

STATUS OF US CONTRIBUTION TO ISO ARCHIVE STANDARDS

PRESENTATION TO THIC MEETING

BY

**DON SAWYER
NASA/SCIENCE OFFICE OF STANDARDS AND TECHNOLOGY
(NOST)**

OCTOBER 18, 1995

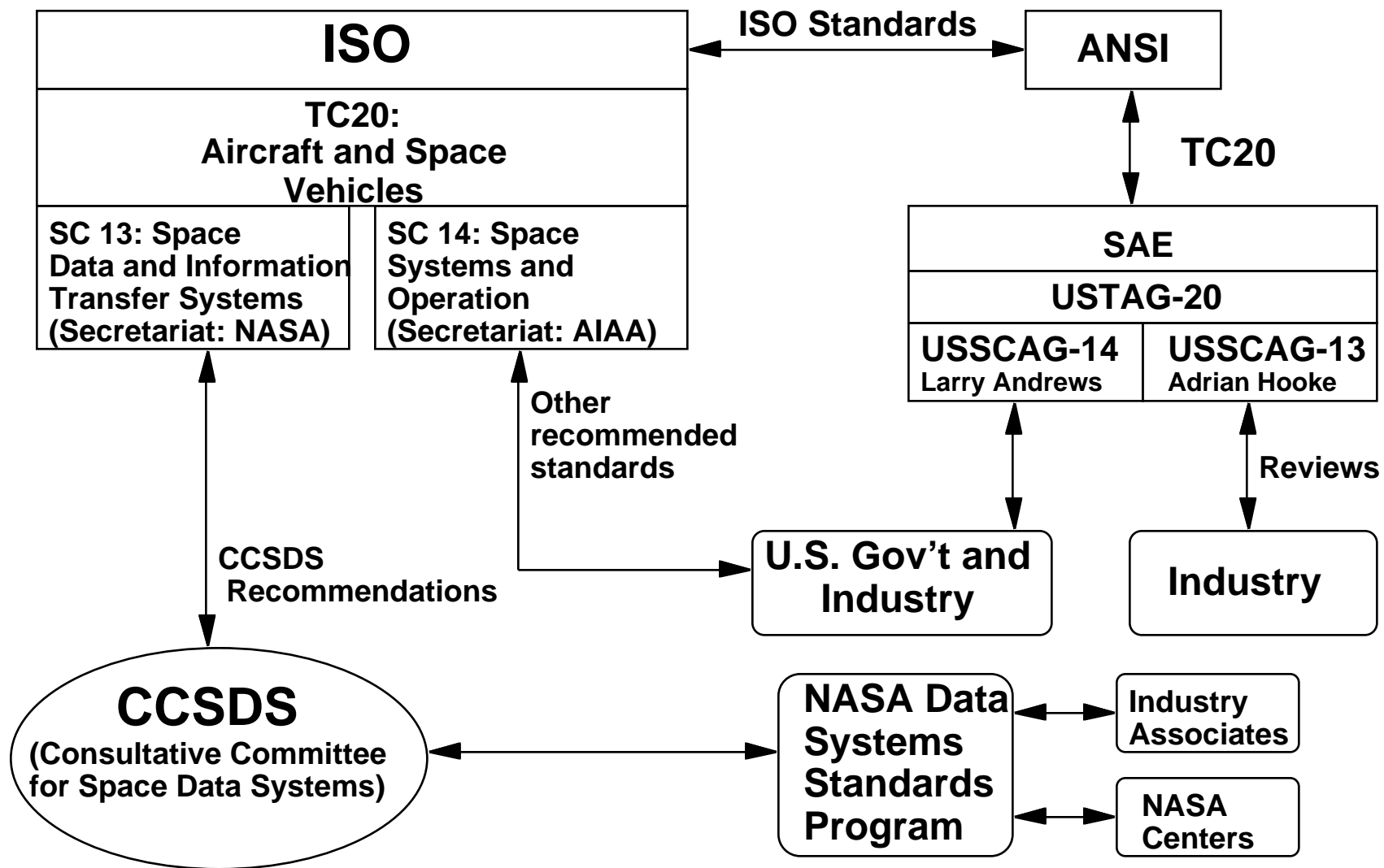
Presentation Outline

- **Organizational context**
- **ISO archive work item highlights**
- **What are some archiving issues?**
- **US organizational approach**
- **Technical approach**
- **Results of first US workshop**
- **Proposed Reference Model Impacts**
- **Proposed schedule**
- **Challenge**

