



The Smithsonian/NASA Astrophysics Data System
Closing the Loop: Linking Datasets to Publications and Back

Closing the Loop: Linking Datasets to Publications and Back

A. Accomazzi, G. Eichhorn, A. Rots

Harvard-Smithsonian Center for Astrophysics

aaccomazzi@cfa.harvard.edu



Overview

- Current linking infrastructure
- Dataset Identifiers in Literature
- Management of links: new infrastructure and proposed enhancements
- How to Participate



Links from Datasets to Literature

- Have existed between Data Centers and ADS since 1994
- Maintained by librarians, data archivists
- Bibcode-URL pairs, linking to either individual observations or aggregates
- Often part of data center's bibliographies



The Smithsonian/NASA Astrophysics Data System

Closing the Loop: Linking Datasets to Publications and Back

Example

MAST Multimission
Archive at Space Telescope

File Edit View Go Bookmarks Tools Help

MAST STScI Tools Mission Search Tutorial Site Search

About MAST Getting Started Suggestion Box

FAQ

High-Level Science Products

Software

FITS

Related Sites

NASA Datacenters

MAST Services

MAST and the VO

Newsletters & Reports

Data Use Policy

Dataset Identifiers

Acknowledgments

Dataset References

A Survey of O VI, C III, and H I in Highly Ionized High-Velocity Clouds
Fox, Andrew J., Savage, Blair D., Wakker, Bart P., 2006ApJS..165..229F

- FUSE:
[I9040104000](#), [Z9090501000](#), [A0460505000](#) (MRK9), [P1071101000](#) (MRK9),
[P1071102000](#) (MRK9), [P1071103000](#) (MRK9), [S6011601000](#) (MRK9),
[A0360202000](#) (MRK59), [B0220201000](#) (POX36), [D1180101000](#) (IZW92),
[P1013501000](#) (3C273), [A0940101000](#) (MRK153), [B0220401000](#) (IIZW70),
[C0810101000](#) (MRK501), [C1490501000](#) (MRK106), [D0280203000](#) (MRK876),
[D0400101000](#) (NGC625), [D0540101000](#) (MRK205), [D0540102000](#) (MRK205),
[D0540103000](#) (MRK205), [D0550101000](#) (MRK586), [D0550102000](#) (MRK586),
[D0760101000](#) (MRK290), [D0760102000](#) (MRK290), [D1540101000](#) (MRK279),
[P1010203000](#) (MRK335), [P1010204000](#) (MRK335), [P1010903000](#) (NGC985),
[P1011201000](#) (AKN120), [P1011202000](#) (AKN120), [P1011203000](#) (AKN120),
[P1012002000](#) (TON951), [P1012901000](#) (MRK421), [P1072901000](#) (MRK290),
[P1073101000](#) (MRK876), [P1073301000](#) (MRK501), [P1080303000](#) (MRK279),
[P1080304000](#) (MRK279), [P1080403000](#) (MRK817), [P1080404000](#) (MRK817),
[P1080601000](#) (MRK509), [P1110101000](#) (IZW1), [Q1060203000](#) (MRK205),
[S6010801000](#) (MRK205), [X0170101000](#) (MRK509), [X0170102000](#) (MRK509),
[Z0100101000](#) (MRK421), [Z0100102000](#) (MRK421), [Z0100103000](#) (MRK421),
[A0230404000](#) (NGC7714), [A0460102000](#) (NGC1705),
[A0460103000](#) (NGC1705), [A0860606000](#) (NGC7714).

File Edit View Go Bookmarks Tools Help

- On-line Data
- References in the article
- Citations to the Article (1) (Citation History)
- Refereed Citations to the Article
- SIMBAD Objects (77)
- NED Objects (65)
- Also-Read Articles (Reads History)
- Translate Abstract

Title: A Survey of O VI, C III, and H I in Highly Ionized High-Velocity Clouds

Authors: [Fox, Andrew J.](#); [Savage, Blair D.](#); [Wakker, Bart P.](#)

Affiliation: AA(Department of Astronomy, University of Wisconsin-Madison, 475 North Charter Street, Madison, WI 53706; Current address: Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago, 75014, Paris, France.), AB(Department of Astronomy, University of Wisconsin-Madison, 475 North Charter Street, Madison, WI 53706), AC(Department of Astronomy, University of Wisconsin-Madison, 475 North Charter Street, Madison, WI 53706)

Publication: The Astrophysical Journal Supplement Series, Volume 165, Issue 1, pp. 229-255. ([ApJS Homepage](#))

Publication Date: 07/2006

Origin: UCP

ApJ Keywords: Galaxy: Halo, Galaxies: Intergalactic Medium, ISM: Clouds, ISM: Kinematics and Dynamics, Ultraviolet: ISM

Abstract Copyright: (c) 2006: The American Astronomical Society

DOI: [10.1086/504800](#)

Bibliographic Code: 2006ApJS..165..229F



The Smithsonian/NASA Astrophysics Data System

Closing the Loop: Linking Datasets to Publications and Back

Example (cont.)

