The Astronomers' Data Manifesto

Ray Norris,
CSIRO Australia Telescope National Facility

Astronomical Data: Good News

- Public-domain data
- Astronomical Data Centres
- The Virtual Observatory

Good News

- Public-domain data
- Astronomical Data Centres
- The Virtual Observatory

Security, confidentiality, and IP protection are not major issues in astronomy

- most data are in the public domain
- few privacy or commercial issues

Good News

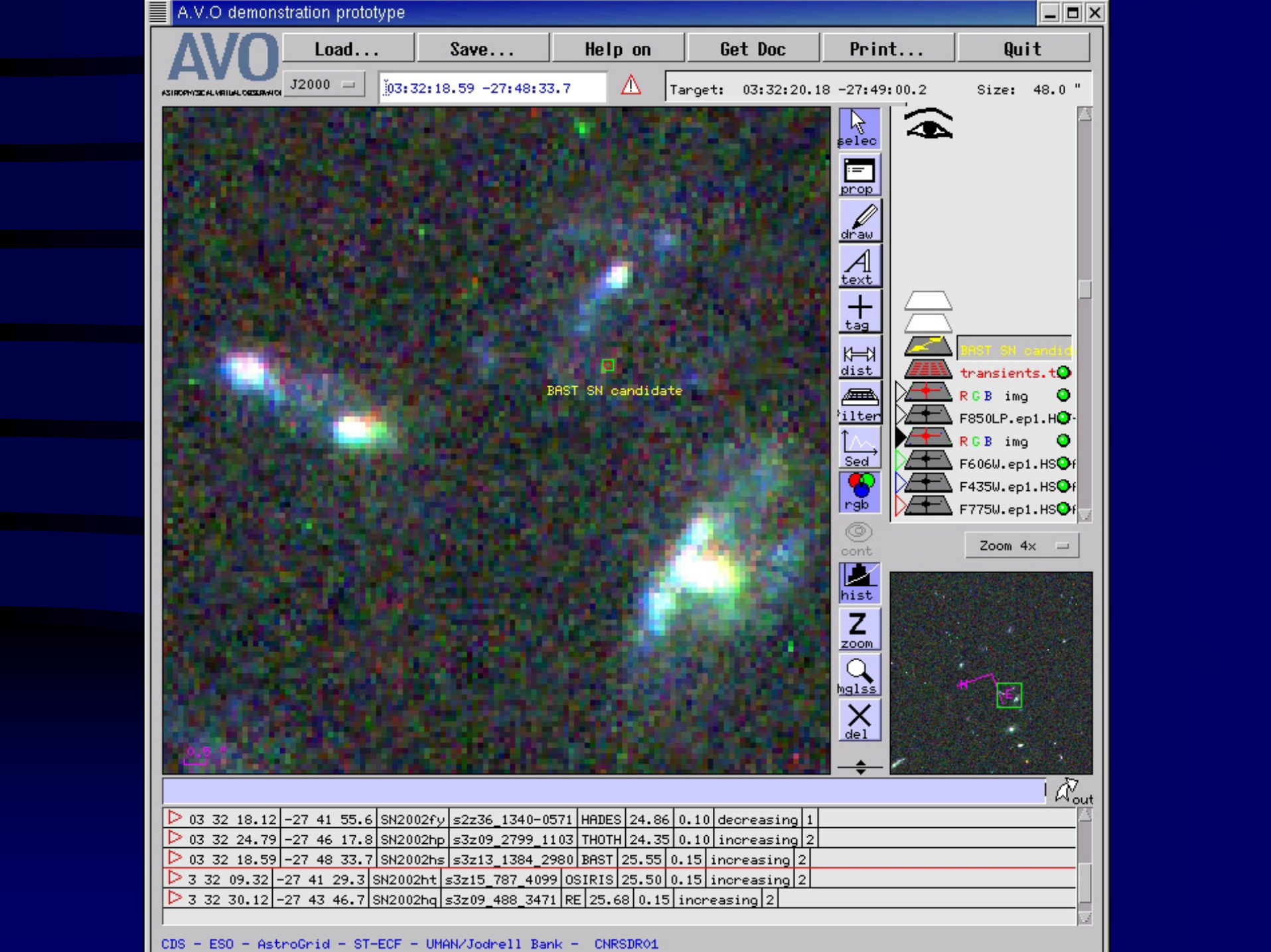
- Public-domain data
- Astronomical Data Centres
- The Virtual Observatory

