

Archiving of Web Information at GSFC

The Digital Archiving System (DAS)

Nikkia Anderson
Gail Hodge

Metadata Review Group Meeting
April 23, 2004



History of the DAS

- Began evaluating need for web capture in 2001
- Developed draft Goddard Core Metadata Element Set with emphasis on project-related objects
- Captured > 230 web sites of scientific and technical information (> 89,000 pages)
- Developed a system for metadata extraction, creation and enhancement
- Developed a system for indexing and searching the metadata
- Extended system to include multiple digital object types – images, videos, and project documents



Welcome to the Digital Archiving System at NASA Goddard Space Flight Center
Library

Home

Update

Retrieve

About

This is one of the Goddard Library projects focused on Digital Preservation and
Knowledge Management.

Searching 91710 Web pages and various digital objects including Balloon
Technology Documents.

Please feel free to explore all the functionality. Any feedback will be greatly
appreciated.



Search for Digital Objects

Search for:

in:

AND

AND

Search

Clear

- All
- Creator.Code
- Creator.Employee
- Description
- Format
- Keyword
- Record No.
- Subject Competencies**
 - EOS Taxonomy
 - NASA Taxonomy
- Title
- Title, Description, Keyword
- Type

[About](#) | [Help](#) | [Home](#)

Home

Update

Retrieve

Retrieve

» Digital
Objects

Retrieve

» by
Record
Number

About



Search for Digital Objects

Home

Update

Retrieve

- Retrieve
- » Digital Objects
- Retrieve
- » by Record Number

About

Search for:

in:

[About](#) | [Help](#) | [Home](#)





Search for Digital Objects

Home

Update

Retrieve

- Retrieve
- Digital Objects
- Retrieve by Record Number

About

Search for:

Space Segment

Space Segment

Instruments

Observatory

Spacecraft

Launch Segment

Flight Support Equipment

Ground Support Equipment

Launch Vehicle

Ground Segment

Calibration

Measurements

Validation

Atmospheric Science

Geophysical/Geologic Science

Oceanographic Science

Instrument Technology

Sensors

AND

AND

in:

EOS Taxonomy

All

All

Clear

[Help](#) | [Home](#)



- Aerospace engineering
- Electrical engineering
- Mechanical engineering
- Bioengineering
- Life Sciences
- Agriculture
- Biology
- Botany
- Ecology
- Genetics
- Geobiology
- Physiology
- Zoology
- Biotechnology
 - Cellular Biotechnology
 - Macromolecular Biotechnology
 - Biotechnology Research
- Earth Sciences
 - Atmosphere
 - Biosphere
 - Cryosphere
 - Geosphere
 - Hydrosphere
 - Systems, Interactions, Feedback Loops
 - Current Issues
 - Earth in Space
 - Geochemistry
 - Geology
 - Geophysics

The Goddard Library

Federal Library of the Year 2002

Objects

Home

Update

Retrieve

- Retrieve
- » Digital Objects
- Retrieve
- by
- » Record Number

About

Search for:

AND

AND

in: NASA Taxonomy

All

All

Search Term: Sensors



Digital Archiving System at

The Goddard Library
Federal Library of the Year 2002

Displaying 1 to 10 of 31

Get	Meta	Preview	Title	Subject.Competencies	Creator.Employee	Creator.Code
		Image	Comparing Two Satellite Sensors	NASA:Earth Sciences, EOS:Sensors, EOS:Instruments		
		Web page	Space Geodesy Networks and Sensors Calibration Office	EOS:Sensors, EOS:Calibration, NASA:Physics, NASA:Geosphere, NASA:Earth Sciences		920.1
		Web page	NPOESS Preparatory Project Sensors	NASA:Earth Sciences, EOS:Sensors, EOS:Instruments, EOS:Observatory		429
		Video	MODIS performance	EOS:Sensors, EOS:Calibration		293
		Web page	Comparison of L7 and EO-1 Sensors	EOS:Spacecraft, EOS:Instruments, EOS:Sensors		427

Home

Update

Retrieve

Retrieve
» Digital
Objects
Retrieve
» by
Record
Number

About

Search Term : Balloon



Digital Archiving System at

The Goddard Library
Federal Library of the Year 2002

Displaying 1 to 10 of 68

Get	Meta	Preview	Title	Subject.Competencies	Creator.Employee	Creator.Code
		Balloon Technology	BT-2943 Report of the Balloon Programme Review Panel	NASA:Astronomy		
		Balloon Technology	BT-2949 Transatlantic Balloon Facility, A Scientific Case	NASA:Physics, NASA:Astronomy, NASA:Atmosphere, NASA:Technology		
		Video	Teaching an old dog new tricks: the ultra-long duration balloon	NASA:Technology		293
		Balloon Technology	BT-1916 Basic Procedures for Small Balloon Flights 1976	NASA:Atmosphere, NASA:Physics, NASA:Communications		
		Balloon Technology	BT-2926 Structural Analysis of a 35.m Diameter Balloon for Long Duration Ballooning	NASA:Physical Sciences	W. W. Schur	
		Balloon	BT-1671 Analyses Conducted to Assess Ballasting Strategy for the	NASA:Atmosphere,	George Conrad,	

Home

Update

Retrieve

Retrieve
Digital
Objects
Retrieve
by
Record
Number

About

Goddard Core Elements



http://media4.umd.edu:8180/ale/src/html/demoHome.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://media4.umd.edu:8180/ale/src/html/demoHome.html

Digital Archival System at **The Goddard Library**
Federal Library of the Year 2007

Goddard Core Metadata Template

Home

Update

- Add New Collection
- Delete Collection
- Update Metadata
- Update Lucene

Retrieve

- Retrieve Digital Objects
- Retrieve by Record Number

About

Help:

- Definitions of the Goddard Core elements
- Follow the link from a field's name for help with completing that field.
- DC Elements, Reference Version

It is possible to exceed the visible limitations of the input boxes.
If you need to repeat a field, just click on the that accompanies the field.