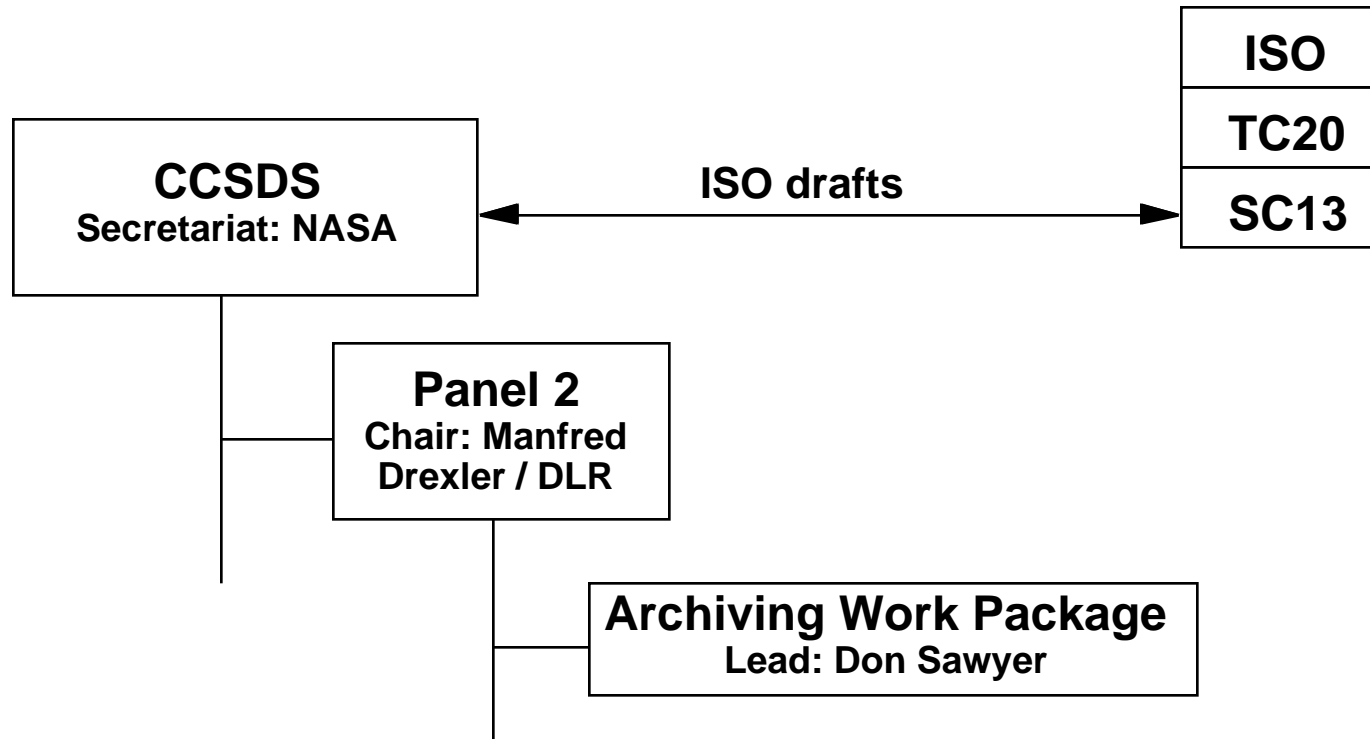
Organizational Focus of Archiving Task

- **Archiving standards being developed under ISO TC20 / SC13**
- **ISO Technical Committee (TC) 20: Aircraft and Space Vehicles**
 - Secretariat: U.S. Aerospace Industries Association
- **TC 20 / Sub-Committee (SC) 13: Space Data and Information Transfer Systems**
 - Chair: Mr. Jean Latour / CNES
 - Secretariat: NASA
- **Organizational focus does NOT mean that we are restricted to space data archiving only**
 - Believe there is widespread commonality in basic archiving issues
 - Need to start with a broad base of disciplines doing archiving

