Anona C. Earls. Designing a Metadata Scheme for the Documentation of Data Rescue Activities. A Master's Paper for the M.S. in L.S degree. July, 2013. 50 pages. Advisor: Jane Greenberg

Rescuing fragile scientific data is crucial to ensuring that this data can continue to be part of the scientific process. The complementary documentation of data rescue efforts is also essential for ongoing scientific study and future data rescue. This paper reports on a study that examined how best to document and describe data rescue activities. A mixed methods approach included a case study of 20 data rescue initiatives and a content analysis of seven descriptive metadata schemes. The results identified 13 metadata elements that were common across schemes and well suited to describing rescue projects. The findings were used to develop a core metadata scheme for the description of data rescue activities. The study was undertaken as part of the Data-at-Risk Initiative (DARI). Results of the study will help inform future DARI efforts and contribute to the knowledge base on data rescue and project-descriptive metadata.

Headings:

Metadata

Data rescue

Data at risk

Endangered data

Scientific data

DESIGNING A METADATA SCHEME FOR THE DOCUMENTATION OF DATA RESCUE ACTIVITIES

by
Anona C. Earls

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

July 2013

Approved by

_____

Jane Greenberg

Table of Contents

Introduction

Scientific data throughout the world are at risk of being lost forever due to factors such as degradation and improper storage conditions, format obsolescence, and a lack of accompanying metadata. In fact, the majority of research data are not properly preserved or archived. (Thompson, 2011) Yet, historical data very often retain scientific value, and current scientific research draws upon data collected from previous studies for this very reason. (Anderson et al., 2011) As Carver (2012, p. 2) points out, "knowledge of a subject's past conditions helps researchers create new models, forecasts, and theories," and historical data can thus be critical to present day scientific advancement. It is understandable, then, that at-risk data poses a serious concern for the scientific community.

In 2010, the Committee on Data for Science and Technology (CODATA)[1], an interdisciplinary Scientific Committee under the International Council for Science (ICSU), appointed the Data at Risk Task Group (DARTG)[2] to work in the general area of data rescue. (Nordling, 2010). A major goal of DARTG has been to create an inventory of scientific data that is at risk of being lost to posterity. (CODATA/DARTG) The prototype Data-at-Risk Inventory[3] was built using the web publishing platform Omeka and was officially launched via a partnership between DARTG, the University of North Carolina at Chapel Hill's (UNC) Metadata Research Center[4], and ibiblio[5] in early November 2011. (Carver, 2012, p.5) With support from the Fed Ex Global Education

Center, DARTG and the Metadata Research Center together formed a student working group called the Data-at-Risk Initiative (DARI)[6] to further spearhead this effort. Another part of this work was to collect information on data rescue. This secondary task led DARTG and DARI to recognize a need to document the efforts being put forth to rescue endangered datasets in a structured way, and the inventory is now beginning to expand its focus to describe these efforts. However, there currently exists no standard way to document or describe rescue efforts or projects per se. This paper addresses the need for a standard descriptive scheme and explores the core descriptive elements needed to describe data rescue activities. Three guiding questions focused on metadata and data rescue needs, as stated in the Research Questions section of the paper, are investigated to this end.

Purpose

The purpose of this study is to design a core metadata scheme to be used in the description and digital documentation of scientific data rescue activities. While there have been efforts to design metadata schemes for the purpose of describing rescued scientific data, approaches to describing the activities undertaken to rescue the data remain largely unexplored. At present, there exist very few (if any) metadata schemes designed to describe and document the characteristics of a project as an entity. The new scheme will consist of elements that reflect the major defining characteristics of data rescue activities in general. These rescue activities may include both specific projects and broader programs established with the aim of rescuing and preserving scientific data. The

scheme will be directly applied to the description of several data rescue projects, which will be documented online through DARI's rescue mission repository.

Literature Review

Faced with the problem of at-risk data, scientists and information professionals have begun to undertake efforts to rescue these data in danger of being lost forever. Initial efforts must as a matter of course involve data archaeology, or "the process of seeking out, restoring, evaluating, correcting, and interpreting historical data sets." (Levitus, 2002) Once the data have been located and properly cared for in their original form, the data rescue can begin. In WCDMP Report No. 49, the World Meteorological Organization (WMO) defines data rescue as "the ongoing process of preserving all data at risk of being lost due to deterioration of the medium and digitizing current and past data into computer compatible form for easy access." (as cited in Diwakar, Kulkarni, & Talwai, 2008, p. 139) The WMO (WCDMP Report No. 49) adds further specifications for continued preservation: 1) data should be stored as image files onto media that can be regularly renewed (cartridges, CDs, DVDs, etc.); 2) data already in computer-compatible media should be constantly migrated to storage facilities that conform to changing technologies; and, 3) data should be key-entered in a form that can be used for analyses." (as cited in Page, et al., 2004) Documentation through metadata "is integral to this work and essential for measuring and assessing high priority data preservation cases." (Anderson et al., 2011, p. 1)

Today, these rescue practices are being carried out by a broad array of organizations and individuals from a number disciplines. Groups such as CODATA and DARTG have spearheaded rescue efforts at the broadest level, working to illustrate the problem of at-risk data through a multi-national Inventory of at-risk data from across scientific disciplines. The World Meteorological Organization (WMO) and the National Oceanic and Atmospheric Administration (NOAA) provide a driving force behind many current rescue projects and initiatives within the realms of atmospheric, climate, and environmental data. The USGS is another major organization behind data rescue efforts within the geological sciences. Smaller-scale efforts are also being carried out around the globe, in locations as diverse as China, Belgium, Cuba, and France. Even crowdsourcing projects for data rescue are now underway via the Web. Such projects include Old Weather[7], a 'citizen science' project aiming to research historical weather variability; Data Rescue at Home[8], a historical weather data digitization project involving volunteers and weather enthusiasts; and the Canadian Historical Data Typing Project[9], a volunteer effort to type up historical weather data from early Canadian observers.

Rescue Efforts

Many present day data rescue activities fall under one of several broad categories correlating to scientific disciplines – in particular, disciplines in which research often depends on historical data and information indicating changes over time. This section provides an overview of various data rescue efforts in the areas of Meteorology & Climatology, Oceanography, and Astronomy, as well as additional efforts relating to biodiversity and geological data.

Meteorological and Climate Data

Numerous ongoing rescue endeavors exist in the areas of meteorology and climatology. The World Meteorological Organization (WMO) Data Rescue projects and initiatives (DARE) is an overarching program that collaborates on a number of other initiatives, including the Expert Team Data Rescue (ET-DARE), the Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative, the NOAA Climate Database Modernization Programme (CDMP), the International Environmental Data Rescue Organization (IEDRO), and the MEditeranean climate DAta REscue initiative (MEDARE). Each of these efforts is focused on the preservation and digitization of climate data into computer compatible form, defining data rescue as "the ongoing process of preserving all data at risk of being lost due to deterioration of the medium and digitizing current and past data into computer compatible form for easy access." (World Meteorological Organization)

The Pacific-Australia Climate Science and Adaptation Planning Program (PACCSAP), which is a part of the International Climate Change Adaptation Initiative (ICCAI) and operates under the government of Australia's Bureau of Meteorology, conducts data rescue and digitization of climate records as they relate to climate change. Small islands are facing serious challenges with respect to climate change, and those in the Pacific region are no exception. There is very limited specific scientific information available to the countries comprised of these islands, and better scientific knowledge is urgently needed to adapt and plan for their future. PACCSAP is actively working with 15 partner countries and regional stakeholders in efforts to help serve this need through an

examination of past climate trends and variation and by providing regional and national climate projections. (PACCSAP) Among the project's goals are security of national climate records for current and future climate study and applications, more efficient data management, higher quality data, more efficient data rescue and data entry, better data availability and improved data analysis, [and] increased availability of historical data through data entry and data rescue efforts. (Martin et al., 2012)

Another rescue effort within the realm of climatology focuses on the ISCCP B1 data record, which refers to geostationary imagery that has been collected from satellites around the world as a part of the International Satellite Cloud Climatology Project (ISCCP) since 1983. The data are archived at the National Climatic Data Center (NCDC), which is a part of the NOAA. The NCDC began data rescue efforts in 2003, though by this time the data were largely unusable, with issues such as unknown formats/lack of format documentation, a lack of software for reading the data files, and nonexistent read and write routines. (Knapp, 2008) The rescue efforts have resulted in the successful use of some ISCCP BI data in additional scientific work studying tropical cyclones, and may also be used to study topics such as rain fall and cloud cover in the future. (Knapp, 2008)

In their description of solar radiation data rescue at Camagüey, Cuba, Antuña et al. (2006) note that a considerable amount of some countries' climate data remains in the form of paper records, despite an abundance of global-scale digitized information. Because these records are at risk of being lost forever, efforts such as those at Camagüey are underway to digitize them into computer-compatible form. Specifically, the

Camagüey project is focused on rescuing the solar radiation measurements dataset collected there over a period of more than 30 years. Due to a lack of funding for computer equipment, the project has developed a low-cost data rescue plan based on older, out-of-service PCs. Work to rescue the data involves entering the original observations, as well as developing the software to process the observations, controlling quality, and improving the original manual processing in order to engage the complete research cycle. (Antuña et al., 2006) The project is illustrative of a meteorological data rescue project functioning successfully in a less developed country, without support or leadership from an international organization, using scare resources and only basic technology, and relying on local expertise. (Antuña et al., 2006)

Additional efforts to rescue and preserve climate data are being carried out in India. Over the past 80-90 years, the Indian Meteorological Department has taken and recorded observations of on barometric pressure, temperature, and humidity, and transcribed them onto preset paper forms called autographic charts. These charts are now held by the Indian Metropolitan department but are at risk of being lost due to the medium's rapid deterioration. Because analyzing and compiling data from the charts is very time consuming and prone to human error, the rescue project has worked to develop methods of automatic extraction and storage of the endangered atmospheric data through use of image processing tools, with a focus on speed, accuracy and space. The Autographic Chart Data Extraction (ACDE) software has been developed for this purpose. (Diwakar, et al., 2008)

Marine/Oceanic Data

A number of rescue activities are focused on historical data relating to the world's oceans. The Global Oceanographic Data Archeology and Rescue (GODAR) Project is one major endeavor in this area, initiated in 1993 under the UNESCO Intergovernmental Oceanographic Commission (IOC). Before the widespread use of computers, oceanographic data were recorded in the form of manuscripts, data reports, and card index files. As electronic data storage became more prevalent, oceanographic observations were increasingly recorded on magnetic media such as tapes and disks. However, these media have been subject to degradation over time, resulting in the loss of unique data in some instances. (Levitus, 2002) The GODAR project is working to locate and digitize or otherwise copy to modern electronic media historical oceanographic data sets (pre-1992) that are at risk of loss due to media decay, and "to incorporate them into a global, comprehensive, integrated, scientifically quality-controlled database with all data in one uniform format." (Levitus, 2012, p. 46)

The EUR-OCEANS (EURopean research on OCean Ecosystems under Anthropogenic and Natural forcingS) network has also supported several data rescue projects. One such initiative has focused on Southern Ocean data rescue, addressing a range of data from research cruises that sailed between 1925 and 1985. The project has successfully retrieved valuable data on the biological components of the Southern Ocean ecosystem. (EUR-OCEANS & British Antarctic Survey, 2008)

Another effort within the marine sciences is the Coastal and Estuarine Data/Document Archeology and Rescue (CEDAR) initiative, backed by NOAA. CEDAR aims to "collect unpublished data and documents on the marine ecosystem; convert and restore information into electronic and printed form, and distribute it electronically to the scientific community, academia and the public." (Pikula, 2001) As the project's name indicates, this effort is specifically focused on coastal and estuarine related documents and data.

