it?: In the XSIL file!

Tools for Scientific Data Archives

Leverage E-Commerce!

Search for this book
In my shopping basket
My authentication
Download
Giftwrap
Make it in size 10, with the pink edging

Web & Java & XML
Servlets & EJB
E-services & E-speak
Software components
Application servers
Security
Databases



Login
Find events like this
Cache from tape to disk
As a Mathematica workbook
Mine this Tbyte with this parametrized search

XML: Extensible Markup Language



Oracle plans XML support

Built-in database support will integrate
DOM on client side

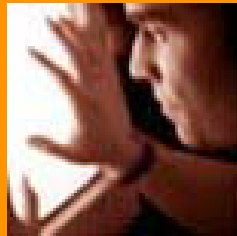
INTRODUCTION AdMarkup Whitepaper

The New York Times advertisement markup prototype



News & Events: Headlines

IBM, e-business, & XML



extreme xml

a monthly column on using Extensible Markup
Language

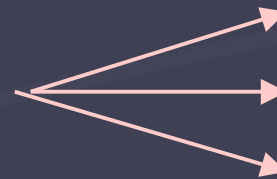
sbnmagaz

XML: “personal” HTML

Separation of structure from presentation

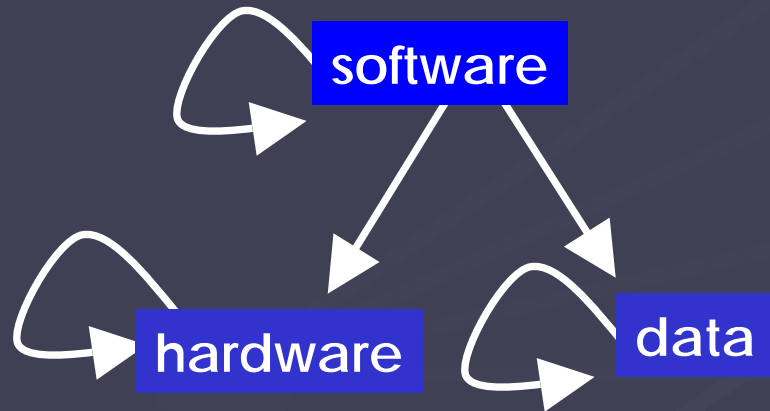


```
<From>Antonio Stadivarius</From>
<To>Domenico Scarlatti</To>
<Date>
  <Day>13</Day>
  <Month>4</Month>
  <Year>1723</Year>
</Date>
<Body>
  lo bisogno una appartamento
  acoglienti a Cremona ...
</Body>
```



13/4/23
April 13, 1723
17.iv.1723

Why XML



The cost is not the hardware
The cost is not the software
The cost is the DEPENDENCIES
The solution is STANDARD INTERFACES

XML is a standard way to represent structured documents,
including metadata and data

- Platform neutral / Open
- Vendor supported / Vendor neutral
- Proven -- decades with SGML
- Extensible
- Syntax checking -- Explicit Schema
- Embraces Java -- the software standard

XML for science

The library or the simulation should output
Binary objects glued together with
Structured documents

XML can provide the syntax for this structure
Like HTML but extensible
for metadata, programs, formatting
for searching, archiving, summarizing

-- now we need to define Document Types
(what are the fields on the catalogue card?)

XSIL

Extensible Scientific Interchange Language

<http://www.cacr.caltech.edu/XSIL/>

Base Objects

Param, Table, Array, Comment, Event, ...

Hierarchy

Streams

Local/Remote, Mirrored
Typed, Mime, Raw

Extensibility

"How can I represent MyObject"

XSIL Example

Pulsar Search in LIGO Data

```
<?xml version="1.0"?>
<!DOCTYPE XSIL SYSTEM "XSIL.dtd">

<XSIL Name="fold32127+11A">
  <Comment>Creator: Stuart Anderson and Roy Williams</Comment>
  <Comment>This time series has been folded at freqfold</Comment>
  <Param Name="Dataset">fold32127+11A</Param>

  <!-- REQUIRED DIMENSIONS AND VARIABLES -->
  <Param Name="n_elem" Type="int">100 </Param>
  <Param Name="n_chan" Type="int">1 </Param>

  <!-- VARIABLES AND ATTRIBUTES: -->
  <Param Name="ligo_type">time series </Param>
  <Param Name="phase" Type="double">21773750.252132</Param>
  <Param Name="tphase" Type="double">0.252132 </Param>
  <Param Name="sample_freq" Type="double">2000.000000 </Param>
```

[continued...]

XSIL Example

Pulsar Search in LIGO Data

[...continued]

```
<!-- HERE IS THE LINK TO THE DATA -->  
  <Stream Type="Remote">  
    file: //fold32127+11A.bin  
  </Stream>  
  
  <Array Name="data" Type="double">  
    <Dim Name="x">100</Dim>  
    <Dim Name="y">20</Dim>  
  </Array>  
  
  <Array Name="spectrum" Type="double">  
    <Dim>131072</Dim>  
  </Array>  
</XSIL>
```

XSIL Base Objects

<XSIL Name="Apple" Type="Fruit">

The container, which may contain other containers

<Param Name=... Unit=...>

<Time Type=...>

<Array Name=... Type=...>

<Dim>...

<Table Name=...>

<Column Name=... Unit=... Type=...>

<URL>

<Stream>

Local/Remote

Typed/Raw/MIME

Encoding

Actuation

XSIL Streams

```
<Stream Type="Local" Delimiter=",">  
  1.28374, 1.23453, 1.94847, 2.148474  
  2.39484, 2.84746, 3.10928, 4.92827  
</Stream>
```


Data is Explicit

```
<Stream Type="Remote" Encoding="Bigendian" Timeout="600"/>  
  http://hpss.cacr.caltech.edu/magval_09_25_97.bin  
</Stream>
```

