



Data at Risk Task Group (DARTG)

Thierry Pauwels

Koninklijke Sterrenwacht van België, Ringlaan, 3, B-1180 Brussel
EMail: Thierry.Pauwels@oma.be,

The Data at Risk Task Group is a task group of CODATA. It is chaired by Elizabeth Griffin (Astrophysicist; IAU Task Force for the Preservation and Digitization of Photographic Plates; Herzberg Institute of Astrophysics, Dominion Astrophysical Observatory), and its first task is, before trying to preserve data at risk, to identify data collections that are at risk. The members are from various disciplines: astrophysics, geology, climatology, botanics, libraries, archives. In a first stage the task group will try to fill the inventory with data collections from the subjects of the task group members.

Definition

We define "data at risk" in this context as scientific data which are not in a format that permits full electronic access to the information which they contain. Such data may be inherently non-digital (e.g. handwritten or photographic), on near-obsolete digital media (such as magnetic tapes) or insufficiently described (lacking meta-data). Some born-digital data can also be considered "at risk" if they cannot be ingested into managed databases because they lack adequate formatting or metadata. Data which are regarded as unuseable tend to be regarded as useless, and then risk being destroyed. Most of the non-electronic data in question pre-date the digital era, and where they complement more modern ones by offering a much longer time-base they are essential, sometimes vital, for studies of long-term trends.

Goals and Objectives

Our overriding goal is to create an Inventory of data that are at risk, and whose unique scientific information is in danger of being lost to posterity. (The Inventory will become the foundation for a Phase II project to design a series of missions to rescue that information.) DARTG will thus accentuate the need to be protective of the scientific content of fragile data, and will illustrate this broader objective by compiling literature describing new science which has emanated from analyses of rescued, historic data. By working through the steps to achieve our Objective, DARTG will demonstrate an approach, process, and practices for building an extensible inventory of scientific data which risk being lost or destroyed and whose information content is therefore seriously endangered.

Procedures

1. Define a set of core metadata properties essential for an inventory.
 - DARi Metadata Schema 1.0 {[PDF](#)}
2. Establish an infrastructure to support inventory data collection and maintenance.
 - The [Data at Risk Inventory](#), built using [Omeka](#), an open-source web publishing platform.
3. Populate the inventory with data at risk in selected target disciplines.
 - Interested parties are encouraged to submit descriptions of their endangered data through the [submission form](#).

Website:

<http://ils.unc.edu/~janeg/dartg/>

If you know of any data collection that is at risk, please have a look at the web site, and contact Thierry Pauwels (thierry.pauwels@oma.be).

[Advanced Search](#)

[Browse Dataset Descriptions](#)

[Submit a Dataset Description](#)

[About](#)

Submit a Dataset Description

Please complete as many of the fields as you can. Put in a "?" if you do not have the requested information.

Please describe your dataset below. [Data at Risk Inventory](#) ▼

1. Research area(s)

2. Title (name associated with the collection)

3. Physical Form of the data (e.g., paper, specimen, photo plate, etc.)

4. Content and context of data, if known (e.g., history, topic, keywords, etc.)

5. Name of current data holder (individual, institution, or organization)

6. Dates associated with the data

7. Size (e.g., 25 specimens, 35mm film, 2 GB, etc.)