Efforts have also been put forth towards rescuing historical marine data in Belgium. With the financial support of the Belgian Science Policy, the Belgian Marine Data Centre launched a 2–year project aimed at identifying Belgian historical marine data recorded on media at risk and rescuing as many of these data as possible. The project specifically focused on data collected during two multi-disciplinary scientific programs. As with many data rescue projects, these efforts will be particularly helpful in cases where evolution over time is under scientific consideration. (Borremans, 2010)

Astronomical Data

Several important rescue projects are focused on the preservation of astronomical plates and the valuable data recorded on them. One project addressing astronomical plate collection and preservation in China began in 2008. Plates from five observatories were physically relocated to a controlled environment, and efforts are now underway to digitize the plates and represent them via an online catalogue. (CODATA/DARTG) Astronomical images are in process of being preserved at the Royal Observatory of

Belgium as well. Wide-field photographic plate images taken throughout the course of the 20th century, which are in danger due to aging emulsions, are being digitized in an effort to both preserve the data and make the information contained in the images widely accessible. Additionally, the Observatory initiated the UDAPAC project, which entailed acting as host to the endangered plates themselves. (CODATA/DARTG) The Dominion Astrophysical Observatory, a part of the Herzberg Institute of Astrophysics in Canada, is also making progress in astronomical plate preservation. These efforts have included the scanning of archival materials, and updates to data acquisition hardware and software. (CODATA/DARTG)

Biodiversity Data

Many important rescue efforts have been pursued within additional scientific domains. The reBiND project is a complementary effort to direct data rescue. The focus of the project is on developing an efficient and well-documented workflow for rescuing legacy biodiversity data. (Güntsch et al., 2012) The workflow consists of "phases for data transformation into contemporary standards, data validation, storage in a native XML database, and data publishing in international biodiversity networks." (Güntsch et al., 2012, p. 752)

Geological Data

The U.S. Geological Survey (USGS) has sponsored a data rescue program focused broadly on rescuing geological data since 2006. Specifically, this effort aims to preserve

and make accessible legacy USGS scientific records at risk of being lost or in need of greater accessibility.

(CODATA/DARTG) Various projects have been executed under the USGS data rescue program, addressing historical images of Alaskan volcanoes; diagnostic records for wildlife disease; expansion of the USGS Landsat archive; historical files from federal government mineral exploration-assistance programs, 1950-1974; water resources records available only in USGS archives; and historical files from USGS North American bird phenology program (BPP), 1880–1970. (U. S. Geological Survey, 2012)

Metadata for Project Documentation

Four guiding principles for the creation of a new metadata scheme are that it be 1) simple; 2) broadly applicable across a range of projects, approaches, or missions; 3) sufficiently describe the project, approach, or mission; and 4) extensible (implementation takes future growth into consideration). (Anderson et al., 2011) Additionally, it is important that the terms and phrases used to create content in metadata records should reflect appropriate and accepted vocabularies within the corresponding community or institution. (DataONE) This section provides an overview of the seven metadata schemes analyzed for this study, in light of best practice guidelines and the schemes' applicability to the description of projects.

The Dublin Core & Project Description

The Dublin Core Metadata Initiative (DCMI) Metadata Element Set, Version 1.1 (Simple Dublin Core) provides the basis for many of the schemes analyzed in this study and is an

important primary source for basic, broadly applicable metadata elements. This point is corroborated by Mourkoussis et al., who point out that the Dublin Core "is of importance for resource discovery across domains and hence of great relevance to any system proposing information retrieval over the Internet. It is also pertinent to issues of interoperability and information exchange." (2003, p.2) The Digital Library Federation encourages the use of Simple Dublin Core as well, along with other supplemental metadata formats as necessary. In addition to the core element set, the DCMI Metadata Terms provide additional options for descriptive characteristics. Along similar lines, the Goddard Core is a metadata element set developed to support evaluation and resource discovery of project-oriented information across the project libraries of NASA's Goddard Space Flight Center (GSFC). (Hodge, Templeton, & Allen, 2005) It is based on qualified Dublin Core, with extensions. Beyond the GSFC, the scheme's applicability for general use with project-oriented information "has been discussed with other NASA Centers, other U.S. Government science agencies, industry and non-governmental organizations internationally." (Hodge et al., 2005, p.22)

The Archaeology Data Service (ADS) and Digital Antiquity have collaborated on a series of Guides to Good Practice, including information on providing project metadata. Specifically, the Guide defines project-level metadata as that which is recorded at a broad level for an entire project/archive and incorporates descriptive and resource discover metadata. The basic Dublin Core metadata elements form the foundation of the ADS project-level metadata. (ADS/Digital Antiquity) The RSLP Collection Description Schema was designed to describe collections within the Research Support Libraries

Programme (RSLP) and is also based on Dublin Core where possible. (Johnston, 2002) Because this scheme is another used to describe collections, or "an aggregation of physical and/or electronic items," only some of the elements will apply to a project as a whole.

Social Science Data Description

The Data Preservation Alliance for the Social Sciences (Data-PASS) is a voluntary partnership of organizations aiming to archive, catalog and preserve social science research data that is at risk of being lost. As a part of these efforts, Data-PASS has drafted a document outlining the project's metadata requirements. (Data-PASS Project) The Data-PASS scheme focuses on studies and their corresponding data. The ISLE Metadata Initiative (IMDI) has outlined a metadata scheme designed to describe "multi-modal multimedia and written language corpora." (IMDI Part 1) It was designed for the linguistic community, who often needed a more extensive and specialized set of elements than a scheme such as Dublin Core could currently provide. (IMDI Part 1) However, like the Data-PASS scheme, only a few of the IMDI metadata elements are mandatory. Again, this is a specialized scheme for a particular community's use, and it is important to note that it is also designed to describe "corpora" rather than a singular item.

Software Project Description

An examination of Apache Maven, a software project management and comprehension tool, provides guidance on project related metadata. The software itself is based on the concept of a project object model (POM), which is the fundamental unit of work in

Maven: an XML representation of a Maven project. (Apache Maven Project) The Apache Maven guidelines on project metadata outlines the recommended eight metadata elements that should be provided for a project. DOAP (Description of a Project) uses an RDF Schema and XML vocabulary to convey semantic information about open source software projects. Although its focus is specific to software projects, the scheme is useful in that it focuses on the projects as entities. The tool "sbt," used to build within Scala and Java, also outlines some of the basic metadata elements to be used for a project. (sbt Documentation) Like Maven, some of its elements are specific to software but are pertinent in that they relate to the description of a whole project.

Research Questions

1) What are the major descriptive characteristics of known data rescue activities?

2) What existing metadata standards can be applied to describing a project as a whole?

3) What metadata elements are essential for describing data rescue projects in particular?

Research Design

The study was designed as a survey of the current landscape and state of the area of data rescue activity description. A mixed methods approach included a case study to examine 20 data rescue projects and programs, a review of best practices in the application of descriptive metadata, and a content analysis of seven metadata schemes relating to preservation and data description. The initial aim in this approach was to determine what means of description, if any, were already in use and to identify areas where further work

was still needed. The ultimate goal, after accomplishing these objectives, was to produce a core metadata scheme based on a synthesis of the findings.

Methods

Data were collected through a review of the relevant literature in the areas of data rescue and descriptive metadata. The review included literature detailing 20 examples of data rescue activities, best practices in the application of descriptive metadata, and seven metadata schemes relating to preservation and data description. Keyword searches used to gather the literature were done through UNC's Articles+ (Summon), Google, and Google Scholar. Search terms included: "endangered scientific data rescue," "scientific 'endangered data' rescue," "scientific 'data at risk' rescue," "description 'project information,'" "documentation 'project information,'" "archiving 'project information,'" "project description," "project description schema," "project description vocabulary," "project description schema," "metadata for project documentation," and "project documentation."

Analysis

A number of schemes were initially considered for this project's analysis, with seven being deemed most applicable to the study's purpose. The metadata elements comprising each scheme were compiled in a spreadsheet along with their descriptions for analysis. Major descriptive characteristics of each data rescue project or program were compiled in a separate spreadsheet for analysis and comparison. The data were then analyzed to identify common trends in descriptive characteristics. Similar and recurring descriptive

elements within and across the two spreadsheets were collected, compiled, and consolidated to form a set of elements. These elements make up the new proposed core metadata scheme for data rescue activity description. A beta-test implementing the new scheme through the DARI online inventory will provide continued analysis of the efficacy of these results.

Challenges

One initial challenge in the selection of appropriate sources was to distinguish between metadata that would describe a rescue project as a whole versus metadata that would describe the data being rescued. For example, it was difficult to draw the line as to where a description of the scientific subject area of the data should fall, and whether it should be reflected in the description of the project or was instead related to the data itself. In this case, the solution was to include ambiguous descriptive information under the Notes element, since it provides important supplemental information but does not explicitly define the project itself.

Locating metadata for the description of an activity or project as a whole in itself also proved to be a challenging task, as this does not appear to be a major focus among metadata schemes. Even metadata relating to "project documentation" tended to address the documentation of a project's steps and progress, rather than an after the fact description. However, some schemes do touch on projects to some degree, and others such as Dublin Core are general enough to have lent a number of useful elements to the scheme design.

Distinguishing between projects and programs for data rescue also presented a minor challenge, primarily because it added a previously unanticipated layer of complexity to the study of these activities. In the end, it did not greatly affect what was required for description of the activity.

Results

Data Rescue Activities

Current data rescue activities tended to fall into one of a few broad scientific categories, notably climatology, oceanography, and astronomy. This result is consistent with the typical context in which historical data preservation is most urgent and valued, as these disciplines in particular often rely on historical data in order to track changes over time. There were many overlapping descriptive characteristics for the rescue activities analyzed. These had the potential to be grouped into several broader categories, equating to a more minimal set of elements, or divided into more granular categories, forming a more extensive and nuanced element set. This section discusses the major characteristics found to be relevant to the description of data rescue activities according to the nature of the activity, who has carried it out, and when (and where) it has taken place. The section concludes by commenting on the current state of self-documentation on the part of these rescue activities overall.

What is it?

