SC4DEVO-3 Workshop Introduction

Bob Mann
University of Edinburgh

Outline

- Introductions
- The Virtual Observatory
- SC4DEVO
- SC4DEVO-3
  - Visualization in the Virtual Observatory

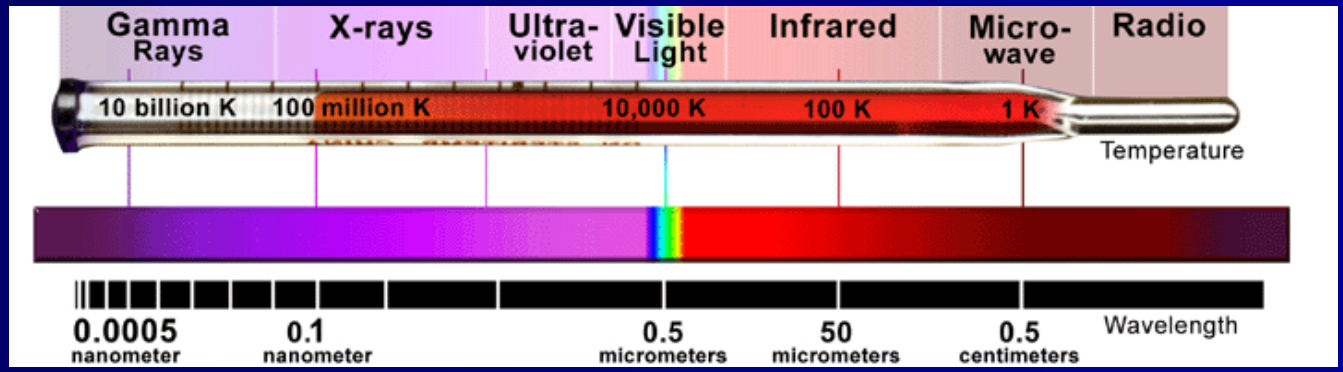
Introductions

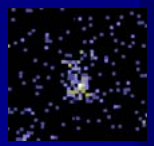
The Virtual Observatory

- Motivation
- Current Status

Observational Astronomy

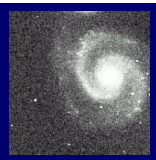
- Electromagnetic spectrum



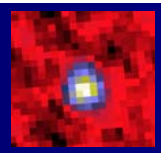


ROSAT ~keV

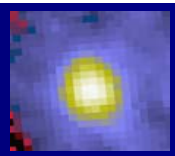
DSS Optical



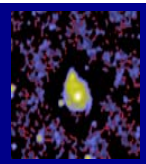
2MASS 2μ



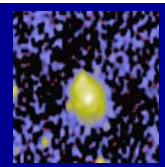
IRAS 25μ



IRAS 100μ



NVSS 20cm



WENSS 92cm

- Differences in:
  - Instrumental characteristics
  - Physical emission mechanism

M51 images from Alex Szalay & Jim Gray

Motivation for the VO

- Multiwavelength astronomy
- Data stored in λ-specific data centres
  - e.g. in the UK:

    optical/IR - Edinburgh & Cambridge
    X-ray - Leicester
    Radio - Jodrell Bank
- Need to federate distributed databases to enable multiwavelength astronomy

Motivation for the VO

- Data volumes – exponential growth
  - Largest sky survey archives ~1-10TB
  - Total astronomical data 100s of TB
  - Volume doubles every ~12 months
- Volumes needed for analysis growing
  - User can't download all data to local disk
  - Need remote storage and processing