Astronomical Data Centres

- Centre de Données astronomiques de Strasbourg, France (CDS)
  - holds electronic copies of published astronomical data at all wavelengths, surveys, etc., with tools to access them.
- NASA Extragalactic Database (NED)
  - Interprets and combines extragalactic data
  - E.g. show the energy distribution of this source over all wavelengths
- Astronomical Data System (ADS)
  - All published astronomical literature
  - Includes links to data centres
- Others (Canadian, Russian, etc…)

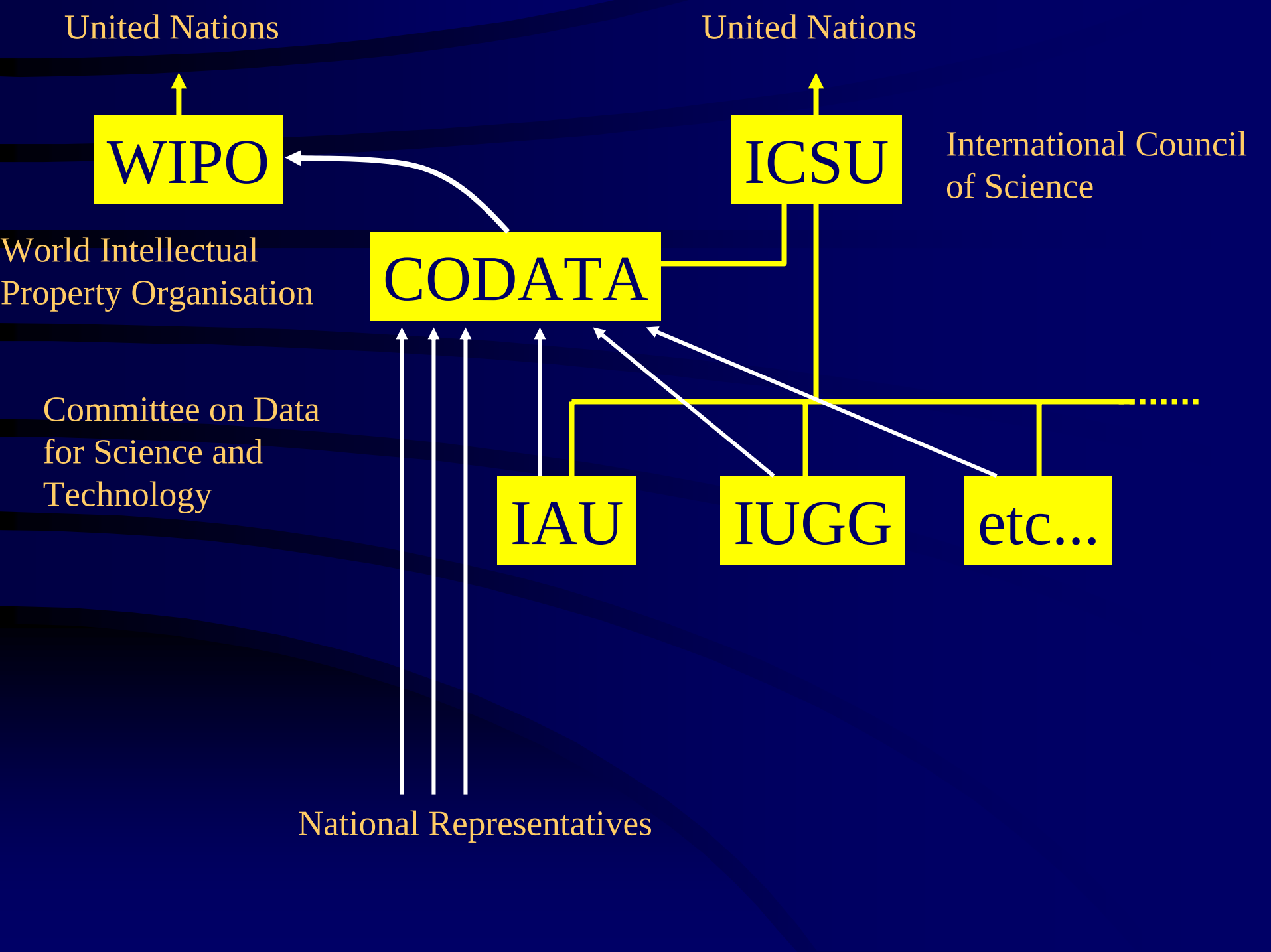
Good News

- Public-domain data
- Astronomical Data Centres
- The Virtual Observatory


A.V.O demonstration prototype
AVO
Load...
Save...
Help on
Get Doc
Print...
Quit
J2000
03:32:18.59 -27:48:33.7
Target: 03:32:20.18 -27:49:00.2
Size: 48.0 "
selec
prop
draw
text
tag
dist
filter
Sed
rgb
cont
hist
zoom
mglss
del
BAST SN candidate
transients.t
R G B img
F850LP.ep1.H
R G B img
F606W.ep1.HS
F435W.ep1.HS
F775W.ep1.HS
Zoom 4x
out
03 32 18.12 -27 41 55.6 SN2002fy s2z36_1340-0571 HADES 24.86 0.10 decreasing 1
03 32 24.79 -27 46 17.8 SN2002hp s3z09_2799_1103 THOTH 24.35 0.10 increasing 2
03 32 18.59 -27 48 33.7 SN2002hs s3z13_1384_2980 BAST 25.55 0.15 increasing 2
3 32 09.32 -27 41 29.3 SN2002ht s3z15_787_4099 OSIRIS 25.50 0.15 increasing 2
3 32 30.12 -27 43 46.7 SN2002hq s3z09_488_3471 RE 25.68 0.15 increasing 2
CDS - ESO - AstroGrid - ST-ECF - UMAN/Jodrell Bank - CNRSDR01