The project or program name was the first descriptive element identified for all activities examined. Some endeavors, such as the MEditerranean climate DAta REscue initiative

(MEDARE), have established titles, while others, such as the Astronomical Plate Collection and Preservation in China project, are referred to informally according to their particular undertaking. Nevertheless, the *Name* or *Title* descriptor was an essential descriptive element for identification purposes. Many of the data rescue endeavors examined also have some form of documentation online, so *Link* was another important aspect to the identification of projects overall. A few of the projects examined had other associated resources – for example, the USGS Data Rescue Program's Toolkit for Managing Electronic Records – which also have the potential to be noted as supplemental description.

A basic description of a rescue activity is also pertinent to its documentation in nearly all cases. Many projects were introduced in their literature according to the driving goal or focus of the endeavor; for example, to "preserve and digitize climate data into computer compatible form." (WMO DARE) Thus, *Project Focus* or *Goal* served in most cases as an accurate and workable summary description of the rescue activity as a whole. An additional aspect of most projects is their approach to the data rescue and methods for accomplishing the rescue and preservation efforts, such as the controlled storage conditions, online metadata cataloging, and plate digitization implemented by the Astronomical Plate project in China. Some rescue efforts also detailed specific software, machinery, or other tools used in preservation and rescue efforts, as in the case of the Camagüey, Cuba project's low-cost methods of data transfer via older computer operating systems. *Approach/Method(s)* on the whole was an important descriptive aspect of the rescue activities examined. A more in depth background description was also an

important component of many of the activities examined, given the complex history and makeup of some endeavors. For many, such as the GODAR project, the *Background Description* element also provided an opportunity for the data being rescued to be noted and discussed.

Throughout the analysis, it became useful to classify the rescue activities according to type of endeavor. While some of the endeavors were projects focused on a particular set of data or materials for rescue, others were broader programs established to undertake data rescue within a scientific discipline and oftentimes had various projects being carried out under the overarching initiative. The *Project/Program* distinction had little effect on the necessities of activity description overall, as both tended to have essentially the same major characteristics, but it was highly useful in terms of understanding the activity being examined and how it related to other endeavors – for example, the Southern Ocean Data Rescue project exists under the umbrella of the EUR-OCEANS Data Rescue program. Identification of an activity's relationship to any others, whether broader or narrower in focus, provided additional layers of description when applicable.

Who is involved?

There are generally many different individuals and organizational bodies involved with a data rescue endeavor. Several distinct roles emerged from this study's examination of rescue activities; however, there was a tendency for roles to blend, vary by project, or remain unclear based on the information available Common questions included: 'Was the project's affiliated organization also it's initiator?' and 'Did that organization also fund

the project?' By necessity each project or program did have an *Initiator*: one or more individuals or an organization behind putting the project into action. In most cases there was also the need for a *Sponsor* or source of funding (and in some cases the source changed over time for a given project, as in the case of IEDRO). Frequently, the initiating body and the funding body were one and the same, especially in the case of a large program such as WMO DARE. In addition to those undertaking the project, there were many instances of secondary involvement, such as those contributing data or materials, or other project partners or collaborating bodies not connected with undertaking the work directly.

When & where?

Dates are another useful and often essential characteristic of data rescue activity description. All current and former efforts necessarily have a date of initiation, at least roughly, which can be used to place them in a temporal context. A majority of the projects examined were ongoing and had no end date or time limit established; however, the date of completion is an equally important temporal marker and was noted when applicable. Geographic location was a distinguishing characteristic for some rescue activities, such as the solar radiation data rescue project in Camagüey, Cuba. However, it was not an applicable descriptor for many others, such as WMO DARE or the GODAR project, which are global efforts and are focused on the type of data rather than a specific region.

1. Name of Project/Program	10. Others Associated (& how so)
2. Link(s)	11. Project Focus/Goal (brief description)
3. Associated Resources	12. Approach/Method(s)
4. Project Initiator/Affiliation	13. Tools (software, devices, etc.)
5. Funding Source /Sponsor (current, past)	14. Location (geographic)
6. Program or Project?	15. Project Length (or ongoing)
7. Projects Carried Out (if under Program)	16. Date Started
8. Project Contributors: Data/Materials	17. Date Ended (or ongoing)
9. Project Contributors: Partners/Collaborators	18. Description (background)

Table 1. Descriptive categories initially determined via analysis of defining aspects of data rescue activities.

At present, data rescue activities primarily are being documented and described through scholarly articles and/or an online presence. While some rescue activities are well documented, there are others for which information is scarce. The level of documentation and description tended to correlate with the scale of the activity and the reach of the entity backing it. The GODAR project, for example, which is affiliated with NOAA, is widely publicized and well documented, whereas much less has been written regarding projects such as the Royal Observatory of Belgium's astronomical plate rescue, which operates unto itself. The end-goal data repositories of some efforts may also be considered a part of their "documentation" – for example, the Astronomical Plate project in China's plan to create an online catalog of digitized plates.

Although many data rescue activities are well documented on an individual level, there remains a need for documentation in the form of a unified overview, such as a directory

or database, of rescue activities. The need for such a directory was acutely apparent throughout this study as challenges initially arose in discovering and correctly defining many of the activities surveyed. This type of compilation would also likely serve as a useful resource for the scientific community as a whole. Such an overview would allow for a greater opportunity to view the whole picture of data rescue efforts and provide important insight into the nature of these efforts via a synthesis of their descriptions. This added perspective could contribute to increased awareness and use of rescued data, as well as further progress with respect to data rescue initiatives in the future.

Metadata Schemes

There are very few, if any, metadata schemes designed to describe a project, activity, or endeavor as a whole. Within the science realm, schemes have more commonly been designed to describe the data that resulted from a study or project, rather than the project itself. Outside of the sciences, schemes similarly tend to describe materials or information as opposed to a project or activity. This section first discusses each of the schemes analyzed and their elements in light of their relevance to data rescue project description, and follows up by discussing the elements relevant to data rescue project description that were found to recur across schemes and guidelines.

Scheme Relevance: Elements and Purpose

Seven metadata schemes were selected for final analysis and synthesis into a proposed set of elements for the description of data rescue projects. Included in the analysis were: the Archaeology Data Service guidelines, the Data-PASS metadata requirements, the DOAP

schema, the Dublin Core Element Set v. 1.1 & the DCMI Terms, the Goddard Core, the IMDI scheme, and the RSLP Collection Description Schema. Several best practices and guidelines for descriptive metadata were also considered, and the essential descriptive categories outlined were compared with the element sets from the schemes analyzed.

The Dublin Core Element Set, Version 1.1 was the most broadly applicable scheme and seven of its 15 elements were determined to be relevant to data rescue activity description. These were *Title, Creator, Subject, Description, Publisher, Contributor,* and *Date*. When the extended DCMI Terms were considered alongside these, the Dublin Core offered 10 applicable element categories. These included *spatial* to represent Geographic Location; *source, isPartOf,* or *isVersionOf* to represent Associated Resources; and *identifier* or *bibliographicCitation* to represent a citation or URL.

The Data-PASS metadata requirements outlined 9 elements relevant to data rescue activity description. The Data-PASS scheme focuses on studies and their corresponding data, which differs slightly from project description and is broader insofar as it includes data description. However, a majority of the elements (*Title, Author, Description, Identifier, Publication Date, Subject, Data Sources, Time Period, Collection Date,* and *Geographic Coverage*) have the potential to apply to the description of projects as well. There could also be potential to draw upon this scheme in the future, if rescue activities and the corresponding rescued data were ever to be described together.

The Goddard Core scheme is made up of a significant number of elements that may easily be applied to rescue project documentation as well, including *Title, Creator, Creator.Organization, Subject.MissionsProjects, Subject.Competency, Subject.Instrument, Subject.BusinessPurpose, Subject.Industries, Subject.Uncontrolled, Description, Date.Created, Identifier.URL, Source,* and *Coverage.Spatial*. Combined into more general categories, these elements address eight applicable element types for rescue activity description, including Title/Name, Description, Creator/Author, Dates, Geographic Location, Associated Resources, Subject Keywords, and Unique ID. Because this scheme was created for information pertaining to the Goddard Space Flight Center, it has some granular elements that differ from the specifics of data rescue endeavors. Yet, it's underlying design goal to support evaluation and resource discovery of project-oriented information has resulted in many aspects that are applicable to this study.

The Archaeology Data Service Guidelines have seven elements relevant to data rescue activities. Because the Archaeology Data Service scheme was designed to apply to archaeological data, some elements are not applicable to scientific data or to projects as a whole. However, a number of the elements (*Project Title, Description, Subject, Creators, Contributors, Source, Dates*) are applicable to projects across disciplines, including data rescue efforts.

Six elements from the RSLP Collection Description Schema were found to be relevant to rescue activities. The schema was designed to describe collections, and thus differs in focus from the goal of rescue activity description. However, some similarity lies in that

both collections and projects could be considered to be types of "bodies" of information. In particular, several of the scheme's General Attributes (*Title, Identifier, Description, Note, Location, Concept*) are applicable in a project description context.

The IMDI scheme offered five applicable elements. This scheme, which was designed to describe multimedia and bodies of written language within linguistics, differs significantly in focus from that of data rescue activity description. Yet, much like the RSLP scheme, similarities can be found in terms of the description of a "body." Many of the elements (including *Name, Title, Date, Creator (Name, Contact, Description), Project (Name, ID, Description)*) are similar in a general sense to what is needed for the description of a data rescue project.

Lastly, the four elements from the DOAP scheme were applicable to rescue activity description. Despite its specified focus on software projects, examination of the scheme and comparison to the description requirements of other projects proved fruitful to some degree. DOAP Properties in particular have some potentially useful elements for rescue description, including *name, homepage, created, short description, description, and helper*.

Apache Maven, a software project management and comprehension tool, offers guidelines on project related metadata, including eight standard elements that should be provided. Although some of these elements are specific to software projects, some (*Project name, Project URL, Description of the project, Group and Artifact ID*) are more

generalizable and add to the knowledge of what is necessary for useful project description. The tool "sbt," which is used to build within Scala and Java, outlines some of the basic metadata elements to be used for a project. Similar to Maven, some of the elements in this framework are software specific, but many (*Name, Organization, Homepage, StartYear, Description*) provide general guidance in standard project description.

The Massachusetts Institute of Technology (MIT) Guide on Data Management and Publishing also provides a list of general aspects that should be documented for a project or project data. These include *Title, Creator, Dates, Location, Data Processing,* and *Sources*. Similarly, the Digital Library Federation (DLF) Best Practices for Shareable Metadata outlines a list of Recommendations for Classes of Data Elements, including *Titles, Names, Dates, Subjects/Topics,* and *Geographic Places.*

The Goddard Core, Dublin Core, Data-PASS requirements, and ADS guidelines offered the four most applicable schemes for the description of data rescue activities. These were the most developed towards describing a project or program, as evidenced by their variety of relevant elements. Notably, two of the schemes are based on Dublin Core where possible, and even Data-PASS uses Dublin Core as a starting point of comparison. As a major standard in the field of metadata, Dublin Core is certainly useful in many ways for data rescue description. At the same time, it is important to note that Dublin Core is a broad scheme not specifically designed for the purpose of project description. In

some cases therefore, description with Dublin Core alone lacks some of the finer nuances that are desirable in detailed description of data rescue.