Data is External

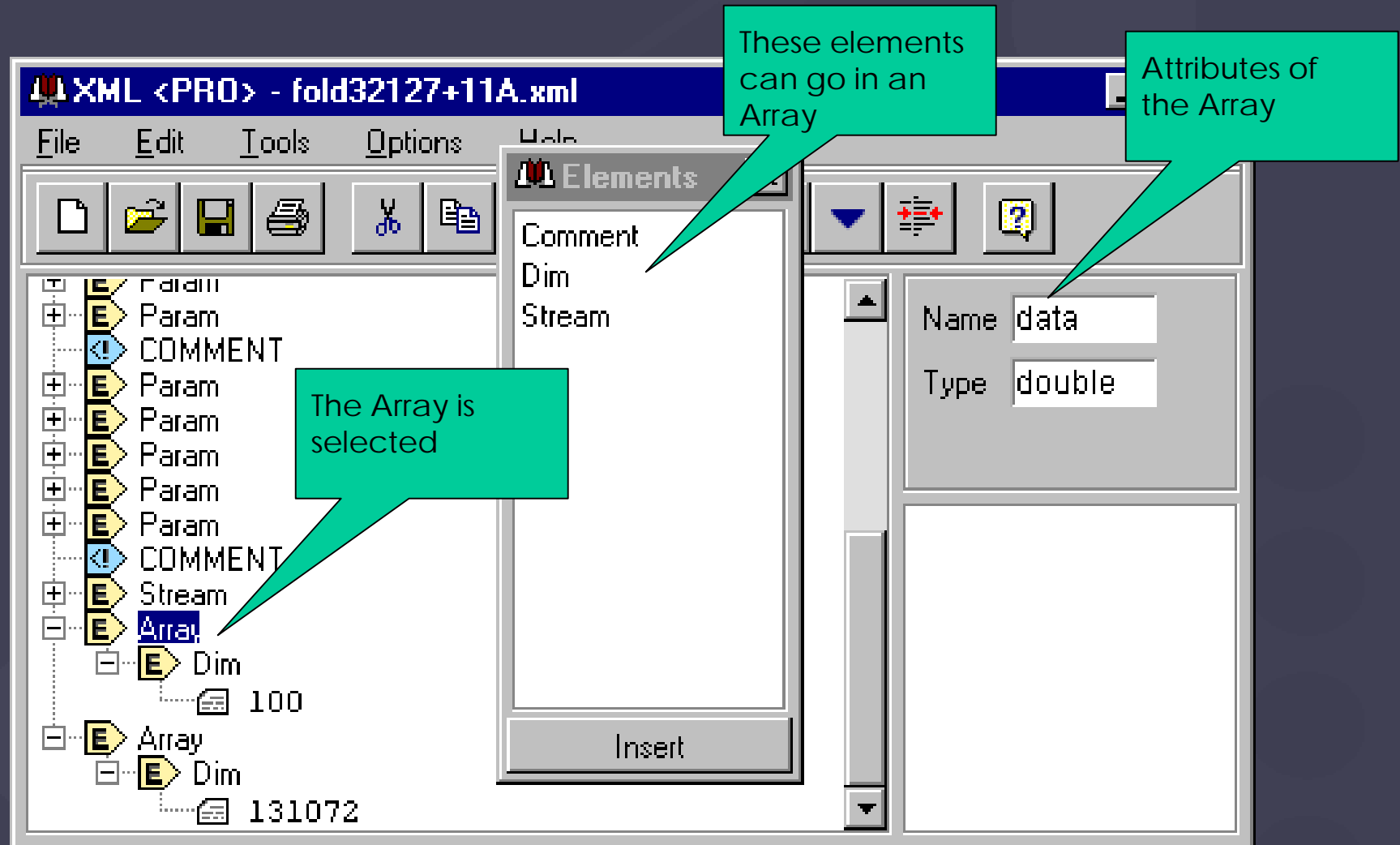
Extending XSIL

Look for Java class of this name
Expected internal structure
Helper application
Queries, Filters

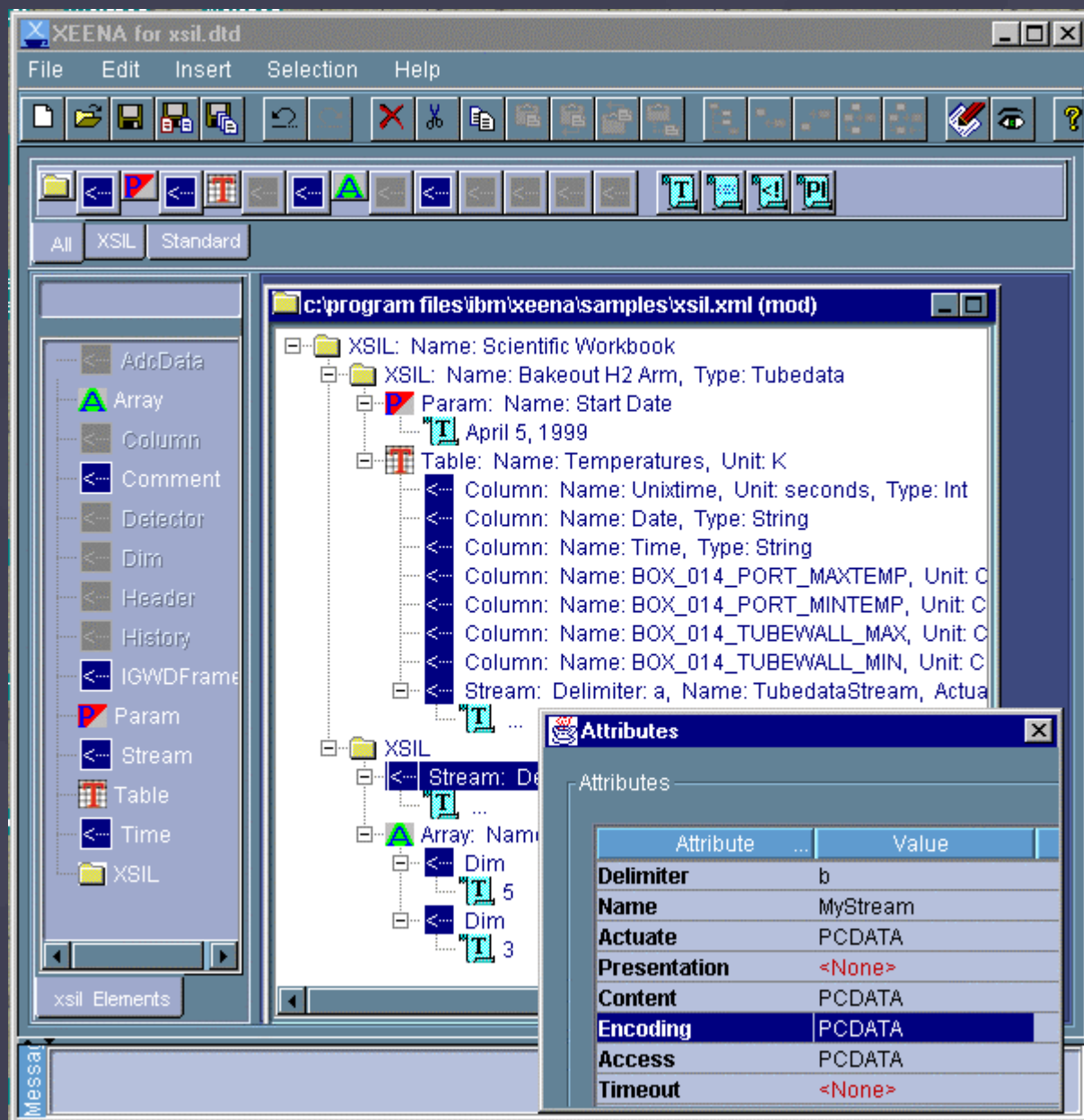


```
<XSIL Type="TimeSeries">  
  <Param Name="StartTime">943873726.92837</Param>  
  <Param Name="EndTime">943873876.83736</Param>  
  <Array Name="Data">  
    <Dim>1024</Dim>  
  </Array>  
  <Stream Type="Remote">  
    file://mydata.dat  
  </Stream>  
</XSIL>
```