So, the VO is

a global federation of astronomical data and computes resources

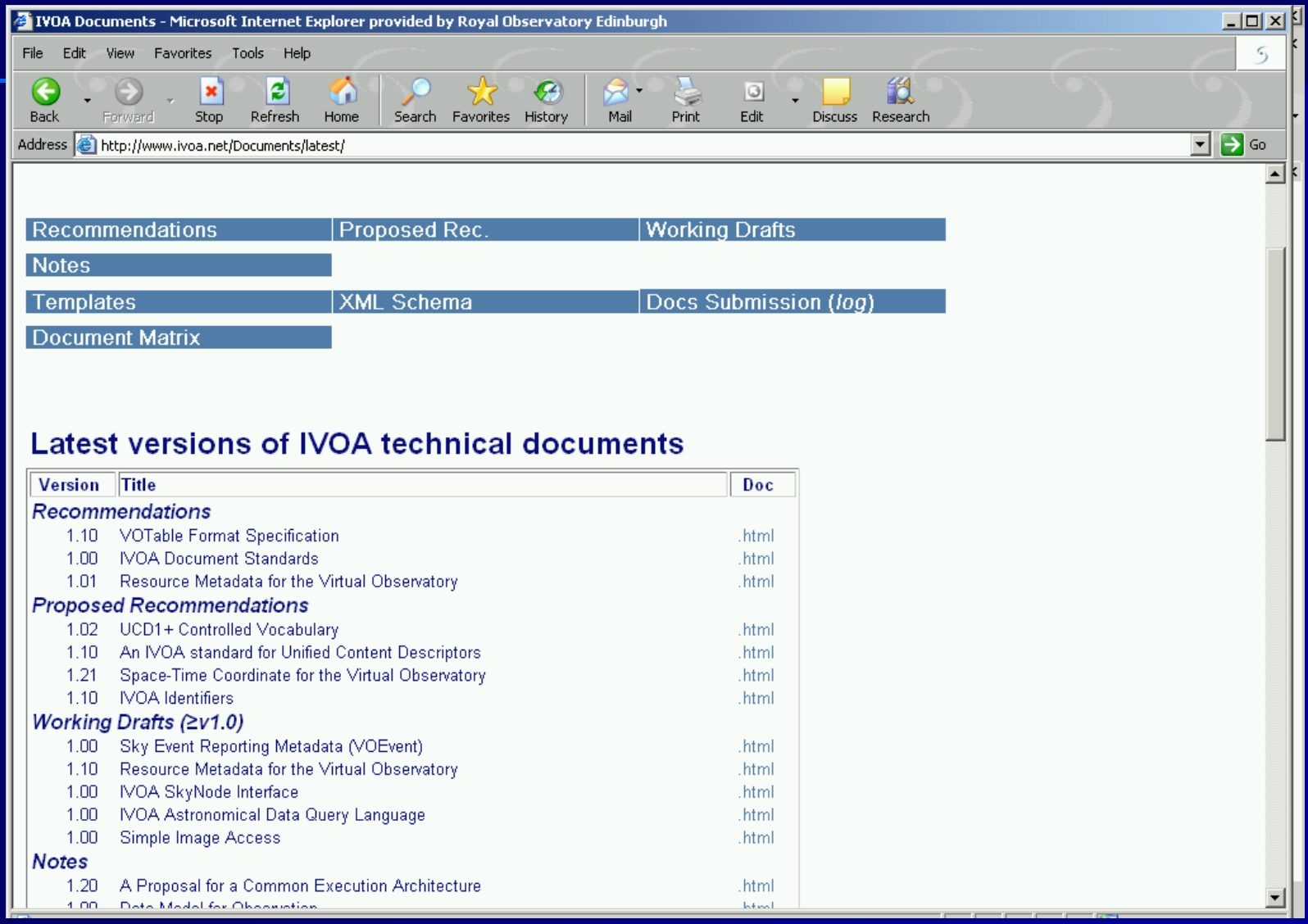
and developing it requires solutions to

- Technical problems
  - building the computational infrastructure
- Sociological problems
  - defining the interoperability standards