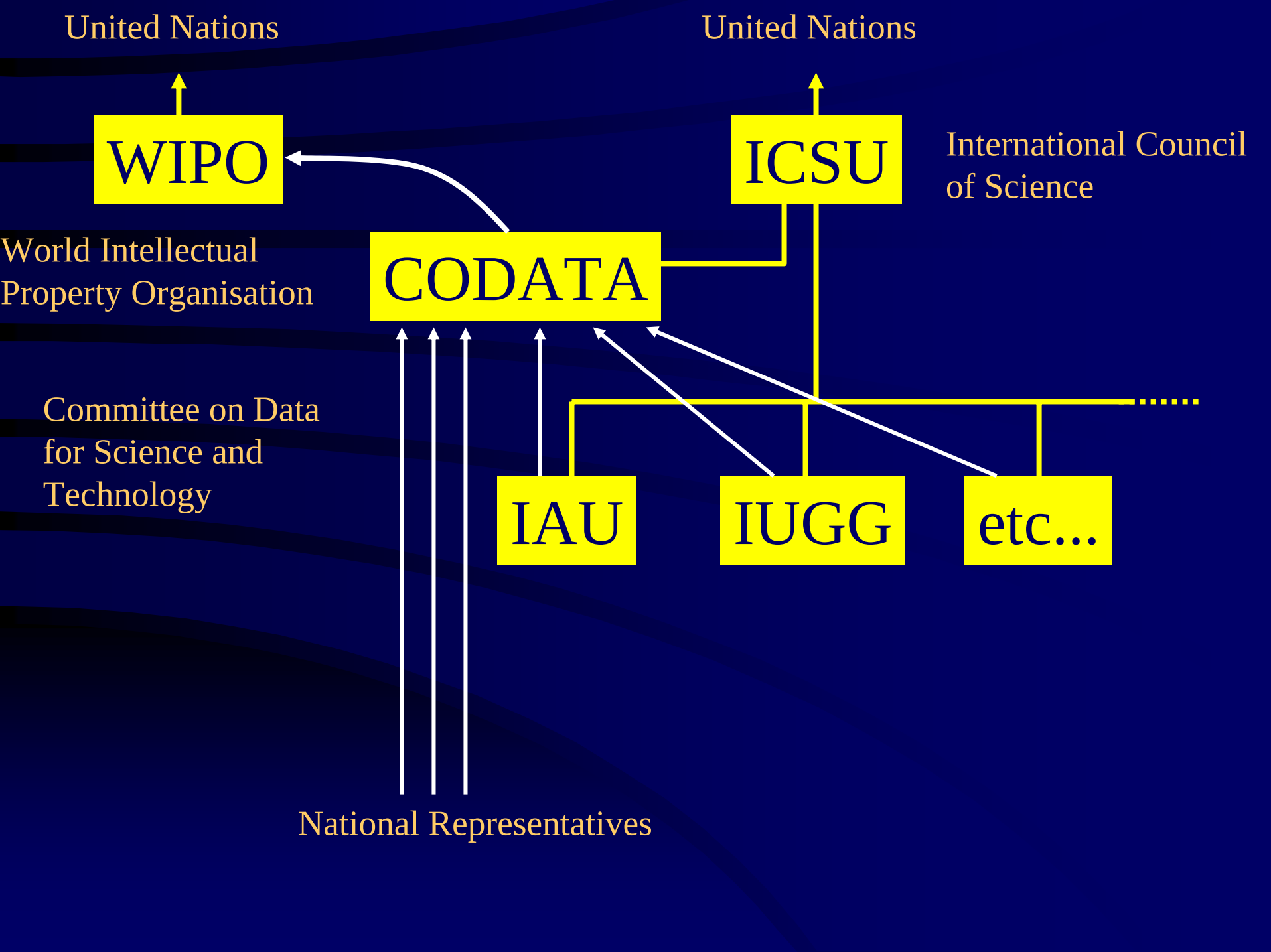
The Challenge

- How well do we manage astronomical data?
- Is there a consensus on how astronomical data should be managed?
- As our data volumes explode, do the old ways still work?
- Can we improve our science by doing better?
- Can we ward off external threats to our data?