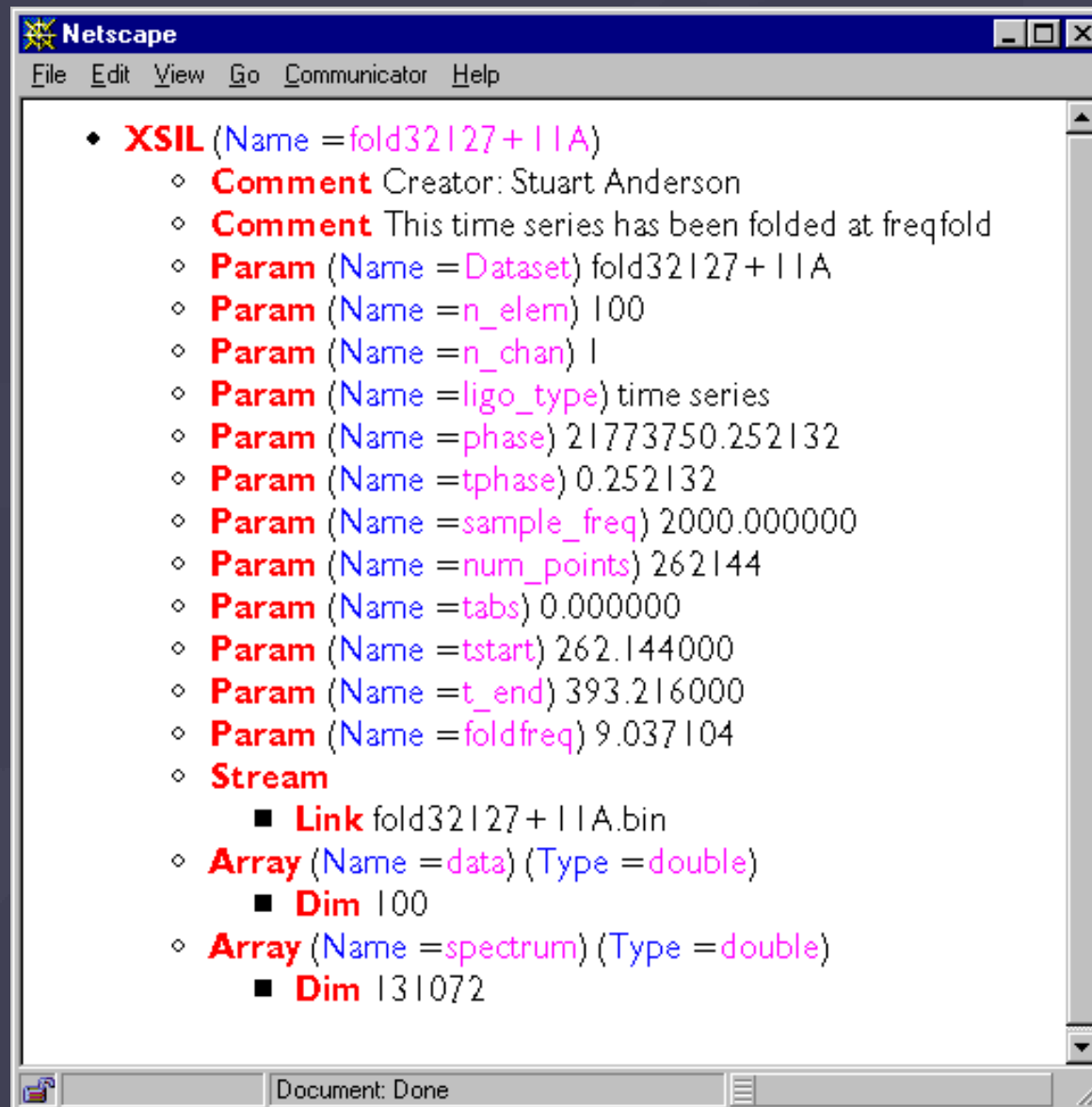

XML Pro -- An XML Editor



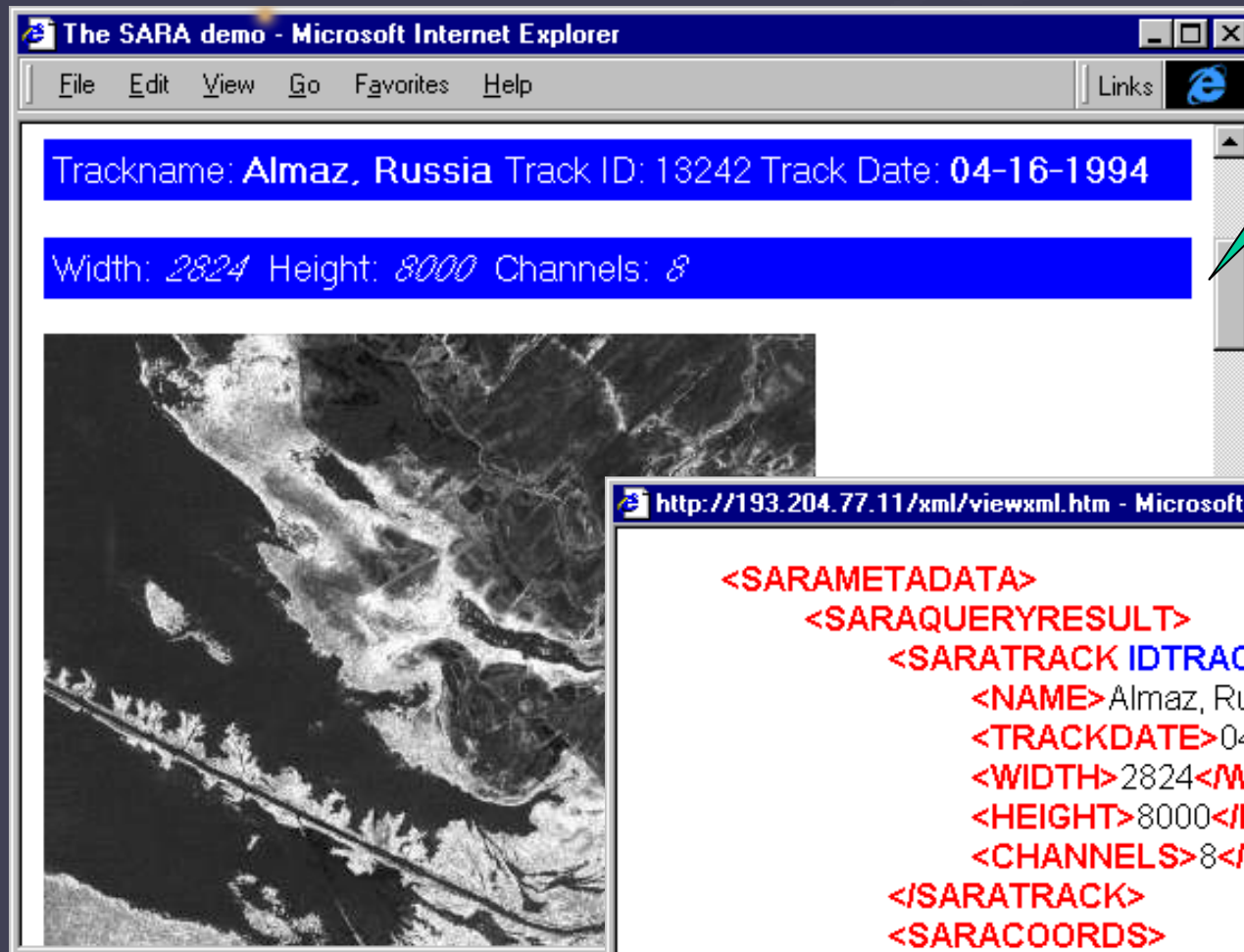
Xeena from IBM



Parsed to HTML

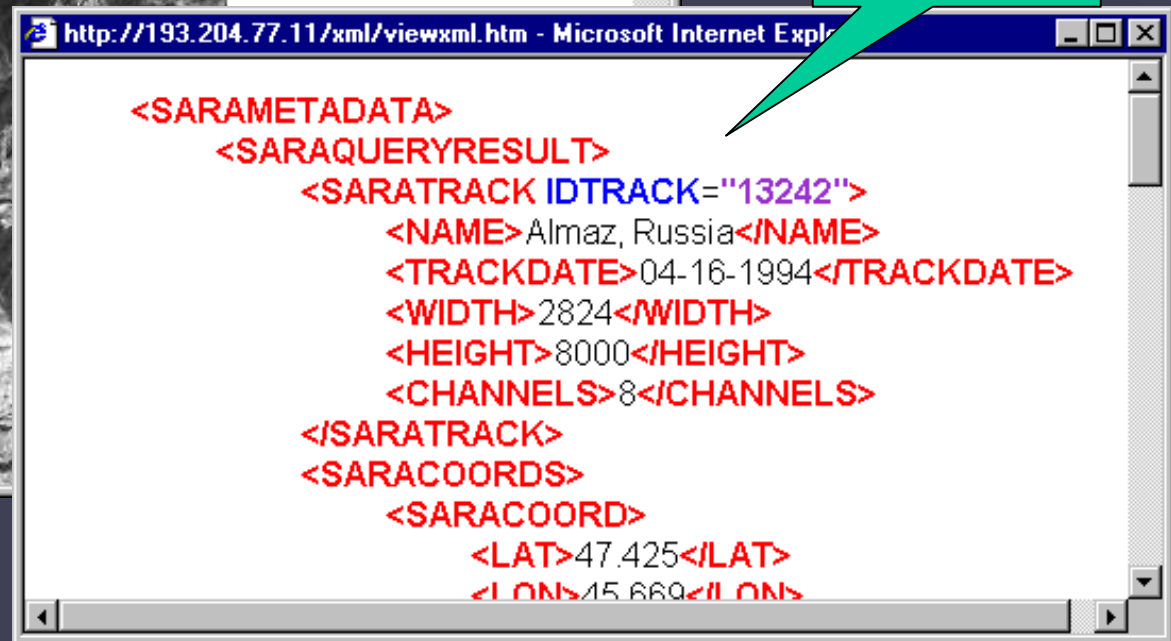


