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
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

WIPO

World Intellectual Property Organisation

Committee on Data for Science and Technology

National Representatives

CODATA

IAU

IUGG

etc...

ICSU

International Council of Science

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

Why can’t someone else do it?
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…
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?