- Example: the WIPO legislation

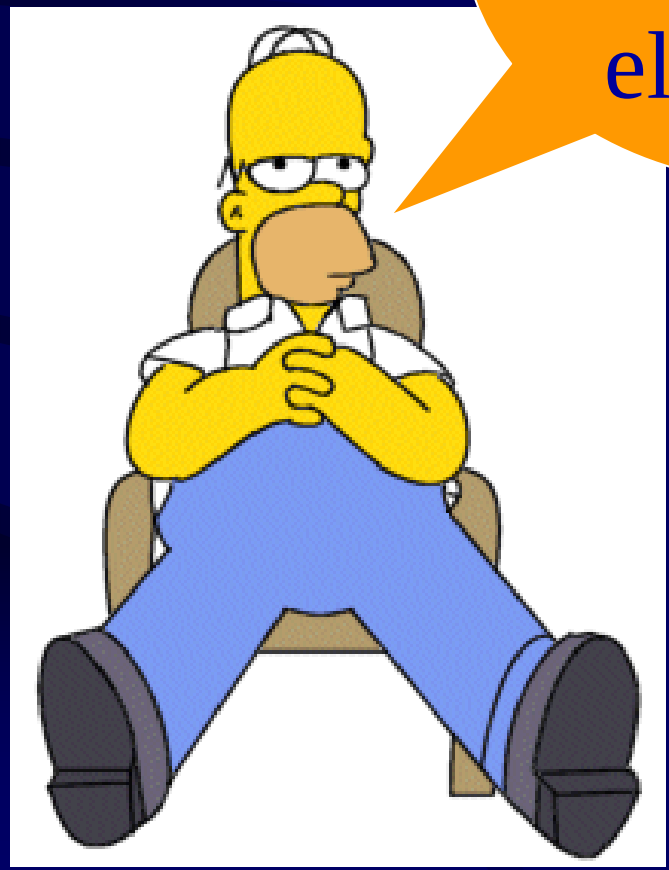

United Nations
United Nations
WIPO
ICSU
International Council of Science
World Intellectual Property Organisation
CODATA
Committee on Data for Science and Technology
IAU
IUGG
etc...
National Representatives