XSL Stylesheets



XML with
XSL
stylesheet
applied

Raw XML



XML Parsing

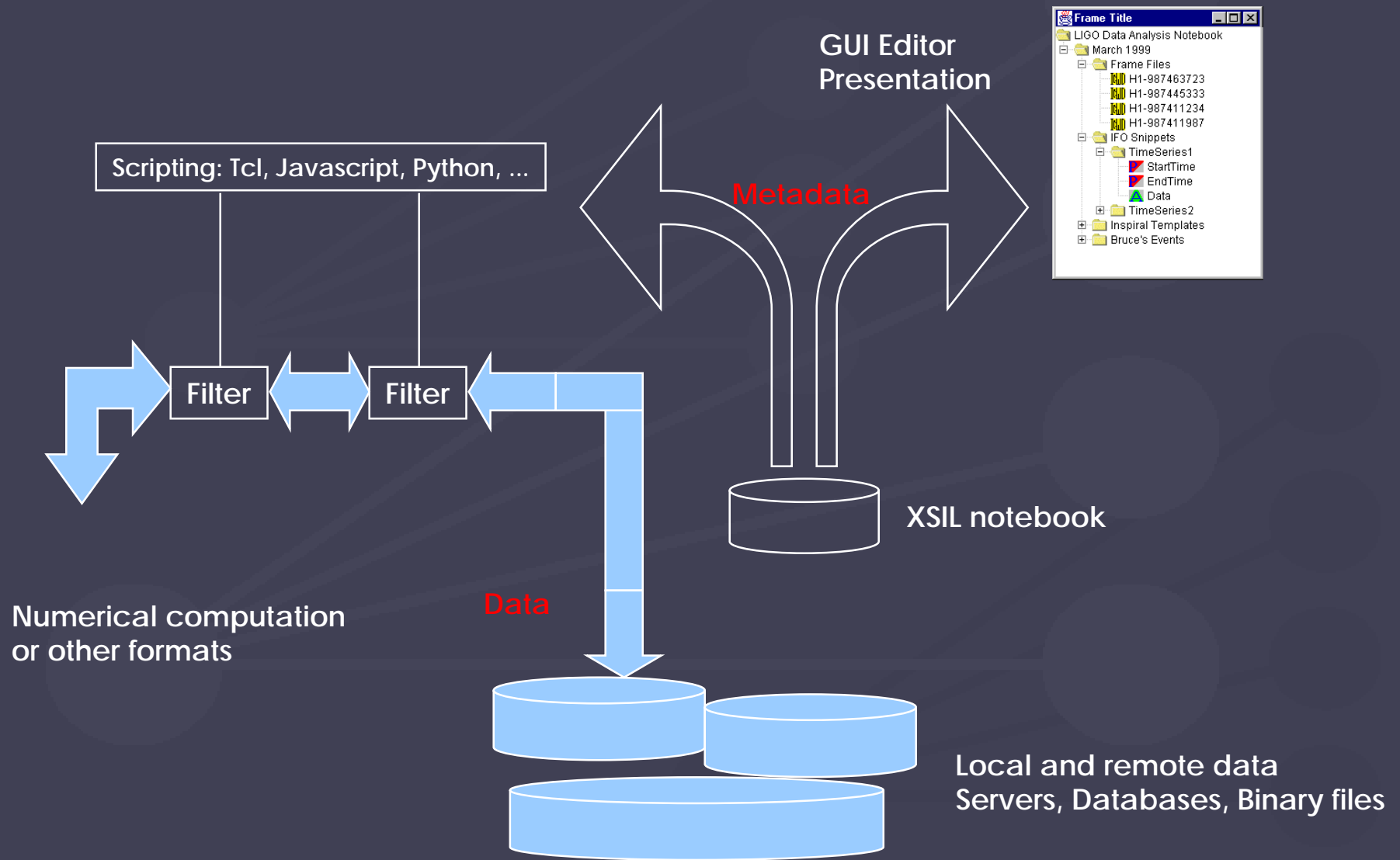
SAX: Event-Based

Handlers for StartElement, Text, EndElement, etc.