Element Relevance: Occurrence Across Schemes

The analysis of metadata schemes resulted in a suggested 11 elements suitable for the description of data rescue activities. The Title/Name element and the Description element were both found in every scheme examined, indicating that these are essential elements to include in the proposed scheme as well. The Dates element was also found to be integral across the schemes, being identified in all but one. The Creators element, the Subject/Topics/Keywords element, and the Unique Identifier element were all found in five of the seven schemes, indicating that these elements would be highly useful to the proposed scheme. Similarly, the Geographic Coverage/Spatial element was found in four of the seven schemes, as was the Source or Associated Resources element, indicating a strong potential usage. The Notes element, the Contributors element, and the URL/Citation element were all found in two of the seven schemes. These were therefore judged to be less essential, but still useful overall to the proposed scheme.

Descriptive Elements	Archaeology Data Service Guidelines	Data-PASS	DOAP	Dublin Core, v. 1.1	DCMI-TERMS	Goddard Core	IMDI (ISLE Metadata Initiative)	RSLP Collection Description Schema
Title/Name	x	x	x	x	x	x	x	x
Description	x	x	x	x	x	x	x	x
Methods								
Notes		x						x
Creator/Author	x	x		x	x	x	x	
Sponsor								
Contributor	x			x	x			
Dates	x	x	x	x	x	x	x	
Geographic Location		x			x	x		x
Associated Resources	x	x			x	x		
URL/citation			x		x			
Subject Keywords	x	x		x	x	x		x
Unique ID		x		x	x	x	x	x

Table 2. Occurrence of potential elements for data rescue activity description across schemes and guidelines.

Proposed Scheme

The final proposed scheme was developed via a synthesis of the results from analysis of rescue activities and metadata schemes (Table 3). The major descriptive categories found to be relevant to rescue activities were compared with the elements derived from analysis of descriptive metadata schemes. In some cases, the element or category from one list had been broken down into several elements in the other list in order to provide more detail. For example, the Contributors element derived from the metadata scheme analysis corresponded to four different sub-categories derived from the rescue activity analysis (Table 3). Such situations were considered carefully in terms of clarity and distinction provided by each element and level of granularity desired for the final scheme. In many cases this led to the formation of a workable middle ground; for example, in the case of

the Contributors category, two final elements resulted (Table 3). This approach was guided in part by the Dublin Core "Dumb-Down" principle[10], which indicates that, when using a qualified element, "the qualifier may be dropped and the remaining value of the element should still be a term that is useful for discovery." (DCMI) Thus, two elements such as Date Started and End Date became simply Dates in the final scheme (Table 3). Each element or category was similarly examined, and the two lists were merged to form the final proposed element set for data rescue activities.

Metadata Elements Analysis	**Rescue Activity Analysis**	**Final Proposal**
Title/Name	Project Name	Title
Description	Project Focus/Goal [brief description]	Description
	Approach/Method(s) to Rescue (incl. software, devices, etc.)	Methods
Notes	Other Description [project background/history]	Notes
Creators	Project Initiator/Primary Affiliation	Creator
Contributors	Funding Source/Sponsor (current & past?)	Sponsor
	Project Contributors: Partners/Collaborators	Contributor
	Project Contributors: Data/Materials	
	Others Associated? (and how so)	
Dates	Date Started	Dates
	End Date (or ongoing)	
Geographic Coverage/Spatial	Location (if applicable)	Location
Source	Associated Projects (if applicable)?	Associated
URL	link/url	URL
Subject/Topic/Keywords		Subject.Keywords
Project ID (optional)		Project ID (optional)

Table 3. Comparison of categories & elements from analyses of rescue activities and metadata schemes, and synthesis to form final scheme proposal.

Final Element Set

The proposed core metadata scheme for the description of data rescue activities consists of 13 metadata elements: *Title*, *Description*, *Methods*, *Notes*, *Creator*, *Sponsor*, *Contributor*, *Dates*, *Location*, *Associated*, *URL*, *Keywords*, and *Project ID.* Many of the element descriptions are based on those from the Archaeology Data Service scheme, which were found to be both detailed and oriented towards the description of whole projects (Table 4).

Metadata Element Name	**Element Description**
Title*	The title (and any alternatives) for the project.
Description	A brief summary of the main focus, goals, aims, and/or objectives of the project.
Methods	A brief summary of the approach, methods, techniques, and/or processes (including tools, software, etc.) being used for the data rescue.
Notes	Other details pertinent to the project, such as background information or project history.
Creator*	Individual(s) or organization(s) who initiated and have overseen the data rescue effort. May include contact information.
Sponsor	Individual(s) or organization(s) who have contributed financially or otherwise endorsed the project.
Contributor	Other individual(s) or organization(s) who have contributed to the project; for example, project partners/collaborators (physical or intellectual efforts), contributors of data/materials, etc.
Dates*	Dates indicating when the rescue effort was initiated and when it was completed (if applicable). May also include important milestones or other significant dates associated with the project.

Location	Location where the project was/is being carried out (if applicable).
Associated	Any other important projects or work (in particular, other data rescue initiatives) associated with this project, or upon which this project has been built.
URL	A link to the project website and/or online documentation of the project.
Keywords	Keywords indicating subject content of the project.
Project ID	A unique ID# assigned to the project by the repository (optional).
*Indicates required element	

Table 4. Final proposed metadata scheme for the description of data rescue activities.

The Title and Description elements were deemed essential to the scheme, as these are the primary means of identification for each rescue activity. Similarly, the Dates and Creator elements designate unique and highly useful information for identifying a rescue activity. Although it is possible to describe a data rescue activity without any additional elements, the Sponsor and Contributor elements also designate unique and important information. The Methods element, although not essential to the basic description of an activity, provides additional details regarding the activity's execution, which could be especially useful as a future data rescue reference point. The Geographic Location element did not apply in all instances but was an important characteristic of some rescue efforts.

Although the Notes element was found in just two of the seven schemes examined, it was found to be useful for all of the rescue activity descriptions. Rescue efforts are often multi-dimensional and have background information such as the type of data being rescued or how the project was developed. Similarly, the URL element was useful for providing additional primary source information on the activities. The Associated

Resources element is also a useful way to provide additional details on activity as it relates to the broader picture of data rescue. The Unique ID element is mainly helpful to the repository or inventory that maintains the descriptions, since it is an arbitrary identification, but it was judged to be potentially useful for the purpose of cataloging rescue activities. The Keywords element is also essential for cataloging an activity and is a useful way to provide an at-a-glance summary.

While each of these elements distinguishes unique and useful information, the majority will be considered optional for initial implementation through the DARI inventory (*Title*, *Creator*, and *Dates* will be required). This distinction is made in order to prioritize the most essential descriptive elements and maintain simplicity, with the particular goal of encouraging contributions to the Inventory through ease of submission. The proposed scheme is currently being implemented through a data entry form on the DARI website (Appendix A).

Benefits

This study contributes to the body of knowledge relating to project documentation and archiving and adds to the realm of existing metadata schemes. Specifically, the results of the study have the potential to aid in the documentation and description of vital data rescue efforts worldwide. The proposed metadata scheme will be directly applied to the description of data rescue activities being documented in the DARI online repository, making this effort more robust. The scheme may also be applied and built upon by other bodies engaging in data rescue documentation. It is hoped that by emphasizing the

significance of data rescue work, the results of this study will help to support further work in this area.

Limitations

This study was conducted within a limited time frame. Although the study attempted to analyze the most applicable descriptive metadata schemes, many other schemes exist that were not a part of the study and could provide further insights. Similarly, although the study made an effort to examine as many data rescue activities as possible, there are still more to be examined. These too could provide further insights that have not yet been incorporated into the creation of the proposed scheme. Additionally, the work to design a new metadata scheme entailed a certain amount of subjectivity. As a result, there is the possibility of bias on the part of the researcher. It is likely that personal preference and circumstance could dictate variations to the scheme. Future study could benefit from a survey of researchers in the areas addressed.

Conclusion

Documentation of efforts to preserve at-risk data is essential for ongoing scientific study and future data preservation efforts. This study has investigated how best to document and describe data rescue efforts through analysis of 20 data rescue activities and seven descriptive metadata schemes, and a review of descriptive metadata best practices. Similar and recurring descriptive elements across schemes and rescue activities were collected, compiled, and consolidated to form a set of 13 final elements. These 13

elements were used to develop a core metadata scheme for the description of data rescue activities.

This study was undertaken as part of the Data-at-Risk Initiative (DARI), which will implement the proposed core scheme through an online data entry template designed to gather information on rescue activities for the online Data-at-Risk Inventory. The study will help to inform future DARI efforts.and contributes to the bodies of knowledge relating to data rescue and project-descriptive metadata. The resulting metadata scheme may also be applied to the documentation and description of data rescue activities worldwide.

Future work in this area could potentially benefit from a study of additional project- and data- oriented metadata schemes in the areas of description and preservation, as well as a review of additional data rescue activities. This study may also be built upon by implementing focus groups and surveys to test the workability of the proposed schema. Additional future consideration should be given to the relationship between data rescue activities and rescued data in the context of documentation; that is, how these two important facets might be best connected within a repository to represent the larger picture of data-at-risk and rescue.

Acknowledgements

The author would like to thank Jane Greenberg, University of North Carolina at Chapel Hill; William L. Anderson, University of Texas at Austin; and all contributing members of the Data-at-Risk Initiative for their valuable input throughout this project.

Notes

[1] http://www.codata.org/

[2] http://ils.unc.edu/~janeg/dartg/

[3] http://ibiblio.org/data-at-risk/

[4] http://ils.unc.edu/mrc/

[5] http://www.ibiblio.org/

[6] http://ils.unc.edu/mrc/dari-2/

[7] http://www.oldweather.org/

[8] http://www.data-rescue-at-home.org/

[9] https://sites.google.com/site/historicalclimatedata/canadian-historical-data-typing-project

[10] http://dublincore.org/resources/faq/#dumbdown

References

Anderson, W., Faundeen, J., Greenberg, J., & Taylor, F. (2011). Metadata for data rescue and data at risk: Ensuring long-term preservation and adding value to scientific and technical data. PV2011, 17 November 2011, Toulouse, France. Proceedings paper retrieved 2013-04-26 from http://ils.unc.edu/~janeg/dartg/final-PV2011-060-DARTG.pdf. Conference presentation retrieved April 26, 2012 from http://www.slideshare.net/2ghouls/metadata-for-data-rescue-and-data-at-risk.