ID

1 TITLE of the resource to be described

Alternative title (Titles other than main title)

2 CREATOR (Name of the person or organization primarily responsible for creating the intellectual content)

Creator Employee

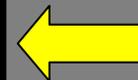
Creator Organization

Done Internet

Goddard Core ID/1503

I. Descriptive Elements

title	The Title element contains a name of the resource.
title.alternate	The Title.alternate element contains some other name of the resource.
creator	The Creator element contains the name of an entity that participated in the creation of the resource.
creator.employee	The Creator.employee element contains the name of an employee that participated in the creation of the resource.
creator.code	The Creator.code element contains a Goddard organizational code under which the work was created.
creator.contract	The Creator.contract element contains a value that identifies a contract under which the work was created.
creator.missionsProjects	The Creator.missionProject element contains a value that identifies a project under which the work was created.
creator.organization	The Creator.organization element contains a value that identifies an organization under whose auspices the resources was created.
creator.instrument	The Creator.instrument element contains a value that identifies an instrument that created the resource.
contributor	The Contributor element contains the name of an entity that contributed to the present state of the resource.
contributor.employee	The Contributor.employee element contains the name of an employee who contributed to the present state of the resource.



Template for Metadata Records

Search Term: Balloon (All)

The screenshot shows a web browser displaying search results for 'Balloon' on the NASA Digital Archive. A preview window is open over the first result, showing a document titled 'Report of the Balloon Programme Review Panel' with a technical drawing of a balloon structure. The main page lists several search results with columns for document type, title, and associated categories.

Get	Meta	Preview			
		Balloon Technology	BT-1916 Basic Procedures for Small Balloon Flights 1976	NASA:Atmosphere, NASA:Physics, NASA:Communications	
		Video	Teaching an old dog new tricks: the ultra-long duration balloon	NASA:Technology	293
		Balloon Technology	BT-2926 Structural Analysis of a 35.m Diameter Balloon for Long Duration Ballooning	NASA:Physical Sciences	W. W. Schur
		Balloon	BT-1671 Analyses Conducted to Assess Ballasting Strategy for the University of California, San Diego	NASA:Atmosphere,	George Conrad,

Search Term: Sensors (All)

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media

Address <http://media4.umd.edu:8180/ale/src/html/demoHome.html>

 **Digital Archiving System**

Displaying 1 to

Get	Meta	Preview	Title
		Image	Comparing Two Satellite Sensors
		Web page	Space Geodesy Networks and Sensors Calibration Office
		Web page	NPOESS Preparatory Project Sensors
		Video	MODIS performance
		Web page	Comparison of L7 and EO-1 Sensors

<http://library01.gsfc.nasa.gov/nasims/GL-2002-00>



NASA:Earth Sciences		
NASA:Earth Sciences, EOS:Sensors, EOS:Instruments, EOS:Observatory		429
EOS:Sensors, EOS:Calibration		293
EOS:Spacecraft, EOS:Instruments, EOS:Sensors		427

Done Internet

Search Term: Sensors (All)

File Edit View Favorites Tools Help

Address <http://media4.umd.edu:8180/ale/sr>

 Digital Archives

Home Update Retrieve Digital Objects by Record Number About

Web page Web page Video Web page Web page

Comparison of L7 and EO-1 Sensors - Microsoft Internet Explorer

LANDSAT

Home Search

Comparison of Landsat 7 and EO-1 Sensor Characteristics

	Landsat 7	EO-1
Manufacturers:		
Comparison of L7 and EO-1 Sensors	EOS:Spacecraft, EOS:Instruments, EOS:Sensors	427
Ultra Sensitive Photon Sensors @ GSFC Technology	NASA:Technology	500
ASAS Sensor and Tilting Platform Details	EOS:Sensors, EOS:Instruments, NASA:Earth Sciences, NASA:Physics	920

Done Internet

Taxonomy of Web Capture Problems

- Problem with spidering (capture) tools
 - Relative references instead of complete, explicit URLs
 - Harvesting process loses the hierarchical information
 - Complexity of web sites outpacing capabilities of spidering tools
- Dynamic Web Sites
 - Fly-over images
 - Pages that are updated daily
 - Frame that is fed dynamically into another web page
- Deep Web
 - Marrying archiving with databases or content management systems
 - Transactional/event based output

Near-Term Activities

- Finalize reduced Goddard Core Metadata Set
- Finalize controlled domain values
- Recapture web sites and update other content
- Get system completely installed in GSFC's environment
- Modify the search interface
- Usability testing
- Roll out the beta version to the GSFC community over the summer

Mid-Term Activities

- Work with GSFC Webmasters, NASA–Wide Taxonomy developers and other centers working on metadata
- Improve metadata generation
- Improve integration with video indexing and ContentDM
- Integrate with full text searching
- Improve generation of metadata
- Implement preservation components (e.g., permanence ratings, persistent ids)

Longer-Term Activities

- Evaluate preservation metadata and formats such as PDF/A
- Address web capture problems by categories, identify categories automatically and initiate programmatic solutions