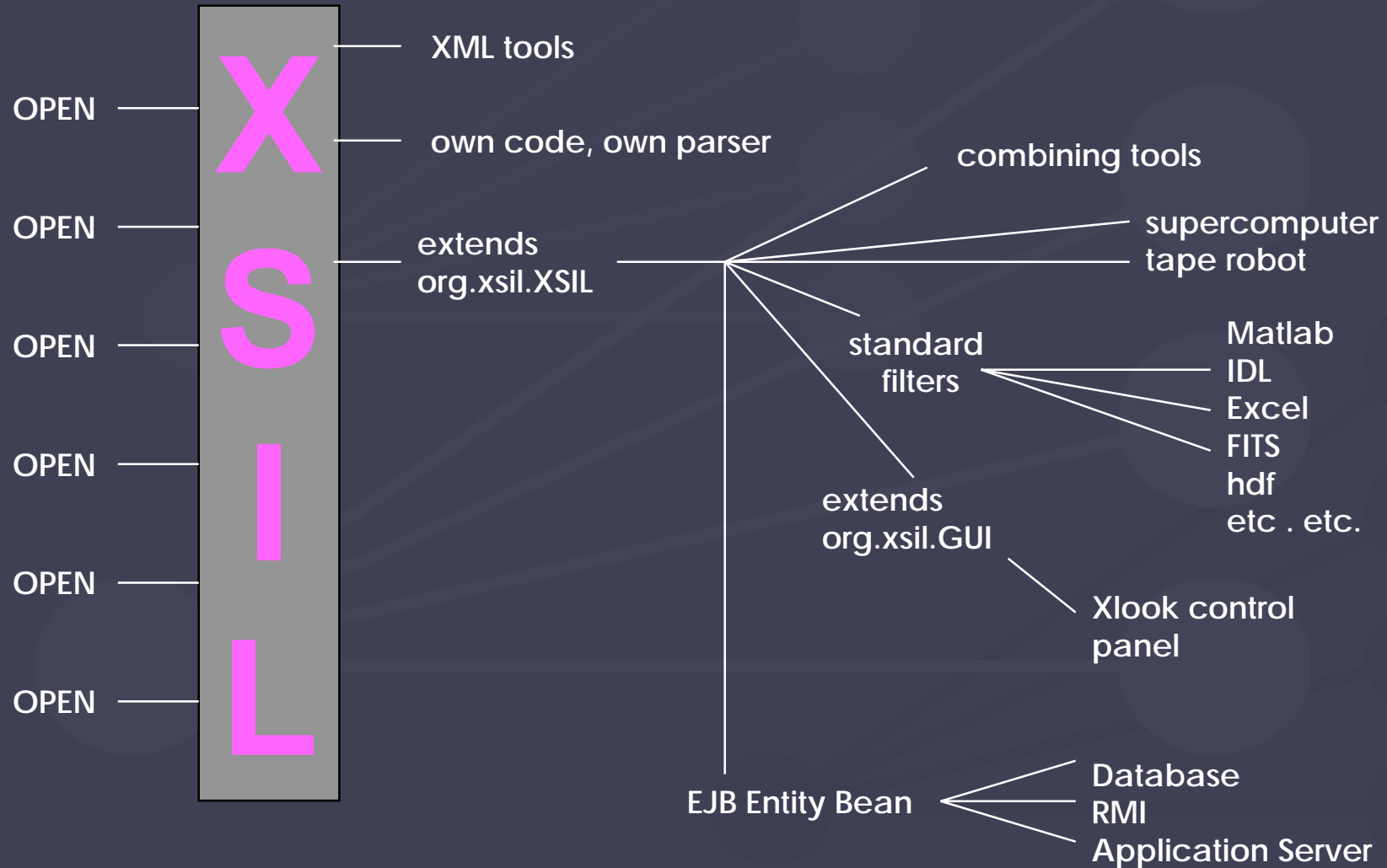
DOM: Document Object Model

Returns a tree-like Document object with data attached

XSIL Information Flow



World Domination Plan



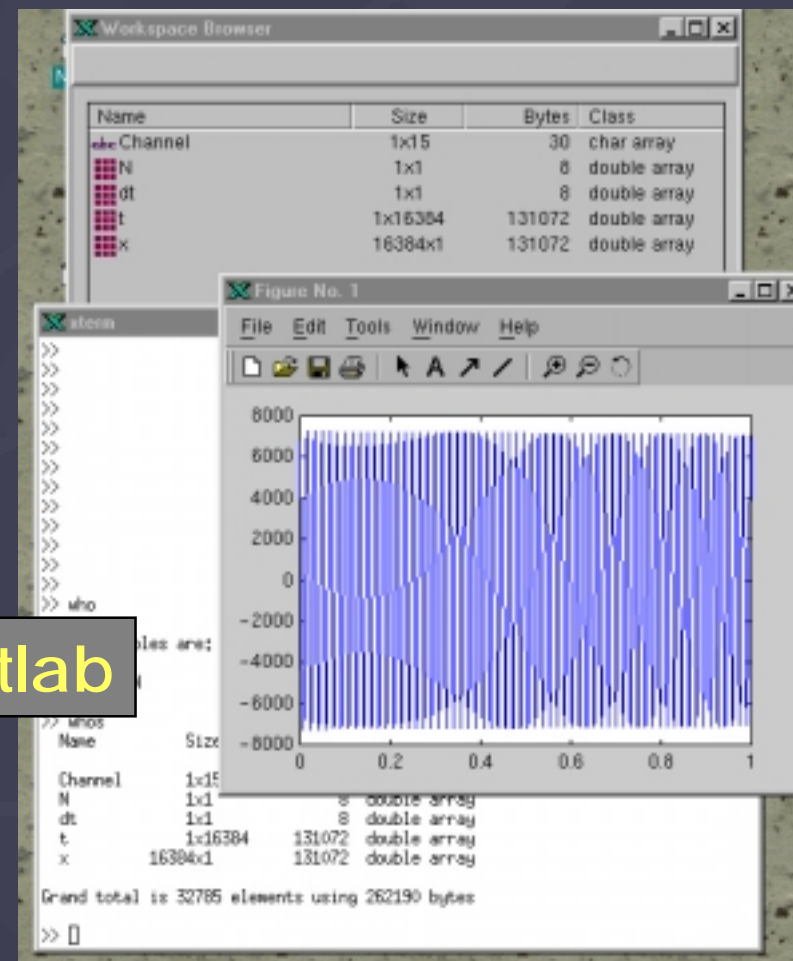
XSIL filter: Matlab

```
xterm
?xml version="1.0"?
<!DOCTYPE LIGO_LW SYSTEM "http://www.cacr.caltech.edu/ligo/XSIL/LIGO_LW.dtd">
<XSIL>
  <Param Name="dt" Type="double" Unit="s">6.10352e-05</Param>
  <Param Name="N" Type="int">16384</Param>
  <Param Name="Channel">H2:LSC-ITMX_EXC</Param>
  <Array Type="float">
    <Dim>16384</Dim>
    <Stream Encoding="BigEndian,base64">
RdNdcEXRdLNFz2/0Rc1PeEXLE4dFyLxvRcZKfkXlvGpFwRdoRb5W9EW7fQxFuIoR
RbV+aEWyWnpFrX6zRavLf0WoYVNFp0ChRaFJ40WdnZNfmdwuRZYGNkWSHC5Fjh6b
RYe0BkWE6utFobYGRXrFcUVyMU1FaWk8RWB0+EVXaS9FTkCWRUT8Z0U7nd9FMiZA
```

XSIL

Matlab

% xsil2matlab < file.xml > file.mat



Some Java

```
package org.xsil.XSIL;
```

```
public class TimeSeries extends XSIL {  
    double startTime = 0.0;  
    double endTime = 1.0;  
    int ndata = 0;  
    double[] data;  
  
    public void constructor(Element e){  
        ...  
        for(int ichild=0; ichild < children.size(); ichild++){  
            Object o = children.elementAt(ichild);  
            if(o instanceof Param){  
                Param p = (Param)o;  
                if(p.getName().equals("StartTime")){  
                    startTime = new Double(p.getText()).doubleValue();  
                }  
                if(p.getName().equals("EndTime")){  
                    endTime = new Double(p.getText()).doubleValue();  
                }  
            }  
            if(o instanceof Array && ((Array)o).getNdim() == 1){  
                ....  
            }  
        }  
    }  
}
```

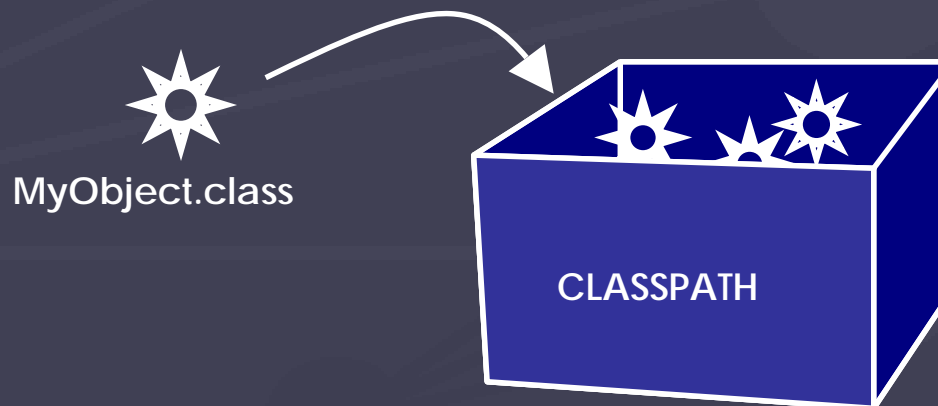
The diagram illustrates the relationship between the code and its structure. A green box labeled "inherited tree structure" has two arrows: one pointing to the `extends XSIL` line, indicating inheritance, and another pointing to the `children` variable, indicating a collection of objects. A blue box labeled "base objects" has two arrows: one pointing to the `Param` class in the `instanceof` check, and another pointing to the `Array` class in the `instanceof` check, indicating the base types of the objects in the tree.