U.S. Space Standards Process



International Development of Archive Standards



- **International effort organized under CCSDS Panel 2 in May, 1995**
 - NASA has lead of Archiving Work Package
 - Example liaisons include CEOS, ISO/TC211, ISO/TC 46, and IACG

ISO Archive Work Item Highlights (WI 14721)

- **Scope**

- Define archive reference model and service categories
- Address intermediate and long term storage of digital data
- Address data used in conjunction with space missions
(BUT WE ARE NOT RESTRICTED TO SPACE DATA ONLY)

- **Purpose**

- Provide framework and common terminology to be used in the provision of archive services
- Encourage commercial support for provision of archive services
- Preserve our valuable data, not only for space related data but for all long term data archives
- First step before adopting or developing specific standards to support archive services

Some Archive Issues

- **Agencies have a significant stewardship responsibility for the information obtained from their observational programs**
 - Observational data are often irreplaceable
 - Taxpayer's investment must be prudently managed
- **Long term (indefinite) preservation of this information is difficult**
 - Data+metadata (i.e., information) must be migrated across new media, operating systems, and management systems
 - Field representations and formats may need to be revised to keep pace with evolving technologies and supported standards
 - What constitutes adequate metadata is not widely understood or standardized
 - Information is becoming ever more widely distributed
 - Information must be readily transportable from archive to archive
- **No consensus on what 'archiving' means, or what services might be available from an archive**
 - Preserving the information is not the same as preserving the data bits
 - Will the customer understand the information in 10, 50 or 100 years?
 - How can we know when effective archiving is being achieved?
 - Lack of consensus impedes establishment of commercial support services

Approach to Issue Resolution

- **Leverage the growing, widely distributed, expertise and interest in archiving of digital information**
 - Government
 - Industry
 - Academia
- **Focus on achieving a useful, near-term (2-3 years), reference model standard**
 - Common concepts and vocabulary
 - Classes of archives
 - Types of services by class of archive
- **At appropriate point, begin additional efforts**
 - Identify relevant existing standards
 - Identify and develop any missing standards
- **Recognize the impacts of rapid technological evolution and special needs of disciplines/sub-disciplines**
- **Promulgate resulting standards broadly**
 - ISO approval
 - Web access

US Organizational Approach

- **Organize US contribution under a framework with NASA lead**
 - Consider existing standards groups as possible umbrella for US effort
 - Establish liaison with Federal Geographic Data Committee (FGDC)
 - Agency archives and users must be represented in this process
 - US contributions to be submitted to CCSDS Panel 2
- **Will be an “Open” process**
 - Important to stimulate dialogue with broad archive/user communities
 - Results of US and International workshops will be put on WEB
 - Will support e-mail comments/critiques
- **Form a core group willing to devote 10-20% of time**
 - Develop material
 - Attend meetings
- **Expect a much larger group to review documents and/or track efforts**

Technical Approach

- **Produce an Archive Reference Model as a first standard**
 - Define what is meant by ‘archiving’
 - Break ‘archiving’ into a few interfaces (e.g., for ingest, storage, dissemination, and searching functions)
 - Define a set of services, some optional, for interfaces
 - Standard Reference Model should be quite stable over time
- **Produce a standard mapping of archive interfaces to applicable standards**
 - e.g., ingest standards “profiles” may include FTP, ISO 12175 (SFDU), BENTO (Open Doc), FGDC Metadata standards, ISO 9660, etc.
 - Give rationale / circumstances for use of each
 - Standard will need updating more often than the Reference Model
- **Identify gaps in needed standards and propose development efforts**
 - e.g., requirements on types of metadata needed
 - » identify what is common and what is discipline specific