Example: the WIPO proposal

- Protects information (about anything)
- No “fair use” provisions
- You cannot cite someone else’s data without obtaining their permission
- Each paper will need a paper-trail showing rights to cite data
- Our data centres and the VO would probably become unworkable


United Nations
United Nations
WIPO
ICSU
International Council of Science
World Intellectual Property Organisation
CODATA
Committee on Data for Science and Technology
IAU
IUGG
etc...
National Representatives

Challenge: how do you stimulate a discussion about data management?

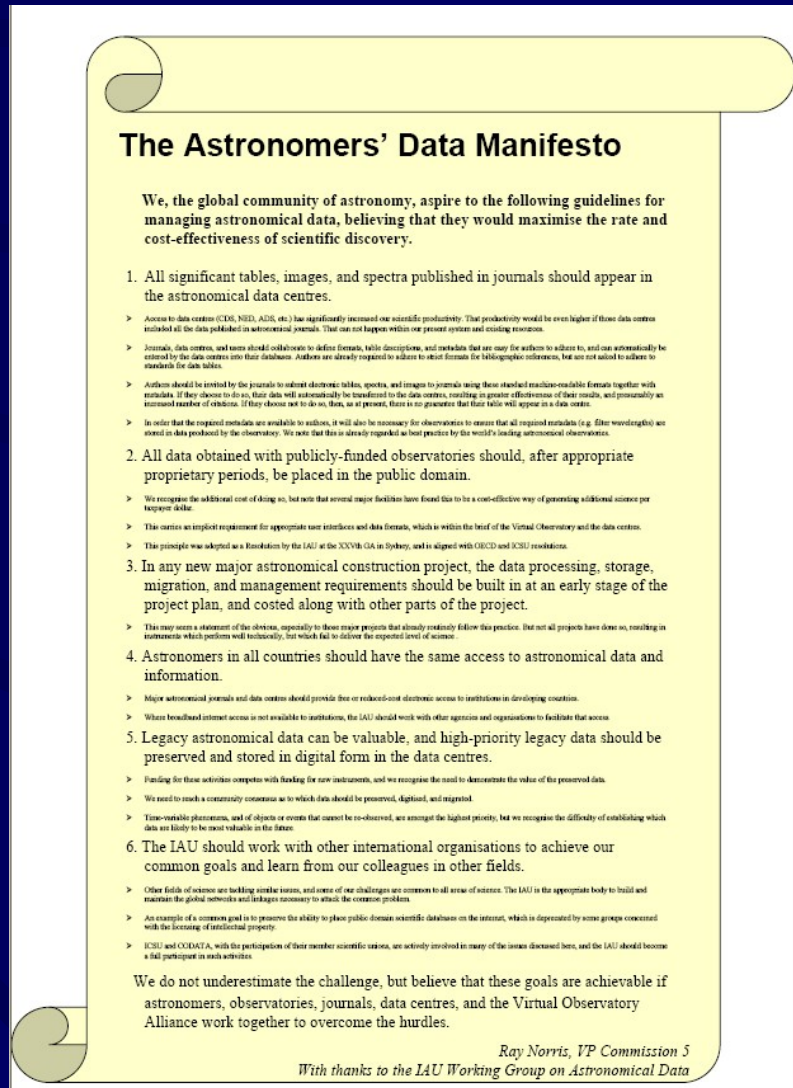


The Astronomers' Data Manifesto

http://www.ivoa.net/twiki/bin/view/Astrodata/AstronomersManifesto

(or just Google on "Astronomers' Data ManifestoAstronomers' Data Manifesto")

We, the global community of astronomy, aspire to the following guidelines for managing astronomical data, believing that this would maximise the rate and cost-effectiveness of scientific discovery…



The Astronomers' Data Manifesto

We, the global community of astronomy, aspire to the following guidelines for managing astronomical data, believing that they would maximise the rate and cost-effectiveness of scientific discovery.