Dynamic Loading

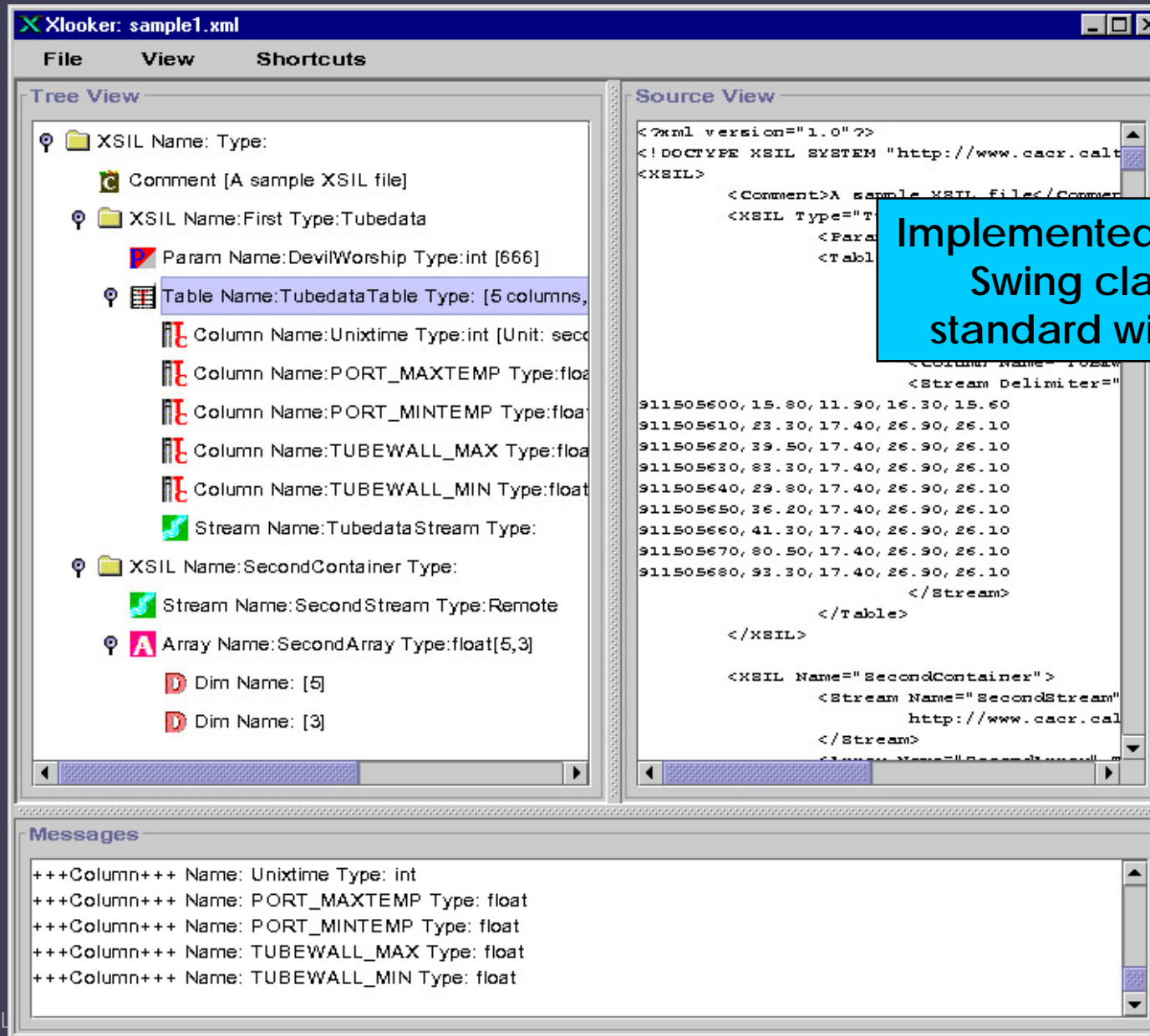
Code is not linked



Just throw it in the box and it will be used!



Xlook Scientific Notebook



Implemented with Java
Swing classes --
standard with Java2

Xlook Scientific Notebook

The screenshot shows the Xlooker application interface. The main window is titled 'Xlooker: sample1.xml'. It has a menu bar with 'File', 'View', and 'Shortcuts'. The 'Tree View' on the left shows a hierarchy of XML elements: 'XSIL Name: Type:', 'Comment [A sample XSIL file]', 'XSIL Name:First Type:Tubedata', 'Param Name:DevilWorship Type:int [666]', 'Table Name:TubedataTable Type: [5 columns, 9 rows]', and 'Column Name:Unixtime Type:int [Unit: seconds] 91'. The 'XSIL Component Viewer' window is open, showing a table with the following data:

Unixtime	PORT_MAXTEMP	PORT_MINTEMP	TUBEWALL
911,505,600	15.800	11.9	16.3
911,505,610	23.3	17.4	26.9
911,505,640	29.8	17.4	26.9
911,505,650	36.200	17.4	26.9
911,505,620	39.5	17.4	26.9
911,505,660	41.3	17.4	26.9
911,505,670	80.5	17.4	26.9
911,505,630	83.300	17.4	26.9
911,505,680	93.300	17.4	26.9

The 'Source View' on the right shows the XML code for the table component. A blue arrow points from the table component viewer to the XML code. A blue box with white text says: 'Implemented with LiveTable Bean from KL Group (www.klg.com)'.

```
<?xml version="1.0" ?>
<!DOCTYPE XSIL SYSTEM "http://www.ca
<XSIL>
  <Table>
    <Column Name
    <Column Name
    <Stream Deli
  </Table>
</XSIL>
<XSIL Name="SecondContainer"
  <Stream Name="Second
    http://www.c
  </Stream>
  <Column Name="Second
```

Xlook Scientific Notebook

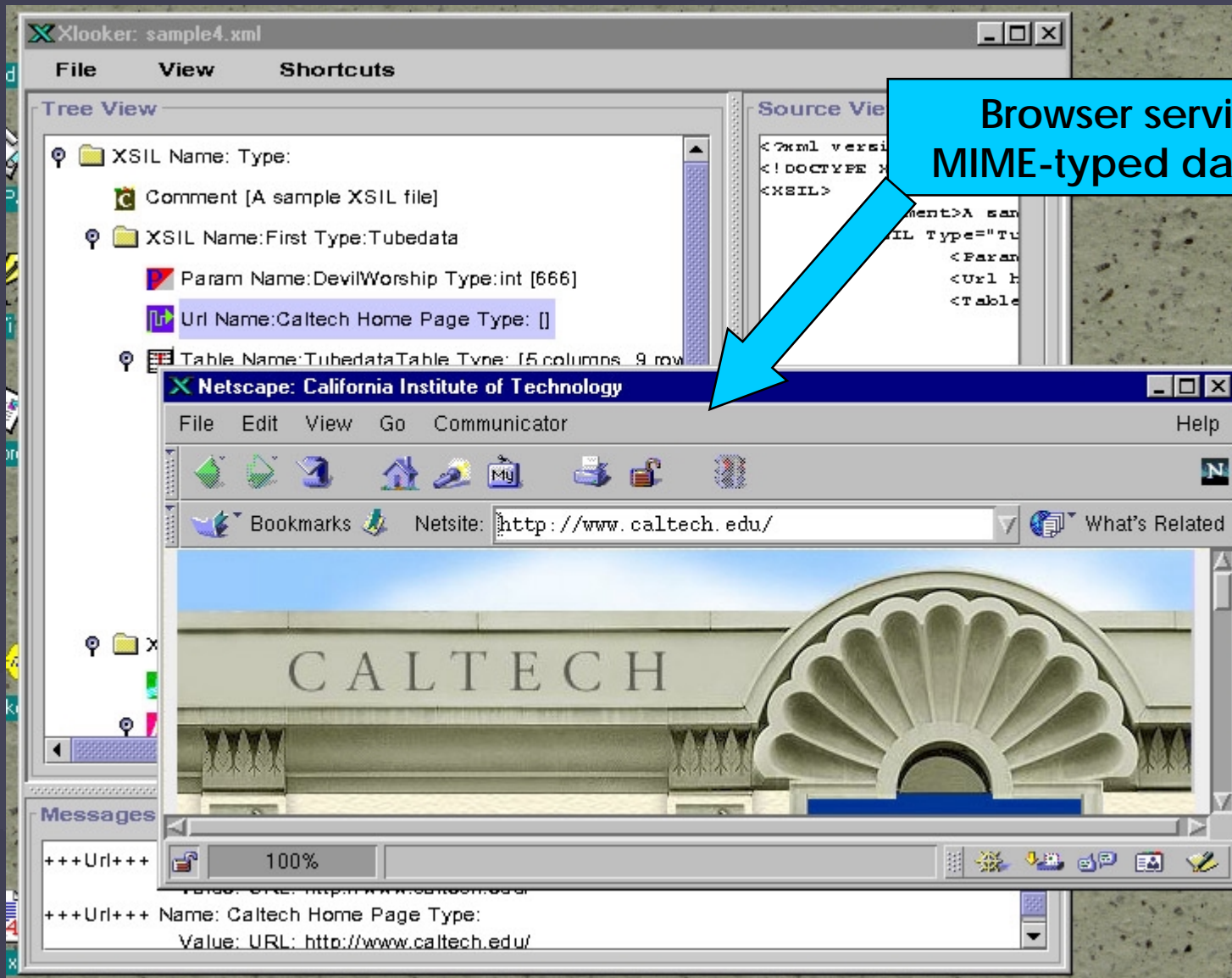
The screenshot displays the Xlook Scientific Notebook application interface. At the top, a menu bar includes 'File', 'View', and 'Shortcuts'. Below this, a 'Tree View' pane lists several parameters: 'Param Name: AverageType Type: int [0]', 'Param Name: Averages Type: int [1]', 'Param Name: Decimation Type: int [1]', 'Param Name: DecimationPremod Type: int [1]', 'Param Name: Channel Type: string [H2:LSC-ITMX_E]', and 'TimeSeries Name: Type: TimeSeries from 0.2 to 1.2'. To the right, a 'Source View' pane shows XML-like code for a 'TimeSeries' object, including attributes like 'channel', 'type', 'start', 'end', and 'array'. A blue callout box with an arrow points from the 'TimeSeries' entry in the Tree View to a 'TimeSeries' component window. This window, titled 'XSIL Component Viewer', shows the component's name and type, and a line graph of a high-frequency sine wave. Below this, another 'XSIL Component Viewer' window shows a 'TimeSeries' component with a line graph of a decaying exponential curve. A 'JClass Chart Properties' dialog is also visible, showing tabs for 'File', 'Look & Feel', 'Chart', 'Data View', 'Axes', 'Chart Area', 'Legend', and 'Titles'. The 'Chart' tab is active, showing 'Chart Styles', 'Labels', and 'Data Binding' sub-tabs. The 'General' sub-tab is selected, showing 'View Name' and checkboxes for 'isShowing', 'isShowingInLegend', and 'FastUpdate'. The 'Data' sub-tab is also visible, showing a 'Customize' button. The 'TimeSeries' component window also has a 'Customize' button. The 'JClass Chart Properties' dialog has a 'Customize' button. The 'TimeSeries' component window has a 'Customize' button. The 'TimeSeries' component window has a 'Customize' button.