File Edit View Go Bookmarks Tools Help

- On-line Data
- References in the article
- Citations to the Article (251) (Citation History)
- Refereed Citations to the Article
- SIMBAD Objects (1)
- Also-Read Articles (Reads History)
- Translate Abstract

Title: XMM-Newton observation of the Lockman Hole. I. The X-ray

Authors: [Hasinger, G.](#); [Altieri, B.](#); [Arnaud, M.](#); [Barcons, X.](#); [Bergeron, Dennerl, K.](#); [Ferrando, P.](#); [Finoguenov, A.](#); [Griffiths, R. E.](#); [H. Lumb, D. H.](#); [Mason, K. O.](#); [Mateos, S.](#); [McMahon, R. G.](#); [M. Ptak, A. F.](#); [Sasseen, T. P.](#); [Schartel, N.](#); [Szokoly, G. P.](#); [Trüm Warwick, R. S.](#); [Watson, M. G.](#)

Affiliation: AA(Astrophysikalisches Institut Potsdam (AIP), An der Stern Germany), AB(XMM-Newton Operations Centre, European 50727, 28080 Madrid, Spain), AC(CEA/DSM/DAPNIA/Serv 91191 Gif-sur-Yvette Cedex, France), AD(Instituto de Fisica Investigaciones Cientificas-Universidad de Cantabria), 39005 Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 C AF(Astrophysikalisches Institut Potsdam (AIP), An der Stern Germany), AG(TeSRE-CNR, Via Gobetti 101, 40129 Bologna)

Done

Smithsonian/NASA ADS Abstract Service

Links for 2001A&A...365L..45H

Resource at [heasarc.gsfc.nasa.gov](#)
[On-line Data Link](#)

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700101

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700201

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700401

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700501

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700601

Resource at [xsa.vilspa.esa.es:8080](#)
XMM-Newton Observation Number 0123700701

Done

File Edit View Go Bookmarks Tools Help

esa XMM-NEWTON SCIENCE ARCHIVE

The XMM-Newton Postcard Server

File Edit View Go Bookmarks Tools Help

Get ODF data for the whole observation (might take long)		Get PPS data for the whole observation (might take long)	
The XMM-Newton Data File Handbook		The SAS User's Guide	
Check for Publications for this observation		Sources identified in this observation	
Target name Lockman Hole	RA(J2000) 163.17633	DEC(J2000) 57.47672	

File Edit View Go Bookmarks Tools Help

Check for Publications for this observation		Sources identified in this observation	
Target name Lockman Hole	RA(J2000) 163.14618000000002	DEC(J2000) 57.485218	
Start time 2000-04-27 02:46:15.0	End time 2000-04-27 21:37:20.0	On-time (s) 67865	
Observation ID 0123700101	SAS version SAS Version Multiple	Proprietary Expiry Date Mar 22 2000	

Exposure Number: 5001
Instrument: MOS1 Filter: Thin
Duration (s): 45999
Mode: Full Frame

Exposure Number: 5002
Instrument: MOS2 Filter: Thick
Duration (s): 61290
Mode: Full Frame

Exposure Number: 5003
Instrument: EPN Filter: Thin
Duration (s): 62010
Mode: Full Frame

Done



Links from Literature to Data

- ADS provides links to datasets from bibcode-URL pairs maintained by Data Centers
- Recently AAS journals introduce option of referencing datasets in papers
- Infrastructure for the management of links provided by ADS



The Name Game

- Current naming schemes and resolution services in use in astronomy include:
 - Bibcodes, eprints, DOIs (ADS)
 - Object names (SIMBAD/NED)
 - Data Catalogs (Vizier)
 - IVO Identifiers (IVOA Registry)
 - **Dataset Identifiers (ADS/ADEC)**



Dataset Identifiers

- In 2003, ADEC approves the definition of dataset identifiers:

ivo://**ADS/FacilityId#PrivateId**

- Properties: unique, permanent, resolvable
- Broad range of granularity
- Both Static and Dynamic Data product support