1. All significant tables, images, and spectra published in journals should appear in the astronomical data centres.

- Access to data centres (CDS, NED, ADS, etc.) has significantly increased our scientific productivity. That productivity would be even higher if those data centres included all the data published in astronomical journals. That can not happen within our present system and existing resources.
- Journals, data centres, and users should collaborate to define formats, table descriptions, and metadata that are easy for authors to adhere to, and can automatically be entered by the data centres into their databases. Authors are already required to adhere to strict formats for bibliographic references, but are not asked to adhere to standards for data tables.
- Authors should be invited by the journals to submit electronic tables, spectra, and images to journals using these standard machine-readable formats together with metadata. If they choose to do so, their data will automatically be transferred to the data centres, resulting in greater effectiveness of their results, and presumably an increased number of citations. If they choose not to do so, then, as at present, there is no guarantee that their table will appear in a data centre.
- In order that the required metadata are available to authors, it will also be necessary for observatories to ensure that all required metadata (e.g. filter wavelengths) are stored in data produced by the observatory. We note that this is already regarded as best practice by the world's leading astronomical observatories.

2. All data obtained with publicly-funded observatories should, after appropriate proprietary periods, be placed in the public domain.

- We recognise the additional cost of doing so, but note that several major facilities have found this to be a cost-effective way of generating additional science per taxpayer dollar.
- This carries an implicit requirement for appropriate user interfaces and data formats, which is within the brief of the Virtual Observatory and the data centres.
- This principle was adopted as a Resolution by the IAU at the XXVth GA in Sydney, and is aligned with OECD and ICSU resolutions.

3. In any new major astronomical construction project, the data processing, storage, migration, and management requirements should be built in at an early stage of the project plan, and costed along with other parts of the project.

- This may seem a statement of the obvious, especially to those major projects that already routinely follow this practice. But not all projects have done so, resulting in instruments which perform well technically, but which fail to deliver the expected level of science.

4. Astronomers in all countries should have the same access to astronomical data and information.

- Major astronomical journals and data centres should provide free or reduced-cost electronic access to institutions in developing countries.
- Where broadband internet access is not available to institutions, the IAU should work with other agencies and organizations to facilitate that access.

5. Legacy astronomical data can be valuable, and high-priority legacy data should be preserved and stored in digital form in the data centres.

- Funding for these activities competes with funding for new instruments, and we recognise the need to demonstrate the value of the preserved data.
- We need to reach a community consensus as to which data should be preserved, digitised, and migrated.
- Time-variable phenomena, and of objects or events that cannot be re-observed, are amongst the highest priority, but we recognise the difficulty of establishing which data are likely to be most valuable in the future.

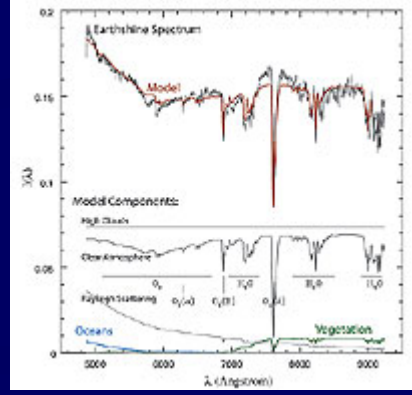
6. The IAU should work with other international organisations to achieve our common goals and learn from our colleagues in other fields.

- Other fields of science are tackling similar issues, and some of our challenges are common to all areas of science. The IAU is the appropriate body to build and maintain the global networks and linkages necessary to attack the common problem.
- An example of a common goal is to preserve the ability to place public domain scientific databases on the internet, which is deprecated by some groups concerned with the licensing of intellectual property.
- ICSU and CODATA, with the participation of their member scientific unions, are actively involved in many of the issues discussed here, and the IAU should become a full participant in such activities.