Antuña, J. C., Fonte, A., Estevan, R., Barja, B., Acea, R., & Antuña Jr., J. C. (2008). Solar radiation data rescue at Camagüey, Cuba. *Bulletin of the American Meteorological Society*, 89, 1507–1511.

Apache Maven Project. Frequently asked questions about project metadata. Retrieved June 23, 2013 from http://maven.apache.org/project-faq.html

Archaeology Data Service / Digital Antiquity. Guides to good practice: Project metadata. Retrieved June 23, 2013 from http://guides.archaeologydataservice.ac.uk/g2gp/CreateData_1-2

Atmospheric Circulation Reconstructions over the Earth (ACRE). Retrieved April 26, 2013 from http://www.met-acre.org/

Borremans, C., Scory, S., De Cauwer, K., & Devolder, M. (2010). Rescuing historical marine data and making them accessible: the case of data collected during the first years of modern oceanography in Belgium. *International Marine Data and Information*

Systems, IMDIS 2010 Id = 36 - 2. Data Circulation and Services in ocean science – Poster.

Brunet, M., & Jones, P. (2011). Data rescue initiatives: bringing historical climate data into the 21st century. *Climate Research*, *47*, 29–40.

Caldwell, Patrick. (2003). NOAA support for global sea level data rescue: International sea level data rescue. *Earth System Monitor*, 14(1), 1-6.

Carver, N., Collins, K., Greenberg, J., Sinclair, J., Thompson, C., Veitch, M., Anderson, W. (2011). Identifying endangered data: A case study supporting inventory design and implementation. ASIST 2011, October 9-13, 2011, New Orleans, LA, USA.

Committee on Data for Science and Technology (CODATA) Data at Risk Task Group (DARTG). Data-at-Risk Inventory. Retrieved from http://www.ibiblio.org/data-at-risk/

Commonwealth of Australia, Bureau of Meteorology. Pacific-Australia Climate Science and Adaptation Planning Program (PACCSAP). Retrieved June 28, 2013 from http://www.bom.gov.au/climate/pacific/

Data-at-Risk Inventory. (n.d.). Retrieved April 26, 2012 from http://www.ibiblio.org/data-at-risk/

DataONE Metadata Best Practices. Retrieved April 26, 2013 from http://www.dataone.org/best-practices/metadata

Description of a Project (DOAP) vocabulary. Retrieved June 23, 2013 from http://xml.mfd-consult.dk/ws/2003/01/rdfs/?rdfs=http%3A%2F%2Fusefulinc.com%2Fns%2Fdoap%23

Digital Preservation Alliance for the Social Sciences (Data-PASS) Project. (2007). Data-PASS Metadata Requirements (version 1.2, approved 12/15/2007). Retrieved April 26, 2013 from http://www.data-pass.org/sites/default/files/metadata.pdf

Digital Library Federation Best Practices for Shareable Metadata. Retrieved April 26, 2013 from http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/ShareableMetadataPublic

Diwakar, A. S., Kulkarni, A. A., & Talwai, P. (2008). Rescue and preservation of climate data by extraction and digitization from autographic weather charts using image processing tools. 8th WSEAS International Conference on SIGNAL, SPEECH and IMAGE PROCESSING (SSIP '08) Santander, Cantabria, Spain, September 23-25, 2008. (p.139-143.)

Downs, R. R. (2009). Managing Risks to Scientific Data. Prepared for Presentation to the NYU/IBM Workshop on Managing Data Risk: Acquisition, Processing, Retention and Governance, New York University New York, NY April 24, 2009. Retrieved April 26, 2012 from http://w4.stern.nyu.edu/emplibrary/Downs-ManagingRisksSciData.pdf

Dublin Core Metadata Element Set, Version 1.1. Retrieved July 5, 2013 from http://dublincore.org/documents/dces/

Dublin Core Metadata Initiative Metadata Terms. Retrieved April 26, 2013 from http://dublincore.org/documents/dcmi-terms

Dupigny-Giroux, L-A., Ross, T. F., Elms, J. D., Truesdell, R., & Doty, S. R. (2007). NOAA's climate database modernization program: Rescuing, archiving, and digitizing history. *Bulletin of the American Meteorological Society*, 88, 1015–1017.

EUR-OCEANS Knowledge Transfer Unit, British Antarctic Survey. (2008). Southern Ocean data rescue: Extending time-series into the past.

Griffin, R. E. (2005). Rescuing and recovering lost or endangered data. CODATA Data Science Journal, 4, 21- 26. doi:10.2481/dsj.4.21.

Güntsch, A., Fichtmüller, D., Kirchhoff, A., & Berendsohn, W. G. (2012). Efficient rescue of threatened biodiversity data using reBiND workflows. *Plant Biosystems – An International Journal Dealing with all Aspects of Plant Biology: Official Journal of the Societa Botanica Italiana, 146(4),* 752-755.

Hodge, G., Templeton, C., & Allen, R. (2005). A Metadata Element Set for Project Documentation. *Science & Technology Libraries, 25:4*, 5-23.

IMDI (ISLE Metadata Initiative). (2003). Part 1 Metadata elements for session descriptions, version 3.0.4. Retrieved June 23, 2013 from http://www.mpi.nl/IMDI/documents/Proposals/IMDI_MetaData_3.0.4.pdf

International Council for Science: Committee on Data for Science and Technology. (2011). CODATA Data At Risk Task Group (DARTG). Data Rescue Efforts. Retrieved April 26, 2012 from http://ils.unc.edu/~janeg/dartg/datarescue.html

International Environmental Data Rescue Organization. Retrieved on June 28, 2013 from http://www.iedro.org/

Johnston, P. (2002). A standard for collection-level description: the RSLP Collection Description Model and Schema; Collection-Level Description & NOF-digitise projects. NOF-digitise programme seminar, London, 22 February 2002. Retrieved June 23, 2013 from *www.ukoln.ac.uk/cd-focus/presentations/stdrslp/stdrslp.ppt*

Knapp, K. R. (2008). Scientific data stewardship of International Satellite Cloud Climatology Project B1 global geostationary observations. *Journal of Applied Remote Sensing*, Vol. 2, 023548.

Levites, S. (2002). The *GODAR* and *WOD* Projects: Past, present, and future. *GODAR-WESTPAC* Workshop Tokyo, Japan 5 March, 2002. p.1-8.

Levitus, S. (2012). The UNESCO-IOC-IODE "Global Oceanographic Data Archeology and Rescue" (GODAR) Project and "World Ocean Database" Projects. *Data Science Journal*, Vol. 11, p. 46-71.

Martin, D., Hutchinson, R., & Howard, A. (2012). Pacific climate change science: PACCSAP – data rescue and digitising. Retrieved June 28, 2013 from http://www.colloque.ird.fr/icshmo-2012/documents/presentations-all/S15-martin-105.pdf

Massachusetts Institute of Technology Libraries. Data management and publishing: Documentation and Metadata. Retrieved July 4, 3013 from http://libraries.mit.edu/guides/subjects/data-management/metadata.html

MEditerranean climate DAta REscue (MEDARE). Retrieved June 28, 2013 from http://www.omm.urv.cat/MEDARE/index.html

Mourkoussis, N., White, M., Patel, M., Chmielewski, J., & Walczak, K. (? ? ?) AMS – Metadata for Cultural Exhibitions using Virtual Reality.

Murillo, A. P., Carver, N., Greenberg, J., Robertson, D. W., Thompson, C. A., & Anderson, W. (2012). Data At Risk Initiative: Scientists' perceptions of endangered data and data reuse. CODATA Abstract for 2012 conference CODATA 2012, Taipei, Taiwan.

Murillo, A. P., Thompson, C. A., Carver, N., Robertson, W. D., Greenberg, J., & Anderson, W. L. (2012). The Data-at-Risk Initiative: Analyzing the current state of

endangered scientific data. *ASIST 2012,* October 28-31, 2012, Baltimore, MD. Retrieved April 26, 2012 from http://ils.unc.edu/~janeg/dartg/AM12ProceedingsFormat_DARI_final_v5.pdf

Nordling, L. (2010). Researchers launch hunt for endangered data: Global effort will catalogue information languishing in drawers and basements. *Nature, 468,* 17. doi:10.1038/468017a.

Page, C. M., et al. (2004). Data Rescue in the Southeast Asia and South Pacific Region: Challenges and Opportunities. *Bulletin of the American Meteorological Society, 85,* 1483–1489.

Pikula, L., Cantillo, A., & Hale, K. (2001). CEDAR: Data and document rescue coastal and estuarine data/document archeology and rescue. (p.146-147.)

RSLP Collection Description Schema. Retrieved April 26, 2013 from http://www.ukoln.ac.uk/metadata/rslp/schema/

sbt Documentation. Project metadata. Retrieved April 26, 2013 from http://www.scala-sbt.org/release/docs/Howto/metadata.html

Schmidt, C. (2007). DOAP A Matic. Retrieved June 23, 2013 from http://crschmidt.net/semweb/doapamatic/

Tan, L. S., Burton, S., Crouthamel, R., van Engelen, A., Hutchinson, R., Nicodemus, L., Peterson, T. C., & Rahimzadeh, F. (2004). Guidelines on Climate Data Rescue. World Meteorological Organization.

Thompson, C. A., Carver, N., Collins, K., Sinclair, J., & Veitch, M. (2011). Supporting scientists in data archiving: emerging roles for information professionals. *ASIS&T SIG*

2011. Retrieved June 21, 2013 from http://ils.unc.edu/~janeg/dartg/DARi_ASIST_SIG_Poster100611_FNL.pdf

U. S. Geological Survey. (2012). Preserving science for the ages – USGS data rescue. *Fact Sheet 2012–3078.*

U.S. Geological Service: "Create a Rescue Request", Retrieved on April 26, 2012 from http://eros.usgs.gov/government/archive_rescue/archive_request.php

World Climate Data and Monitoring Programme (WCDMP). Project Plan for Data Rescue of Climate and Hydrological Data. Retrieved April 26, 2013 from http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP49_Annex2.pdf

World Meteorological Organization. Data Rescue projects and initiatives (DARE). Retrieved June 28, 2013 from http://www.wmo.int/pages/prog/wcp/wcdmp/CDM_2.php

World Meteorological Organization. Expert Team Data Rescue (ET-DARE). Retrieved June 28, 2013 from http://www.wmo.int/pages/prog/wcp/ccl/opace/opace1/et-dare.php

Appendix A

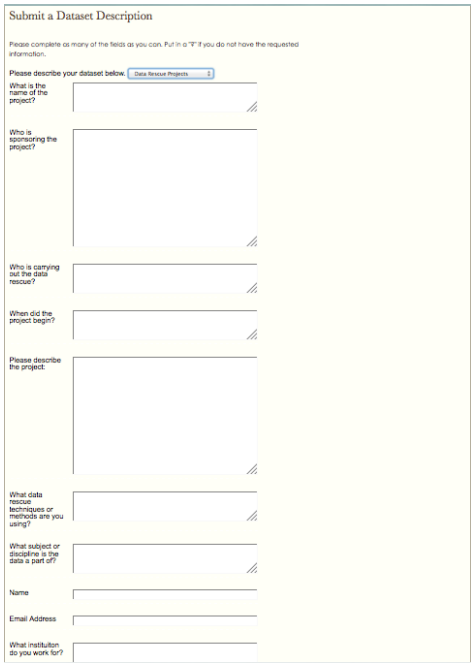


Figure 1. Screenshot of the prototype entry form for contributors of data rescue project information on the DARI website.