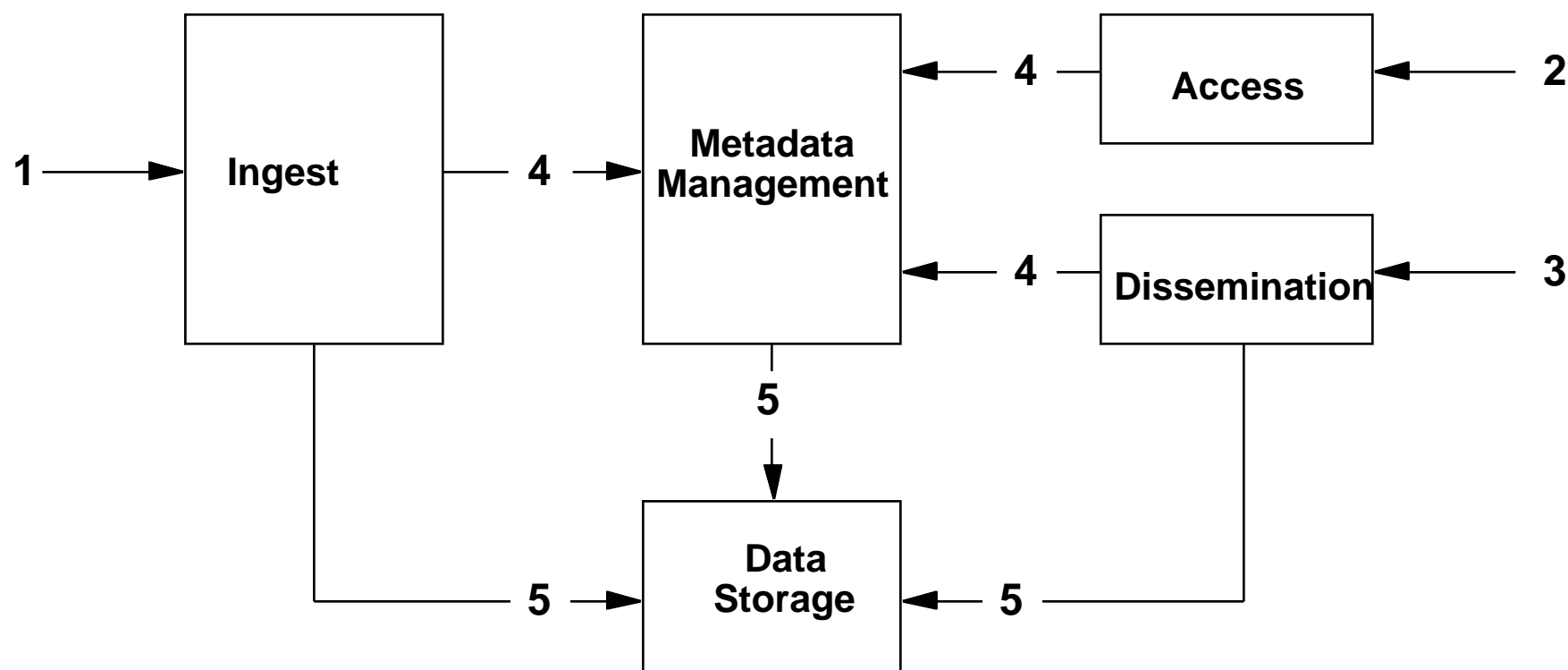
Results of First US Workshop: Overview

- **Very successful workshop**
 - Held October 11-12 at U. of MD Conference Center
 - Over 50 attendees (some turned away for lack of space)
 - Good range of government, industry, and academia participation
 - Enthusiastic support for continuing the effort
 - Advanced the draft Reference Model and terminology
 - Identified a number of issues
 - Scheduled next US workshop for 19-20 December
- **Continue effort toward broader awareness and participation**
 - Oil services
 - Environmental protection agency
 - More vendors
 - Others?!

Results of First US Workshop: Reference Model

- **Draft reference model accepted as good start**
 - Classes of archives needed
 - Ingest, Data Storage, Metadata Management, Access, Dissemination
- **Recommended enhancements**
 - Incorporate explicit “policy” function
 - Include issues of security, distributed functions
 - Cast into an object oriented framework
- **Some issues**
 - Include physical items and analog storage as well as digital storage?
 - Replication only or maintenance and dissemination of depletable samples/specimens?
 - Role of transactions (internal interfaces?) vs versioning (external interfaces?)
 - Migration - media,metadata,systems/archive/application software
- **Prototyping**
 - Attempt to describe several archives in terms of the Reference Model as soon as practical