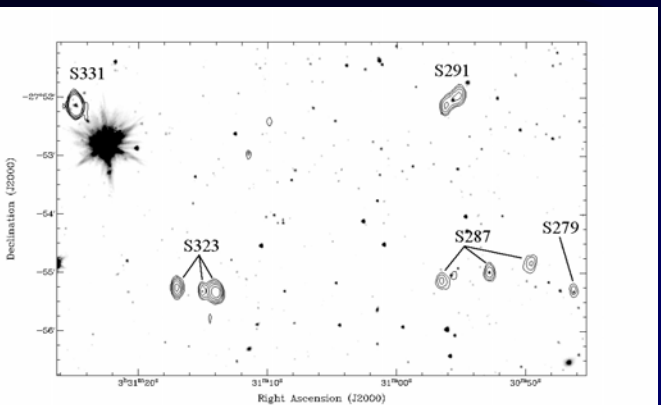
We do not underestimate the challenge, but believe that these goals are achievable if astronomers, observatories, journals, data centres, and the Virtual Observatory Alliance work together to overcome the hurdles.

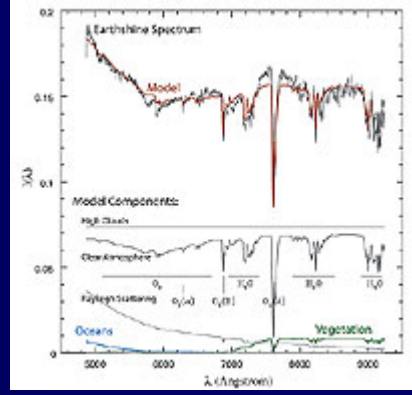
Ray Norris, VP Commission 5

With thanks to the IAU Working Group on Astronomical Data

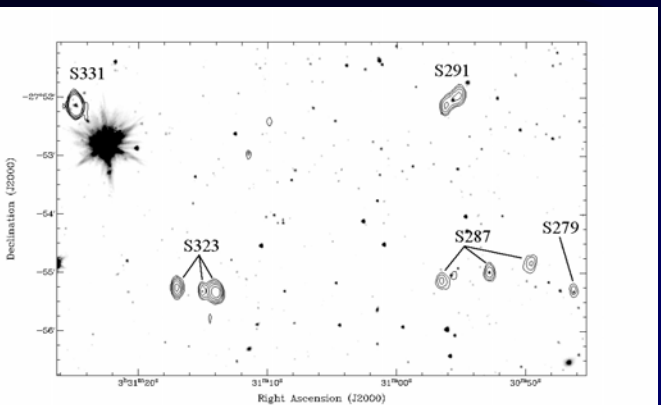


1. All significant tables, images, and spectra published in journals should appear in the astronomical data centres.





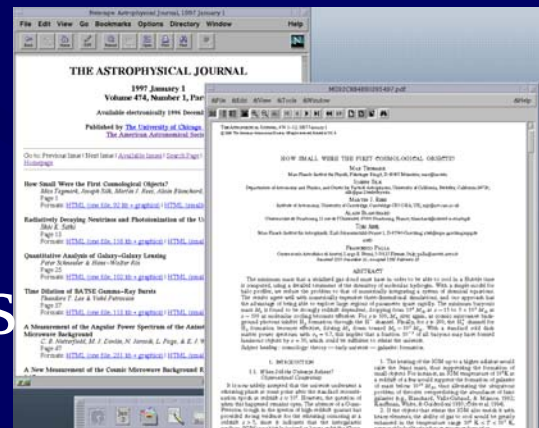
1. All significant tables, images, and spectra published in journals should appear in the astronomical data centres.



Doesn't this happen already?

Journal Data

- Many data published in journals never make it to the data centres
- When they do appear in data centres, they often don't carry the metadata or ontology that enable machine-understanding
  - e.g. plot SED (Spectral Energy Distribution)
- One solution: standards agreed by authors, journals, data centres.
- Would need to be optional
- Guarantees appearance in data centres



2. All data obtained with publicly-funded observatories should, after appropriate proprietary periods, be placed in the public domain.

Problem:

- "Why should I share my data with my competitors?“
- (Because that’s how science works, stupid!)

- This principle endorsed by a resolution at Sydney IAU GA
- Consistent with ICSU recommendations
- OECD Science Ministers have also said they want this

- But: still not yet there!

3. In any new major astronomical construction project, the data processing, storage, migration, and management requirements should be built in at an early stage of the project plan, and costed along with other parts of the project

Isn’t that obvious?