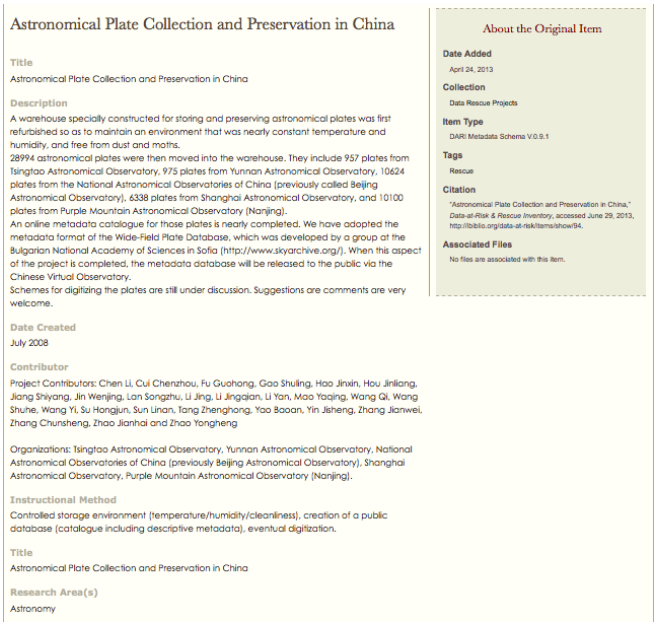


Figure 2. Screenshot of the documentation of a data rescue project on the DARI website, utilizing the metadata scheme for the description of data rescue activities.

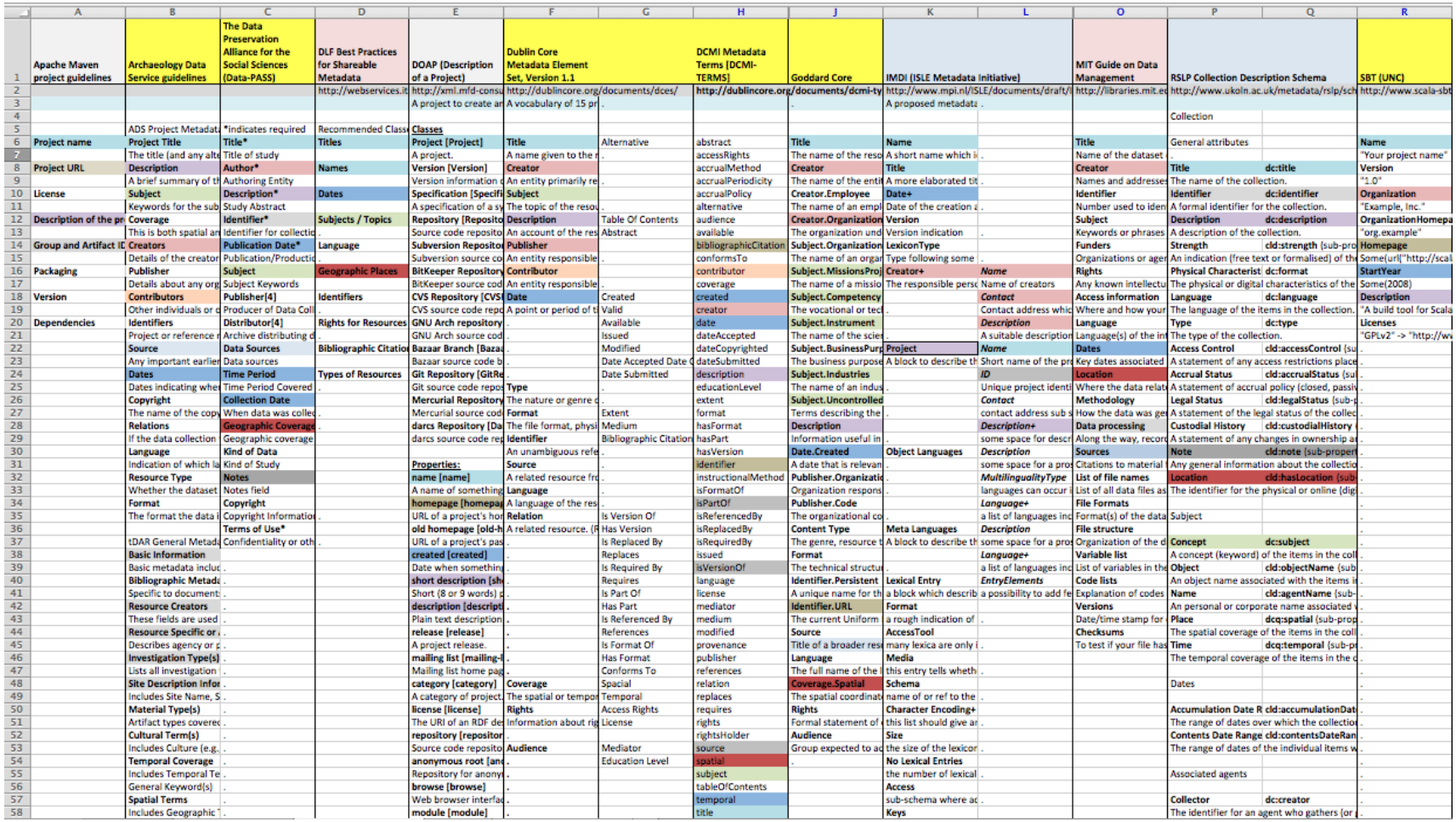
Appendix B
Comparison and Overview of Data Rescue Activities

	A	B	C	D	E	F	G	H	I	J	K	L
1	Name	Astronomical Plate Collection and Preservation Project	Dominion Astrophysical Observatory Astronomical Plate Preservation	reBiND Project	Rescue and Preservation of Climate Data by Extraction and Digitization from Autographic Weather Charts Using Image Processing Tools	Coastal and Estuarine Data/Document Archeology and Rescue (CEDAR)	French NMS	UNESCO-IOC-IODE "Global Oceanographic Data Archeology and Rescue" (GODAR) Project [incl. "World Ocean Database" (WOD)]	Historical Marine Data: Belgium	ISCCP B1 Data Rescue: National Climatic Data Center: NOAA	Pacific Australia Climate Change Science and Adaptation Planning Program (PACCSAP)	UDAPAC (Preserving Astronomic Images at the Royal Observatory Belgium)
2	Link(s)	http://ils.unc.edu/~janeg/dartg/datar	http://ils.unc.edu/~janeg/dartg/datar	http://ils.unc.edu/~janeg/dartg/datar .pdf		www.aoml.noaa.gov/general/lib/CED	http://climatheque.meteo.fr/	http://www.nodc.noaa.gov/General/	http	http://www.ncdc.noaa.gov/oa/rsad/i..ppt		http://ils.unc.edu/~janeg/dartg/da
3	Associated Resources	?	?	?	?	?	?	?	?	?	?	?
4	Project Initiator/Affiliation	?	Dominion Astrophysical Observatory (DAO)	Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM)	India Meteorological Department (IMD)	The NOAA/NODC/Library and Information Science Division (LSD) and the N O M a t i o n a l Ocean	Climatology department (DCLIM) in Toulouse	National Oceanographic Data Center (NODC) and World Data Center-Silver Spring for Oceanography	Belgian Marine Data Centre	NOAA	?	Royal Observatory of Belgium
5	Funding Source (Sponsor) - current?	?	?	LIS-program of DFG - Deutsche Forschungsgemeinschaft (German Research Foundation)	?	South Florida Ecosystem Restoration Prediction and Modeling Program	?	?	Belgian Science Policy	?	?	none yet
6	Program or Specific Project?	project	project	workflow initiative	project	project	project	project	project	project	?	project
7	Projects Carried Out (if under Program)?	n/a	?	?	n/a	n/a	n/a	n/a	n/a	n/a	?	n/a
8	Project Contributors: Data/Materials	Tsingtao Astronomical Observatory, Yunnan Astronomical Observatory, National Astronomical Observatories of China (previously Beijing	Dominion Astrophysical Observatory (DAO)	?	India Meteorological Department (IMD)	?	National Climatological Database (BOCLIM)	?	"Project Sea" (1970-1976); "Concerted Research Actions" (1977-1982)	?	?	Royal Observatory of Belgium, on outside source (so far)
9	Project Contributors: Partners/Collaborators	Chen Li, Cui Chenzhou, Fu Guohong, Gao Shuling, Hao Jinxin, Hou Jinliang, Jiang Shiyang, Jin Wenjing, Lan Songzhu, Li Jing, Li Jingqian, Li Yan,	?	?	?	NOAA	ACRE	?	?	?	?	?
10	Others Associated (and how so)	Chinese Virtual Observatory	Herzberg Institute of Astrophysics (HIA)	CODATA Data at Risk Task Group, Freie UniversitÃ¤t Berlin	?	?	?	managed by a Project Leader: Sydney Levitus, Director of the WOC Oceanography, Silver Spring, USA; International Oceanographic Data	?	?	?	Belgian Federal Space Policy, seve Belgian Federal Scientific Institute
11	Project Focus/Goal [brief description]	Astronomical Plate Preservation	Preservation of astronomical data recorded on photographic plates.	To develop cost-efficient workflows for rescuing legacy biodiversity databases. To develop an efficient and well- documented workflow for	Extraction and digitization from autographic weather charts using image processing tools	Coastal and estuarine data and documents: To collect unpublished data and documents on the marine ecosystem; convert and restore	Climate data preservation Digitization Quality control	GODAR: Locating (archaeology) and digitizing or copying to modern electronic media (rescuing) historical (pre-1992) oceanographic data that	To identify the Belgian historical marine data recorded on media at risk ["Data Archaeology"] and to "rescue" as many of these data as	recovery and use of ISCCP B1 data	• The security of National climate records for current and future climate study and applications • More efficient data management	Host collections of endangered photographic plates.
12	Approach/Method(s)	Temperature & humidity controlled warehouse; online metadata cagalogue; plate digitization.	Digitization [scanning] of archival material (plates/spectra) via POS microdensitometer	reBiND Workflows: transforming data stored in outdated database systems into well-documented,	digitization - image processing...	digitization, etc	Digitization Quality control	digitization, online sharing	• locate the data at risk of being lost, • track and rebuild the corresponding meta-information,	Data are collected at Satellite Processing Centers (SPCs) worldwide.	• Digitisingofpaperrecords • Qualitycontrolofexistingkeyeddata	Air-conditioned archive room // Digitization of photographic plate (images) via high precision scanne
13	Tools (software, devices, etc.)			BioCASE, eXist-db, Xquery	Autographic Chart Data Extraction (ACDE) software	?	?	?	?	?	CIDE database	
14	Location	China?	Dominion Astrophysical Observatory (DAO), Canada	n/a	India	South Florida	France	n/a	Belgium	n/a	n/a	Royal Observatory of Belgium
15	Project Length	3+ years	ongoing	3 years	?	?	1994-2003, 2003-present	ongoing	2 years	?	?	ongoing
16	Started	Jul-08	1990s?	?	?	?	1994	1992-1993	?	2003	?	10+ years ago
17	Ended?	?	ongoing	?	?	?	ongoing	ongoing	?	ongoing	?	ongoing
18	Description	A warehouse specially constructed for storing and preserving astronomical plates was first refurbished so as to maintain an environment that was nearly constant temperature and humidity, and free from dust and moths. 28994 astronomical plates were then moved into the warehouse. They include 957 plates from Tsingtao Astronomical Observatory, 975 plates from Yunnan Astronomical Observatory, 10624 plates from the National Astronomical Observatories of China (previously called Beijing Astronomical Observatory), 6338 plates from Shanghai Astronomical Observatory, and 10100 plates from	The Astrophysical Observatory (DAO), now also known through the title Herzberg Institute of Astrophysics (HIA), of which it is part, is Canada's largest [though not oldest] optical astronomical observatory. Dating from 1918, it grew from Ottawa astronomer J.S. Plaskett's desire for a site which experienced better observing conditions than were routine at the Dominion Observatory in Ottawa, and which would be equipped with the largest telescope then known - a reflector of diameter 72 inches, or 1.8 metres - where its users could better carry out forefront research. While the environs of Victoria in BC were not of the quality of those of	The Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM, http://www.bgbm.org) based at Freie UniversitÃ¤t Berlin (FUB) provides a combination of international collaborative efforts and scientific production in both systematic research and biodiversity informatics. The BGBM is one of the leading natural history research institutions in Europe and maintains extensive scientific collections of herbarium specimens (about 3.5 million), one of the world's largest living plants collections, as well as the most complete botanical library in Germany. It early specialised in the development and implementation of protocols, data	The procedure will be applied to the huge number of charts which have been recorded for the last 80- 90 years.	CEDAR is focused on coastal and estuarine data and documents. CEDAR differs from simple document scanning in that scientific selection, review and editing of the final product is necessary to insure clarity and completeness.		The "Global Oceanographic Data Archaeology and Rescue" project was initiated in 1993 under the auspices of the UNESCO Intergovernmental Oceanographic Commission (IOC). The project has the goal of locating (archaeology) and digitizing or copying to modern electronic media (rescuing) historical (pre-1992) oceanographic data that exist in manuscript or electronic media form that are at risk of loss due to media decay. The IOC "World Ocean Database" project initiated in 2001 focuses on encouraging international data exchange for the post-1991 period and the development of regional atlases. The GODAR project was established by	Belgium has a long oceanographic data collection history. However, a lot of these data are still only available on media that are not suitable for electronic data processing. Recently, the Belgian Marine Data Centre launched, with the financial support of the Belgian Science Policy, a 2-year project aiming at identifying the Belgian historical marine data recorded on media at risk ("Data Archaeology") and to "rescue" as many of these data as possible ("Data Rescue"). That work will contribute to safeguard and make available an old dataset particularly important for any marine or climate research where evolution over time is under	History Data collection for the International Satellite Cloud Climatology Project (ISCCP) began in July 1983. Since then data from 28 satellites have been incorporated into the ISCCP B1 data record. Data collection Data are collected at Satellite Processing Centers (SPCs) around the world. The centers receive full resolution data for the satellite for which they are responsible to provide data. The data are sampled to approximately ~10km and 3-hourly resolution and sent to the ISCCP Central Archive (NCDC) for storage. The SPCs further subsample the data to approximately 30km		At the Royal Observatory of Belgiu we have a collection of ~25,000 wi field photographic plates (images), taken in the course of the 20th century. They constitute a unique record of the past appearances of the sky, and enable the study of transient phenomena as well as th study of phenomena for which a long time-base is needed. These photographic plates are in danger because of the ageing of th emulsions. Most are stored in plac that are difficult to access, and are thus not easily exploitable scientifically. Some have been stor under incorrect conditions and has lost their emulsions, and therewith