Implemented with JChart
Bean from KL Group
(www.klg.com)

XSIL

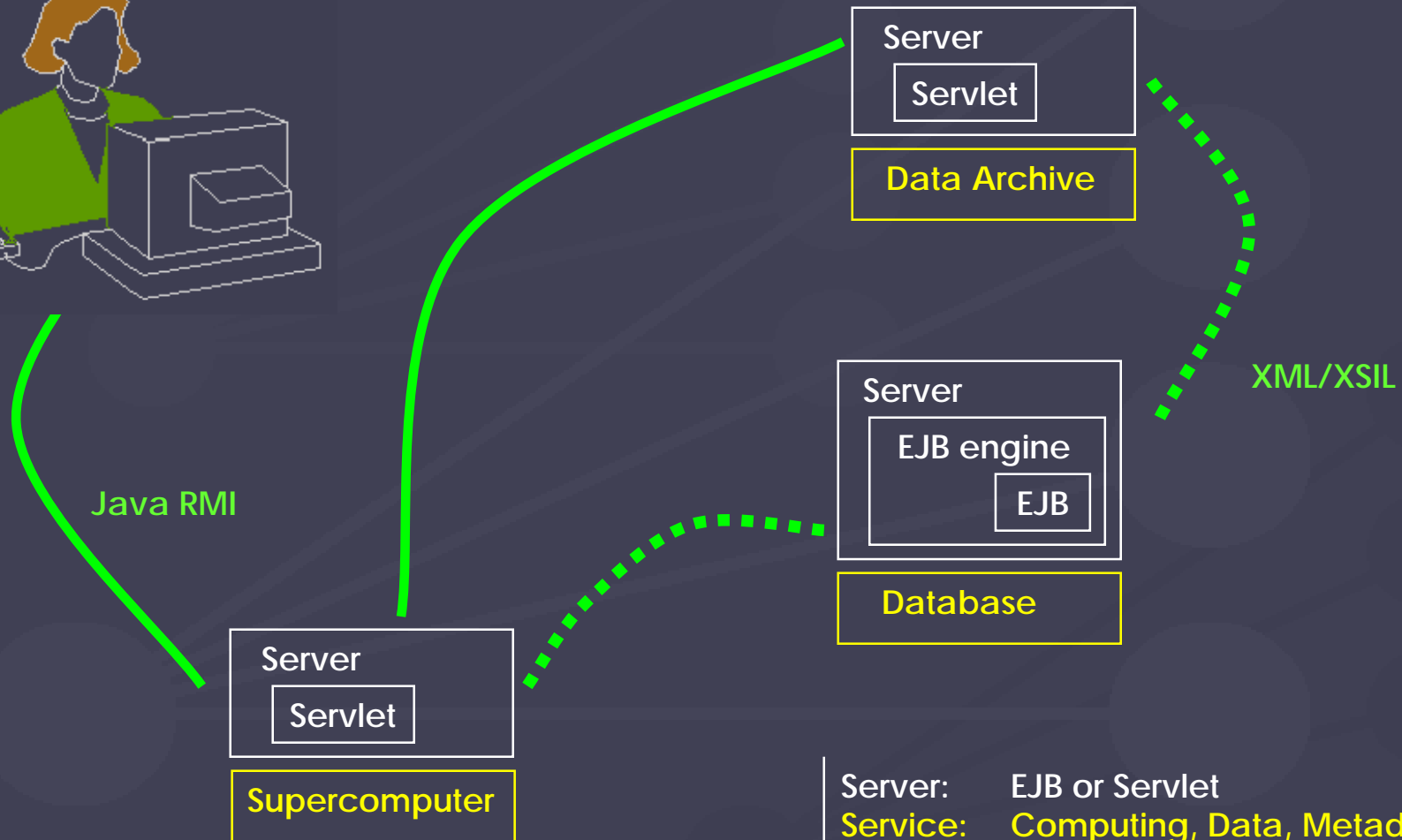
Williams 1999

Xlook Scientific Notebook



Browser services for
MIME-typed data (URL's)