Many new instruments are planned without sufficient planning or funding for data management
(decreasing scientific productivity)

4. Astronomers in all countries should have the same access to astronomical data and information.

We take for granted instant access to literature and databases.
Our colleagues in developing countries still dream of it
(thus disadvantaging them even further)

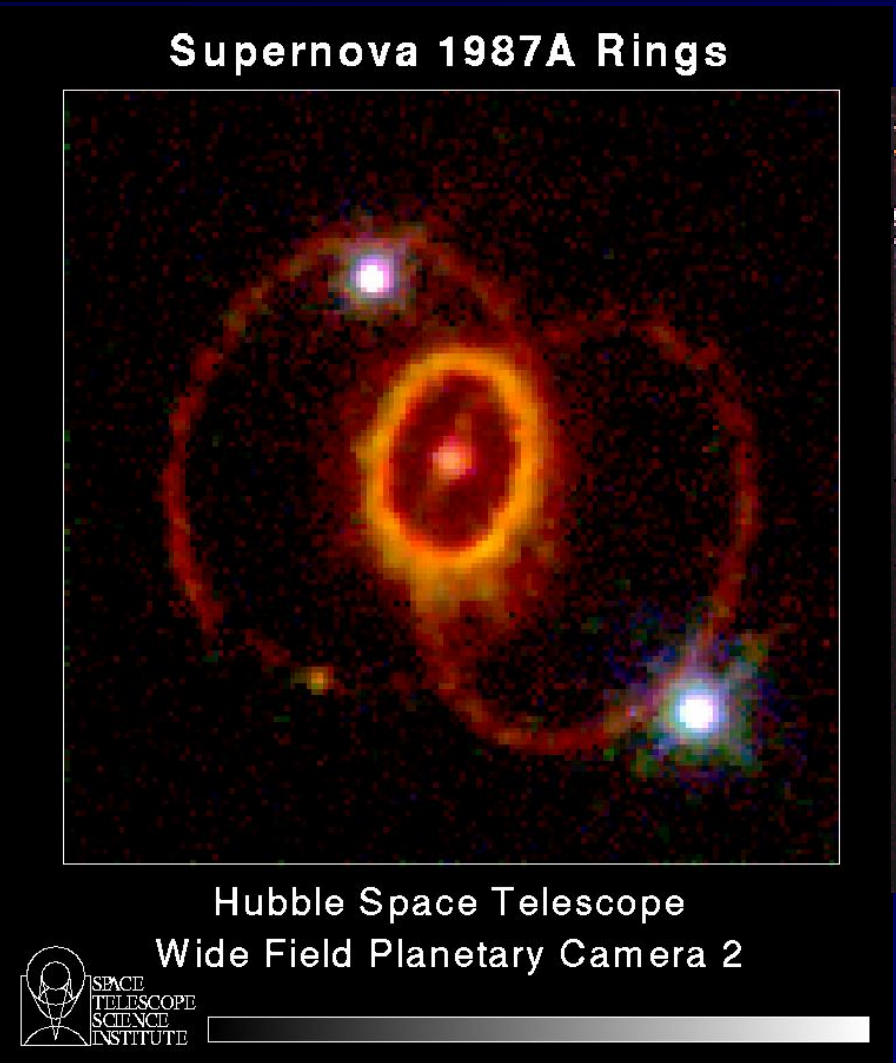
We can’t leave this for someone else to fix!
All scientists have a part to play in this process

5. Legacy astronomical data can be valuable, and high-priority legacy data should be preserved and stored in digital form in the data centres.

How do you prioritise?

E.g. SN1987A

(Closest recorded supernova since invention of telescope)



Challenge: Digitising old data competes for funding with new instruments

Challenge: how to prioritise?

We need to develop standards and guidelines

6. The IAU should work with other international organisations to achieve our common goals and learn from our colleagues in other fields.

- Use bodies such as CODATA to cross-fertilise

Where do we go from here?

- All this will happen only if the astronomical community makes it happen.
- Do data enthusiasts in other disciplines face this same challenge?
- Can we use CODATA as a forum to pool our expertise and success/failure/stories, and learn from each others' experience?