(**Comparison and Overview of Data Rescue Activities**, cont.)

M	N	O	P	Q	R	S	T	U
NARA: Preserving United States Government Records of the Past, Present and Future	**Solar Radiation Data: Camagüey, Cuba**	**EUR-OCEANS Data Rescue**	**USGS Data Rescue Program**	**World Meteorological Organization (WMO) Data Rescue projects and initiatives (DARE), including the Expert Team Data Rescue (ET-DARE)**	**Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative**	**NOAA Climate Database Modernization Programme (CDMP)**	**International Environmental Data Rescue Organization (IEDRO)**	**MEditeranean climate DAta REscu initiative (MEDARE)**
http://ils.unc.edu/~janeg/dartg/datar	www.lidar.camaguey.cu/sdr/index.ht	http://www.antarctica.ac.uk/bas_res	http://ils.unc.edu/~janeg/dartg/datar	http://www.wmo.int/pages/prog/wcp	http://www.met-acre.org/	http://www.ncdc.noaa.gov/oa/climat	http://www.iedro.org/	http://www.omm.urv.cat/MEDARE
?	.pdf	?	ookit for Managing Electronic Record	http://www.wmo.int/pages/prog/wcp	?	?	?	?
National Archives and Records Administration (NARA)	Camagüey Lidar Station, Camagüey Meteorological Centre, Camagüey, Cuba	EUR-OCEANS	USGS	WMO	WMO	NOAA, WMO	Dr. Richard Crouthamel, WMO	WMO
?	Territorial Meteorological Research Program Grant 0901007, from the Camagüey Chapter of the Ministry of	EUR-OCEANS	USGS	WMO	the University of Southern Queensland in Australia; the Met Office Hadley Centre (MOHC) in the	?	private donations, some federal funds (and initially, U.S. donations to the World Meteorological	?
program	project	program	program	program	program	program	program	program
?	n/a	**Southern Ocean Data Rescue** (http://www.antarctica.ac.uk/bas_re search/data/access/esodap/)	**Volcanoes; Diagnostic Records for Wildlife Disease; Expansion of USGS Landsat Archive; Historical Files from Federal Government Mineral**	**Asia and South Pacific Region, International Surface Temperature Initiative (ISTI) Climate of the Carpatian Region**	n/a	n/a	n/a	n/a
?	Camagüey Lidar Station, Camagüey Meteorological Centre, Camagüey, Cuba	?	?	?	?	?	National Meteorological Services of Kenya, Malawi, Mozambique, Niger, Senegal, and Zambia.	?
?	n/a?	?	?	?	the University of Southern Queensland in Australia; the Met Office Hadley Centre (MOHC) in the UK; the US National Oceanic and	?	World Meteorological Organization, NOAA, and the weather services of many other countries	?
United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), National Aeronautics and Space	Roberto Acea; Universidad Complutense of Madrid; Dr. Ricardo Garcia-Herrera and his team	?	?	?	World Meteorological Organization Commission for Climatology, Global Observing Systems Information Centre	?	Dr. Sharon Nicholson (Florida State University), National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service's	?
Developments and implementation of long-term preservation and access strategies of scientific electronic records; identify and determine high-	rescuing the solar radiation measurements dataset collected at Camagüey, Cuba	?	Preserve and make accessible legacy USGS science records -- address data sets at risk of loss due to obsolescence of media or format,	**Climate and Hydrological Data**	undertakes and facilitates the recovery of historical instrumental surface terrestrial and marine global weather observations to underpin	preserve and digitize climate data into computer compatible form	Locating and rescuing historic weather observations in Africa; environmental data rescue and digitization.	preserve and digitize climate dat into computer compatible form
?	Step one: Transferred the original handwritten records to digital form. Software was designed and created	?	Digital preservation, inventorying, boxing and sending records to proper Federal records storage at a	?	linking international meteorological organisations & data rescue infrastructures to facilitate the	?	Comuters, digital cameras, copy stands, and software; data imaging and digitization.	?
	FORTRAN 77 for discontinued PCs - runs on the original MS-DOS and several emulated MS-DOS versions	?	?	?	?	?		?
n/a	Camagüey, Cuba	?	n/a	n/a	n/a	?	USA, Kenya, Malawi, Mozambique, Niger, Senegal, and Zambia	?
ongoing	ongoing	?	ongoing	ongoing	ongoing	ongoing	ongoing	ongoing
?	?	?	2006	1979	?	?	2004 officially (roots in 2000)	?
ongoing	ongoing	?	ongoing	ongoing	ongoing	ongoing	ongoing	ongoing
The U.S. National Archives and Records Administration (NARA) is the Federal government agency responsible for preserving and making available to the public the records created in the course of business conducted by the United States Federal government. The National Archives was established in 1934, but its major holdings date back to 1775. They capture the sweep of the past: slave ship manifests and the Emancipation Proclamation; captured German records and the Japanese surrender documents from World War II; journals of polar expeditions and photographs of Dust Bowl farmers; Indian treaties making transitory	Engaging the complete research cycle is fundamental to Cuban science philosophy, and thus to the Camagüey Lidar Station team, so we have also gone further than just rescuing the original observations. The project also has involved developing the software for processing the observations, controlling quality, and improving the original manual processing. tion Algorithm. This algo- rithm calculates the solar zenith and azimuth angle with uncertainties equal to ±0.0003° from the years -2000 to 6000 CE. The rest of the tables and plots for manual interpolation of the		The U.S. Geological Survey (USGS) is a science organization that provides impartial information on the health of our ecosystems and environment, the natural hazards that threaten us, the natural resources we rely on, the impacts of climate and land-use change, and the core science systems that help us provide timely, relevant, and useable information. As the United States' largest water, earth, and biological science and civilian mapping agency, the USGS collects, monitors, analyzes, and provides scientific understanding about natural resource conditions, issues, and problems. The diversity of our scientific expertise enables us to carry out large-scale, multi-	The initial WMO Data Rescue (DARE) project launched in 1979 was aiming at: -assisting countries in the management, preservation and use of climatic data over their own territories. - commiting to microfilm and microfiche, and eventually to digital media the original historical written manuscript records which in many cases were in danger of deteriorating and of being lost. ---------- -- From 1979 to 1997, the WMO/Belgium-funded Data Rescue projects assisted more than 40 African countries in preserving their climate data, at least on microfilms or microfiches.	The international Atmospheric Circulation Reconstructions over the Earth (ACRE) project is collaborating with a number of international data rescue projects and activities.	The CDMP Programme is a partnership between NCDC and private industry to image and key paper and microfilm records and to make them available on the Web to members of the climate and environmental research community. NCDC also provides funding under CDMP for other NOAA agencies with a need to place images online and/or digitize data.	The data rescued by the International Environmental Data Rescue Organization (IEDRO) enables the meteorological and scientific communities to provide more accurate severe weather forecasting and to understand climate change. This knowledge offers the world community a greater ability to more accurately predict long-range weather patterns. In 2000, Dr. Sharon Nicholson representing Florida State University approached the National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service's International Activities Office (IAO) to set up a project for locating and rescuing historic weather	The MEditerranean climate DAta REscue (MEDARE) project is an initiative, born under the auspice o the World Meteorological Organization, with the main objective is being to develop, consolidate and progress climate data and metadata rescue activitie across the Greater Mediterranean Region (GMR).