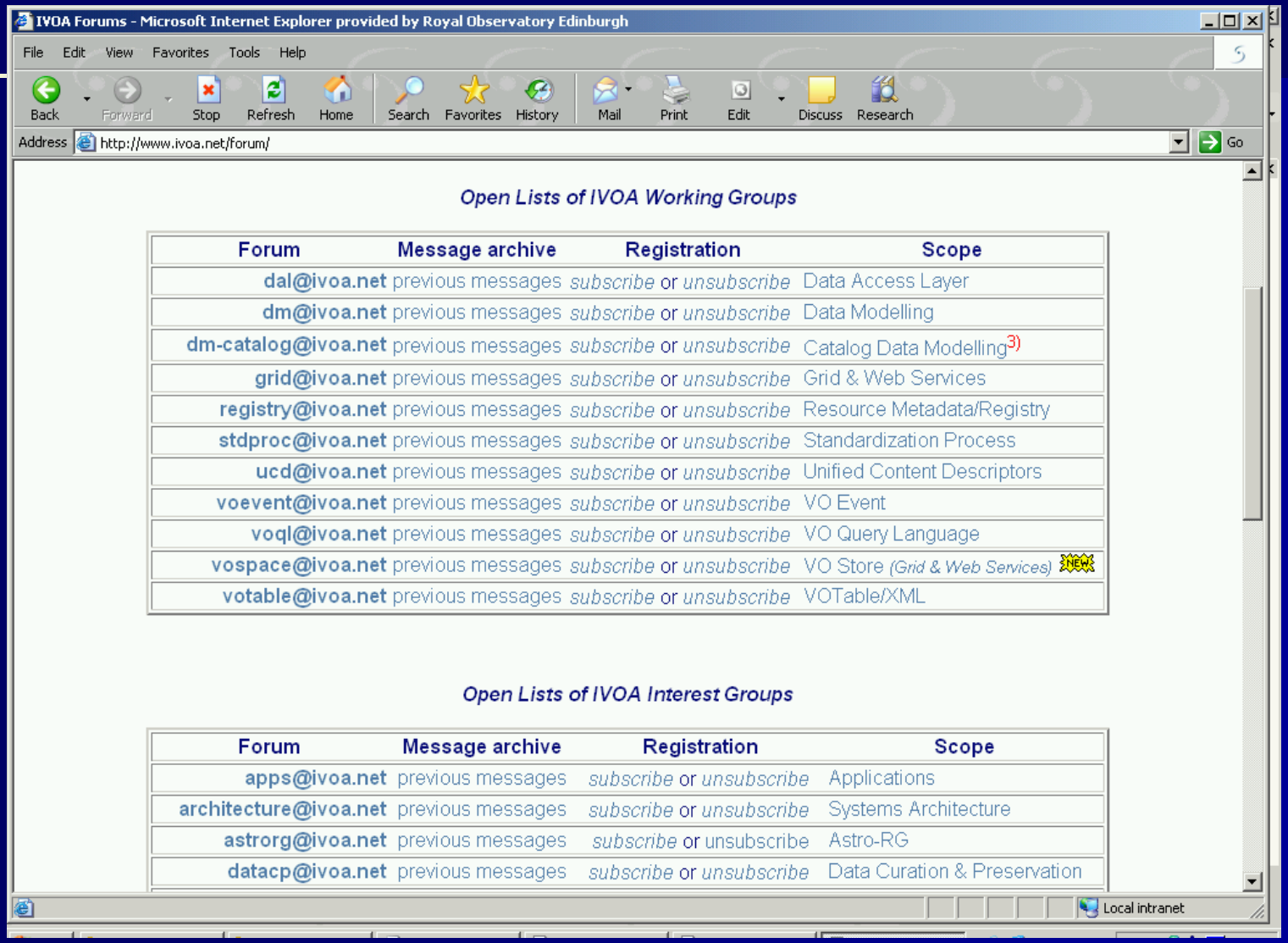
International Virtual Observatory Alliance

EU, China, India, Canada, Spain, Italy, Armenia, France, Germany, Hungary, Japan, Korea, USA, Russia, UK, Australia

IVOA standards process



Working & Interest Groups



Open Lists of IVOA Working Groups

Forum	Message archive	Registration	Scope
dal@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Data Access Layer
dm@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Data Modelling
dm-catalog@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Catalog Data Modelling[3)]
grid@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Grid & Web Services
registry@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Resource Metadata/Registry
stdproc@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Standardization Process
ucd@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Unified Content Descriptors
voevent@ivoa.net	previous messages	*subscribe* or *unsubscribe*	VO Event
voql@ivoa.net	previous messages	*subscribe* or *unsubscribe*	VO Query Language
vospace@ivoa.net	previous messages	*subscribe* or *unsubscribe*	VO Store *(Grid & Web Services)* NEW
votable@ivoa.net	previous messages	*subscribe* or *unsubscribe*	VOTable/XML

Open Lists of IVOA Interest Groups

Forum	Message archive	Registration	Scope
apps@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Applications
architecture@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Systems Architecture
astrorg@ivoa.net	previous messages	*subscribe* or unsubscribe	Astro-RG
datacp@ivoa.net	previous messages	*subscribe* or *unsubscribe*	Data Curation & Preservation

VO Status

- IVOA is three years old
  - Some VO projects started second three years of funding
- Good international collaboration
- Basic VO infrastructure agreed
  - Details in talk by John Taylor
- VO taking shape
  - Maybe ~3 more years until it's really there

SC4DEVO

- What the acronym means
- Where the money came from
- Progress to date

SC
4
DE
VO

Service **C**omposition

for

Data **E**xploration in the

Virtual **O**bservatory

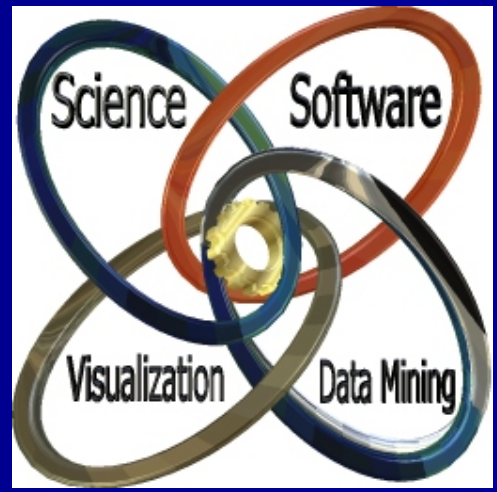
SC – Service Composition

- VO based on service-oriented architecture
  - e.g. resources (databases, compute nodes, analysis applications) represented logically as *services*
  - Services interact by exchanging messages
  - Standard message types comprise interface
- Model used widely in e-Science/Grid Computing
  - Good for loosely-coupled heterogeneous distributed systems
- Services must be composed to make a larger system, like the VO