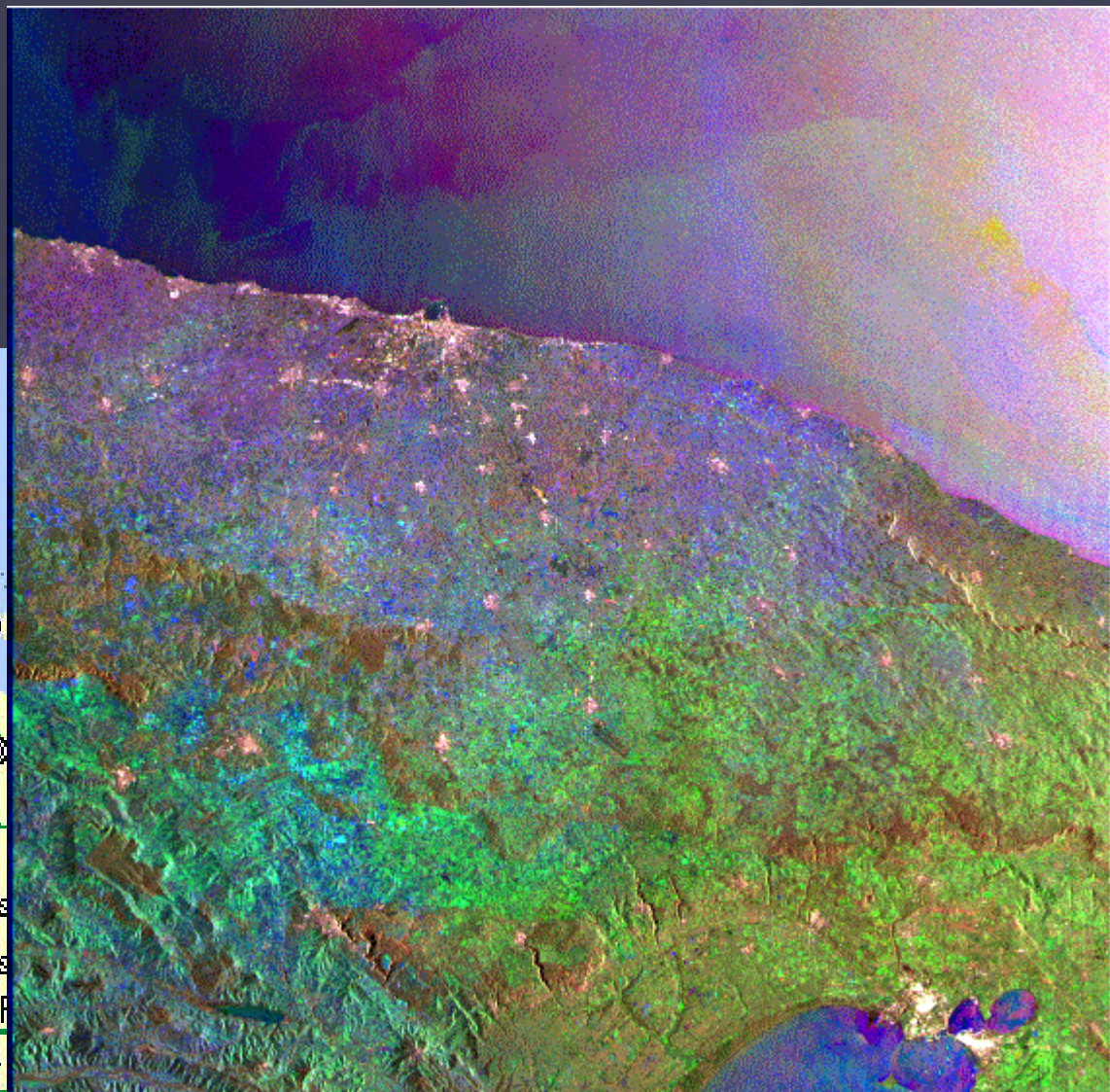
Distributed Science Data



Digital Puglia

Synthetic Aperture Radar Atlas

Multichannel
Image Retrieval
Format Selection
Processing Menu



NPACI Digital Sky Project



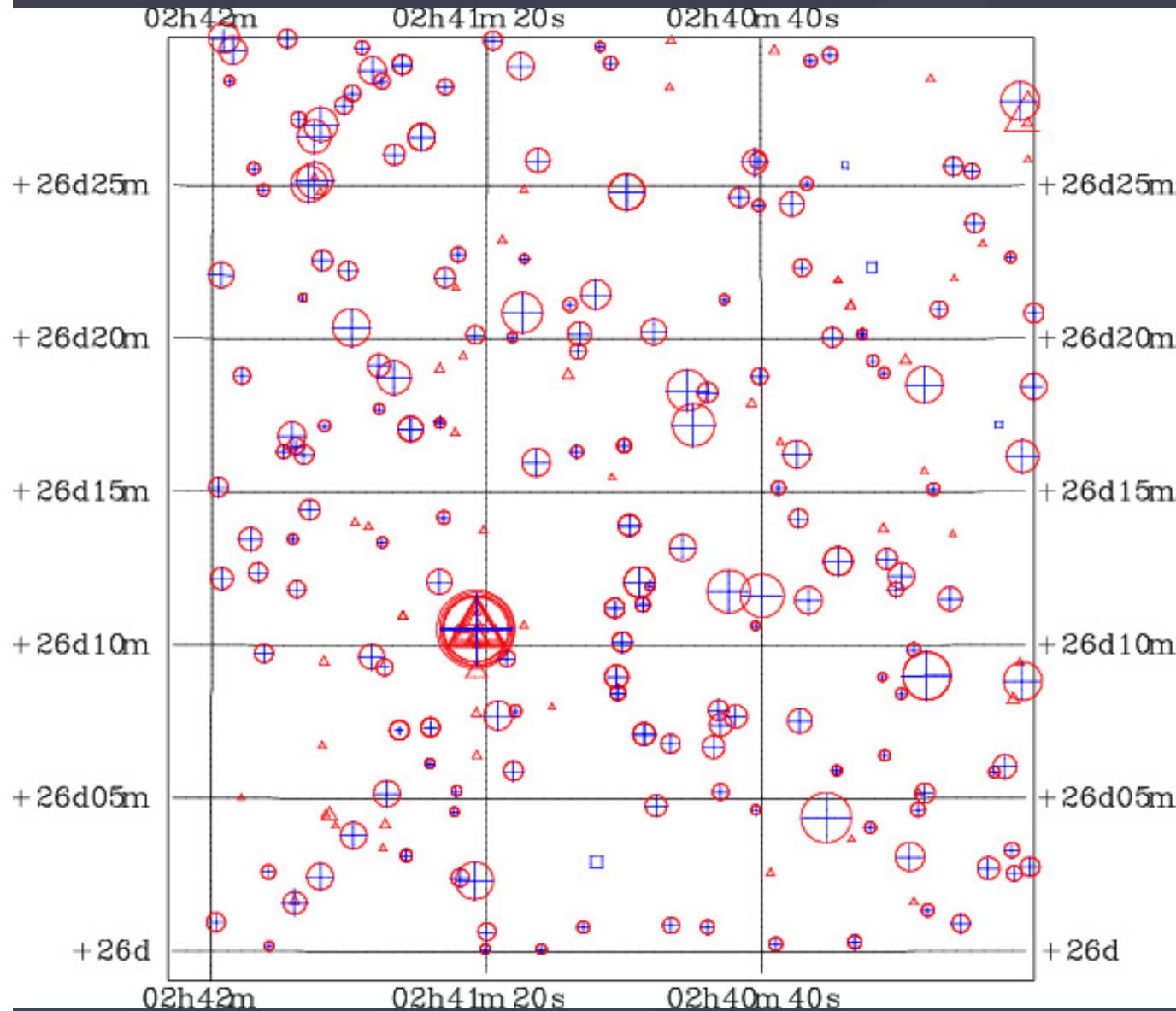
The Sky is a Database

```
\char DATABASE='SQL: FROM tmassr.pts_samp_cat (2MASS Sampler Point Source Catalog)'
\char EQUINOX='J2000'
\char SKYAREA='No constraint'
\char WHERE='SQL: WHERE k_m BETWEEN 12.0 AND 12.1'
\char SELECT='SQL: SELECT (Selected column names follow in next row.)'
```

ra	dec	j_m	j_msig	h_m	h_msig	k_m	k_msig	rd_flg	cc_flg	extd_flg
346.995577	27.962376	12.282	0.031	11.990	0.035	12.000	0.034	222	000	
40.202781	28.413649	12.679	0.030	12.152	0.033	12.000	0.030	222	000	
40.956840	29.456654	12.645	0.029	12.165	0.032	12.000	0.031	222	000	
106.165728	20.975992	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
106.556947	20.133484	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
134.320346	14.391330	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
135.038423	17.798285	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
344.606603	29.220026	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
344.850216	25.119240	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
346.759144	28.661133	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
43.087673	29.482285	12.359	0.030	12.055	0.031	12.000	0.022	222	000	
43.960498	28.390268	12.359	0.030	12.055	0.031	12.000	0.022	222	000	
106.135655	19.268063	12.259	0.030	12.055	0.031	12.000	0.022	222	000	
106.334142	17.934162	12.259	0.030	12.055	0.031	12.000	0.022	222	000	

<input checked="" type="checkbox"/>	ra	right ascension (J2000 decimal deg)
<input checked="" type="checkbox"/>	dec	declination (J2000 decimal deg)
<input type="checkbox"/>	err_maj	major axis of position error ellipse
<input type="checkbox"/>	err_min	minor axis of position error ellipse
<input type="checkbox"/>	err_ang	position angle of error ellipse major axis (E of N)
<input checked="" type="checkbox"/>	j_m	J selected "default" magnitude or 95% confidence upper limit
<input checked="" type="checkbox"/>	j_msig	J "default" mag uncertainty
<input checked="" type="checkbox"/>	h_m	H selected "default" magnitude or 95% confidence upper limit
<input checked="" type="checkbox"/>	h_msig	H "default" mag uncertainty
<input checked="" type="checkbox"/>	k_m	K selected "default" magnitude or 95% confidence upper limit
<input checked="" type="checkbox"/>	k_msig	K "default" mag uncertainty
<input checked="" type="checkbox"/>	rd_flg	source of JHK "default" mags (AKA "read flag")
<input type="checkbox"/>	bl_flg	indicates # JHK components fit to source (each digit=0 1 2)
<input checked="" type="checkbox"/>	cc_flg	indicates artifact contamination and/or confusion
<input checked="" type="checkbox"/>	extd_flg	indicates src associated with or contaminated by an ext. src
<input type="checkbox"/>	id	id

Cross-Identification



2MASS versus DPOSS cross-identification

2MASS : $j_m + 15$

DPOSS: $l_mtotn \leq 18$

□ DPOSS unmatched

○ 2MASS matched

⊕ DPOSS matched

△ 2MASS unmatched