Appendix C
Comparison and Overview of Metadata Schemes, Guidelines, & Best Practices



	A	B	C	D	E	F	G	H	J	K	L	O	P	Q	R
1	Apache Maven project guidelines	Archaeology Data Service guidelines	The Data Preservation Alliance for the Social Sciences (Data-PASS)	DLF Best Practices for Shareable Metadata	DOAP (Description of a Project)	Dublin Core Metadata Element Set, Version 1.1		DCMI Metadata Terms [DCMI-TERMS]	Goddard Core	IMDI (ISLE Metadata Initiative)		MIT Guide on Data Management	RSLP Collection Description Schema		SBT (UNC)
2				http://webservices.it	http://xml.mfd-consu	http://dublincore.org/documents/dces/		http://dublincore.org/documents/dcmi-ty	http://www.mpi.nl/ISLE/documents/draft/I			http://libraries.mit.ed	http://www.ukoln.ac.uk/metadata/rslp/sch		http://www.scala-sbt
3					A project to create an	A vocabulary of 15 pr	.	.	.	A proposed metadata	.				
4													Collection		
5		ADS Project Metadat	*indicates required	Recommended Class	Classes										
6	Project name	Project Title	Title*	Titles	Project [Project]	Title	Alternative	abstract	Title	Name		Title	General attributes		Name
7		The title (and any alte	Title of study		A project.	A name given to the r	.	accessRights	The name of the reso	A short name which i	.	Name of the dataset	.		"Your project name"
8	Project URL	Description	Author*	Names	Version [Version]	Creator	.	accrualMethod	Creator	Title		Creator	Title	dc:title	Version
9		A brief summary of th	Authoring Entity		Version information c	An entity primarily re	.	accrualPeriodicity	The name of the entit	A more elaborated tit	.	Names and addresses	The name of the collection.		"1.0"
10	License	Subject	Description*	Dates	Specification [Specifi	Subject	.	accrualPolicy	Creator.Employee	Date+		Identifier	Identifier	dc:identifier	Organization
11		Keywords for the sub	Study Abstract		A specification of a sy	The topic of the resou	.	alternative	The name of an empl	Date of the creation a	.	Number used to iden	A formal identifier for the collection.		"Example, Inc."
12	Description of the pr	Coverage	Identifier*	Subjects / Topics	Repository [Reposito	Description	Table Of Contents	audience	Creator.Organization	Version		Subject	Description	dc:description	OrganizationHomepa
13		This is both spatial an	Identifier for collectio	.	Source code repositor	An account of the res	Abstract	available	The organization und	Version indication	.	Keywords or phrases	A description of the collection.		"org.example"
14	Group and Artifact ID	Creators	Publication Date*	Language	Subversion Repositor	Publisher		bibliographicCitation	Subject.Organization	LexiconType		Funders	Strength	cld:strength (sub-pro	Homepage
15		Details of the creator	Publication/Productio	.	Subversion source co	An entity responsible	.	conformsTo	The name of an organ	Type following some	.	Organizations or ager	An indication (free text or formalised) of the		Some(url["http://scal
16	Packaging	Publisher	Subject	Geographic Places	BitKeeper Repository	Contributor		contributor	Subject.MissionsProj	Creator+	*Name*	Rights	Physical Characterist	dc:format	StartYear
17		Details about any org	Subject Keywords		BitKeeper source cod	An entity responsible	.	coverage	The name of a missio	The responsible perso	Name of creators	Any known intellectu	The physical or digital characteristics of the		Some(2008)
18	Version	Contributors	Publisher[4]	Identifiers	CVS Repository [CVS	Date	Created	created	Subject.Competency		*Contact*	Access information	Language	dc:language	Description
19		Other individuals or c	Producer of Data Coll	.	CVS source code repo	A point or period of t	Valid	creator	The vocational or tec	.	Contact address whic	Where and how your	The language of the items in the collection.		"A build tool for Scala
20	Dependencies	Identifiers	Distributor[4]	Rights for Resources	GNU Arch repository	.	Available	date	Subject.Instrument	.	*Description*	Language	Type	dc:type	Licenses
21		Project or reference r	Archive distributing d	.	GNU Arch source cod	.	Issued	dateAccepted	The name of the scie	.	A suitable description	Language(s) of the int	The type of the collection.		"GPLv2" -> "http://w
22		Source	Data Sources	Bibliographic Citatio	Bazaar Branch [Bazaa	.	Modified	dateCopyrighted	Subject.BusinessPurp	Project	*Name*	Dates	Access Control	cld:accessControl (su	.
23		Any important earlier	Data sources		Bazaar source code b	.	Date Accepted Date C	dateSubmitted	The business purpose	A block to describe th	Short name of the pr	Key dates associated	A statement of any access restrictions place		.
24		Dates	Time Period	Types of Resources	Git Repository [GitRe	.	Date Submitted	description	Subject.Industries		*ID*	Location	Accrual Status	cld:accrualStatus (su	.
25		Dates indicating whe	Time Period Covered	.	Git source code repos	Type	.	educationLevel	The name of an indus	.	Unique project identi	Where the data relate	A statement of accrual policy (closed, passiv		.
26		Copyright	Collection Date		Mercurial Repository	The nature or genre c	.	extent	Subject.Uncontrolled		*Contact*	Methodology	Legal Status	cld:legalStatus (sub-p	.
27		The name of the copy	When data was collec	.	Mercurial source cod	Format	Extent	format	Terms describing the	.	contact address sub s	How the data was ger	A statement of the legal status of the collec		.
28		Relations	Geographic Coverage	.	darcs Repository [Da	The file format, physi	Medium	hasFormat	Description		*Description+*	Data processing	Custodial History	cld:custodialHistory	.
29		If the data collection	Geographic coverage		darcs source code rep	Identifier	Bibliographic Citation	hasPart	Information useful in	.	some space for descr	Along the way, record	A statement of any changes in ownership a		.
30		Language	Kind of Data			An unambiguous refe	.	hasVersion	Date.Created	Object Languages	*Description*	Sources	Note	cld:note (sub-propert	.
31		Indication of which la	Kind of Study		Properties:	Source	.	identifier	A date that is relevan	.	some space for a pros	Citations to material	Any general information about the collectio		.
32		Resource Type	Notes		name [name]	A related resource fr	.	instructionalMethod	Publisher.Organizatio	.	*MultilingualityType*	List of file names	Location	cld:hasLocation (sub-	.
33		Whether the dataset	Notes field		A name of something	Language	.	isFormatOf	Organization respons	.	languages can occur i	List of all data files as	The identifier for the physical or online [dig		.
34		Format	Copyright		homepage [homepag	A language of the res	.	isPartOf	Publisher.Code		*Language+*	File Formats			.
35		The format the data i	Copyright Information	.	URL of a project's hor	Relation	Is Version Of	isReferencedBy	The organizational co	.	a list of languages inc	Format(s) of the data	Subject		.
36			Terms of Use*		old homepage [old-h	A related resource. [F	Has Version	isReplacedBy	Content Type	Meta Languages	*Description*	File structure			.
37		tDAR General Metada	Confidentiality or oth	.	URL of a project's pas	.	Is Replaced By	isRequiredBy	The genre, resource t	A block to describe th	some space for a pro	Organization of the d	Concept	dc:subject	.
38		Basic Information			created [created]	.	Replaces	issued	Format		*Language+*	Variable list	A concept (keyword) of the items in the coll		.
39		Basic metadata includ	.		Date when something	.	Is Required By	isVersionOf	The technical structur		a list of languages inc	List of variables in the	Object	cld:objectName (sub	.
40		Bibliographic Metada	.		short description [sh	.	Requires	language	Identifier.Persistent	Lexical Entry	*EntryElements*	Code lists	An object name associated with the items it		.
41		Specific to document	.		Short (8 or 9 words) g	.	Is Part Of	license	A unique name for th	a block which describ	a possibility to add fe	Explanation of codes	Name	cld:agentName (sub-	.
42		Resource Creators	.		description [descript	.	Has Part	mediator	Identifier.URL	Format		Versions	An personal or corporate name associated v		.
43		These fields are used	.		Plain text description	.	Is Referenced By	medium	The current Uniform	a rough indication of	.	Date/time stamp for	Place	dcq:spatial (sub-prop	.
44		Resource Specific or	.		release [release]	.	References	modified	Source	AccessTool		Checksums	The spatial coverage of the items in the coll		.
45		Describes agency or p	.		A project release.	.	Is Format Of	provenance	Title of a broader res	many lexica are only i	.	To test if your file has	Time	dcq:temporal (sub-pr	.
46		Investigation Type(s)	.		mailing list [mailing-l	.	Has Format	publisher	Language	Media			The temporal coverage of the items in the c		.
47		Lists all investigation	.		Mailing list home pag	.	Conforms To	references	The full name of the l	this entry tells wheth	.				.
48		Site Description Infor	.		category [category]	Coverage	Spacial	relation	Coverage.Spatial	Schema			Dates		.
49		Includes Site Name, S	.		A category of project.	The spatial or tempor	Temporal	replaces	The spatial coordinat	name of or ref to the	.				.
50		Material Type(s)	.		license [license]	Rights	Access Rights	requires	Rights	Character Encoding+			Accumulation Date R	cld:accumulationDat	.
51		Artifact types coverec	.		The URI of an RDF de	Information about rig	License	rights	Formal statement of	this list should give ar	.		The range of dates over which the collectio		.
52		Cultural Term(s)	.		repository [repositor	.		rightsHolder	Audience	Size			Contents Date Range	cld:contentsDateRan	.
53		Includes Culture (e.g.	.		Source code repositc	Audience	Mediator	source	Group expected to ac	the size of the lexicor	.		The range of dates of the individual items w		.
54		Temporal Coverage	.		anonymous root [ane	.	Education Level	spatial	.	No Lexical Entries					.
55		Includes Temporal Te	.		Repository for anonyr	.		subject		the number of lexical	.		Associated agents		.
56		General Keyword(s)	.		browse [browse]	.		tableOfContents		Access					.
57		Spatial Terms	.		Web browser interfac	.		temporal		sub-schema where ac	.		Collector	dc:creator	.
58		Includes Geographic T	.		module [module]	.		title		Keys			The identifier for an agent who gathers (or		.