DE – Data Exploration

- DE = Data Mining + Visualization
- The *coupling* of data mining and visualization is the key
- A route into the data
  - Finding significant patterns to follow-up
- Is this situation unique to astronomy?
  - No, e-science is driven by data avalanche

SDMIV Workshop



- *Scientific Data Mining, Integration and Visualization*
- Edinburgh, October 2002
- 50 participants
  - astronomy, atmospheric science, bioinformatics, chemistry, digital libraries, engineering, environmental science, experimental physics, marine sciences, oceanography, *plus* CS - data mining, visualization, Grid computing

http://www.nesc.ac.uk/talks/sdmiv/report.pdf

Lessons from SDMIV

- CS and Apps people want to interact
  - See mutual benefit from collaboration
- Common problems in all disciplines
  - Lots of distributed data in many formats
- Lots of DM and Vis software out there, but…
  - Doesn’t match how we work now
  - Don’t know what to use or where to find it
  - How does it fit into the computational infrastructure we’re building?...VO, Grid, etc

The SC4DEVO Proposal

- UK e-Science Programme launches "International Sister Projects" initiative
  - Money for advancing collaboration
- Aim:
  - Work out how to do VO data exploration
  - How to generalise to e-science
- Anglo/Australian/US consortium
  - UK: AstroGrid (Bob Mann et al)
  - Aus: CSIRO Grid Computing group (Dave Abel)
  - US: GRIST project (Roy Williams et al)
  - *plus* DM, Vis, workflow, Grid researchers

SC4DEVO progress to date

- One of four projects funded in 1st year
  - Funding for 4 workshops in 2004 & 2005
- SC4DEVO-1: Caltech, July 2004
  - General overview of SC4DEVO issues
- SC4DEVO-2: Edinburgh, Nov/Dec 2004
  - Focus on data mining
- SC4DEVO-3: now, visualization
- SC4DEVO-4: late 2005, summary

SC4DEVO-3

- Goals
- Issues
- Format
- Practical Info

SC4DEVO-3 Goals

- Foster interaction between people interested in visualization in the VO
  - Get visualization researchers thinking about requirements from astronomy
  - Help astronomers learn more about what is already done in the vis community
  - Share experiences amongst those already working in VO visualization

SC4DEVO-3 Issues

- Visualization of image data
- Visualization of multivariate data
- Coupling data mining and visualization
- Re-using existing vis tools and techniques in the VO

(More on detailed science requirements from Mark Allen and Andrew Hopkins)

Image visualization

- Much experience within astronomy
  - RVS: Malte Marquarding
  - Aladin: Thomas Boch
- Issues:
  - Combining distributed data sources
  - Large images: server-side vis engines
  - Collaborative visualization

Multivariate visualization

- Less experience in astronomy
  - VisIVO: Claudio Gheller
  - Existing vis work: Richard Holbrey
- Issues:
  - Coupling data mining and visualization
  - Interactive data exploration in the VO
  - Scalability (Masa Takatsuka)
    - Sample size and dimensionality

Re-using vis tools in the VO

- What useful tools are out there?
  - Can they be VO-enabled? (John Taylor)
    - Data formats, use in worflow, standards for data access and data storage
- Can the Grid help? (Garry Smith)
- Are there wider principles from the vis world of which we should be aware?
  - Don't want to reinvent wheels

SC4DEVO-3 format

- ***Informal discussion meeting***
  - so, please question, heckle, interrupt, etc
- One hour slots
  - 45 mins talk, 15 mins discussion
- Identify major issues as we go along
  - Discussion session on Friday afternoon

Practical Info

- Toilets
- Dinner tonight:
  - 7.00 Lachlan's Restaurant,
    Macquarie Graduate School of Management
- Presentations
- Anything else?

Session 1: Introduction and Science Drivers

- 10.30 Challenges and scientific requirements for visualization in the VO (***Mark Allen***)
- 11.30 Tea/Coffee
- 12.00 Looking into the dark: a visualization wish-list for the Virtual Observatory (***Andrew Hopkins***)