Results of First US Workshop: Initial Reference Model Schematic



Archival Information Services and Interfaces

Results of First US Workshop: Initial Reference Model Interfaces

- 1 Ingest:** the external interface which accepts information into the archive. This may include staging of information in preparation for full acceptance, confirmation of receipt, and validation. [Key technology areas include data packaging and metadata. Existing standards include SFDU, BENTO(OpenDoc),FGDC metadata standards.]
- 2 Access:** the external interface for querying, browsing and product ordering by information consumers including other archive sites. [Key technology areas include request brokers, WWW browsers, WAIS, DBMS, data mining(AI), human machine interface(HMI) and data modeling. Existing standards include SQL, Z39.50, and OMG CORBA and services.]
- 3 Dissemination:**the external interface which sends copies of requested information to information consumer. This may include staging to on-line storage, notification to users of availability or direct dissemination via a protocol such as http. [Key technology areas include transportable media, network, file transfer protocols, and data/object packaging techniques and relocatable code (e.g. Java).Existing standards include CDF, HDF, BENTO, SFDU, SDTS, TCP/IP, FTP,OMG CORBA and services, and Remote Data Access(RDA).]
- 4 Metadata Management:** the internal interface for populating and maintaining catalogs, directories, inventories and related metadata stores. [Key technology areas include database management systems, metadata, hypertext, data modeling. Existing standards include SQL, HTML, POSC,STEP/EXPRESS.]
- 5 Data Storage:**the internal interface for storage and retrieval of data products from long-term stores (i.e. media). This may include transfer from staging storage to long term storage, and management of multiple types of media. [Key technology areas include file storage management systems (FSMS), media evolution, data compression, and database management systems (DBMS). Existing standards include the IEEE Mass Storage Reference Model , POSIX, and SQL92 (and SQL3).]

Results of First US Workshop: Terminology

- **Archive:** A repository that intends to preserve information for a designated community;
 - An archive holds data (information) that are correctly and independently useable by the intended user community when withdrawn.
 - Metadata must be sufficient (for the data) to be used by the intended user community.
 - An archive has a single management / stewardship domain
- **Data:** The representation forms of information
- **Information:** Any type of knowledge that can be exchanged
- **Metadata:** Data about other data

Many other terms need definitions in the context of the Reference Model

Proposed Impacts from Reference Model in Three Years

- **Improves our ability to understand and evaluate all archives with regard to:**
 - Use of new technologies
 - Individual archive experience
 - Migration issues, etc.
- **Will assist in education of new archive developers**
- **Will facilitate improvement of archive functions for new and old archives**
- **Will improve our ability to cost the development, maintenance, and operation of archives**
- **Will be able to identify what is an archive**
- **Will be a step closer to providing for common public interfaces to archives**
- **Better vendor support in addressing archive needs**

Proposed Schedule

- **First US Workshop - October 11-12, 1995**
 - Seek broad Gov't and industry committment to ongoing effort
 - Draft materials to be submitted to first international workshop
- **First international workshop - October 26-27, 1995 in Oxford, England**
 - CCSDS Panel 2 workshop
- **Second US workshop - core group - December 19-20, 1995**
 - Washington area (tentatively at Mitre facilities)
- **Third US workshop - core group - March ??, 1996**
- **Second international workshop - May, 1996 in Pasadena, CA, USA**
 - ISO working draft
- **Begin to relate Reference Model interfaces to existing standards - October, 1996**
 - Break into discipline subgroups as needed
- **Fourth international workshop - May, 1997**
 - ISO committee draft
- **Sixth international workshop - May, 1998**
 - ISO Draft International Standard

Challenge

- **Off to a good start with US participation in 11-12 October workshop**
 - We know not all interested parties are involved or even aware
- **Will you participate? Sign-up at:**
 - <http://www.gsfc.nasa.gov/nost/isoas/us/overview.html>
 - Or, contact Don Sawyer (sawyer@nssdca.gsfc.nasa.gov)
- **Will you alert colleagues who may be interested?**
- **We have opportunity to lead the international effort**
- **Scope of progress that can be made is unknown**
 - I believe an archiving reference model is definitely achievable
 - How much farther we can go depends on the effort provided
 - The organizers are open to all input

Acronyms

BNSC	British National Space Agency
CEOS	Committee on Earth Observing Satellites
CCSDS	Consultative Committee for Space Data Systems
CNES	Centre National d'Etudes Spatiales
DIS	Draft International Standard
FGDC	Federal Geographic Data Committee
IACG	InterAgency Consultative Group
ISO	International Organization for Standardization
NASA	National Aeronautics and Space Administration
SFDU	Standard Formatted Data Unit
TC	Technical Committee
WI	Work Item