Dataset Identifiers: Examples

- ADS/Sa.CXO#obs/123
- ADS/Sa.CXO#defset/COUP
- ADS/Sa.CXO#bibcode/2005ApJ...635L..81A
- ADS/IRSA.Atlas#2006/0701/121559_24406

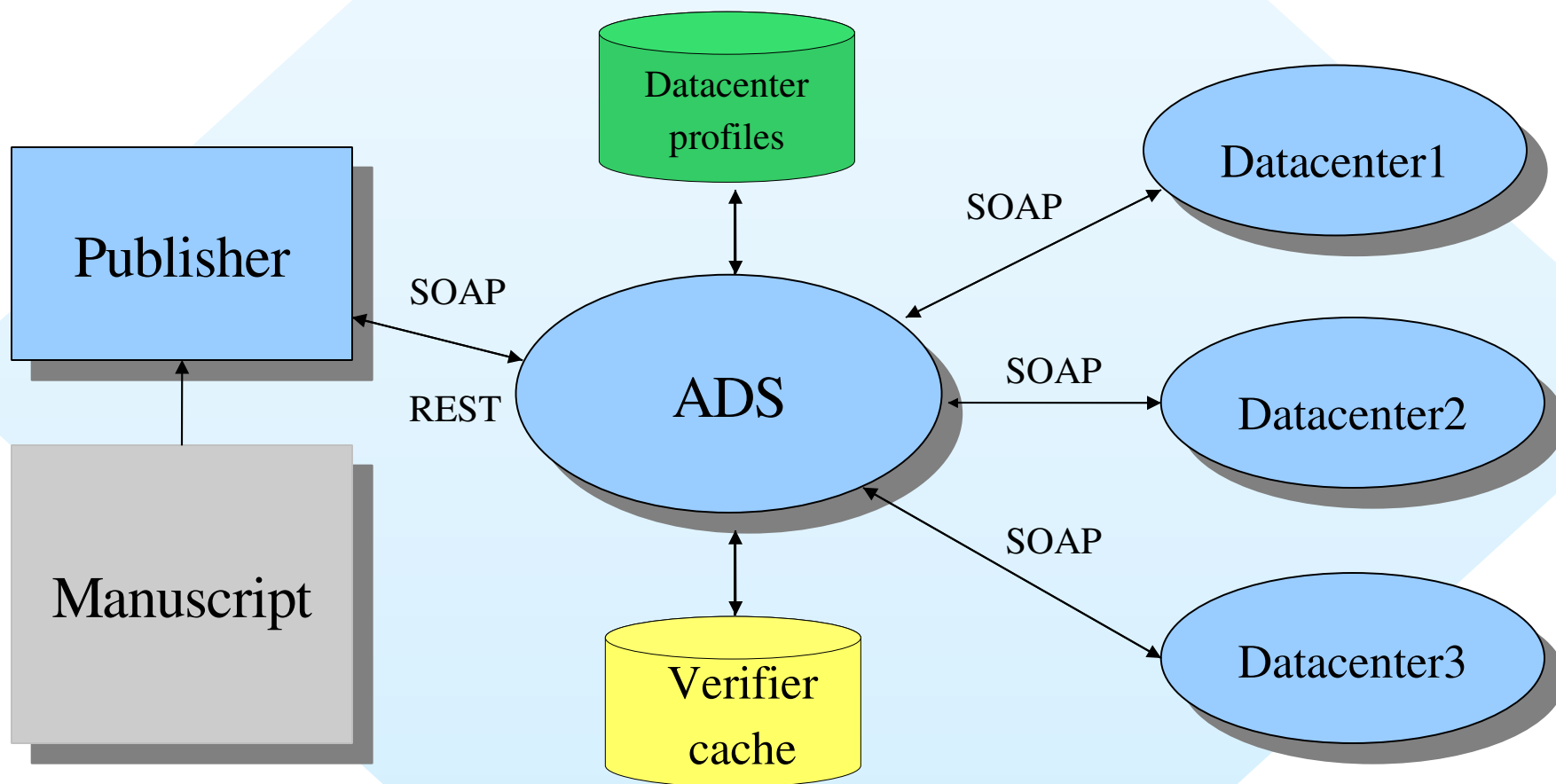


Current Use of Dataset Identifiers

- Data centers assign data set identifiers
- Author references data set identifiers in paper
- Publishers verify identifiers during copy-editing and create links to datasets in manuscripts
- ADS extracts identifiers from published papers and provide links from bibcodes to datasets



Dataset Identifier Verification

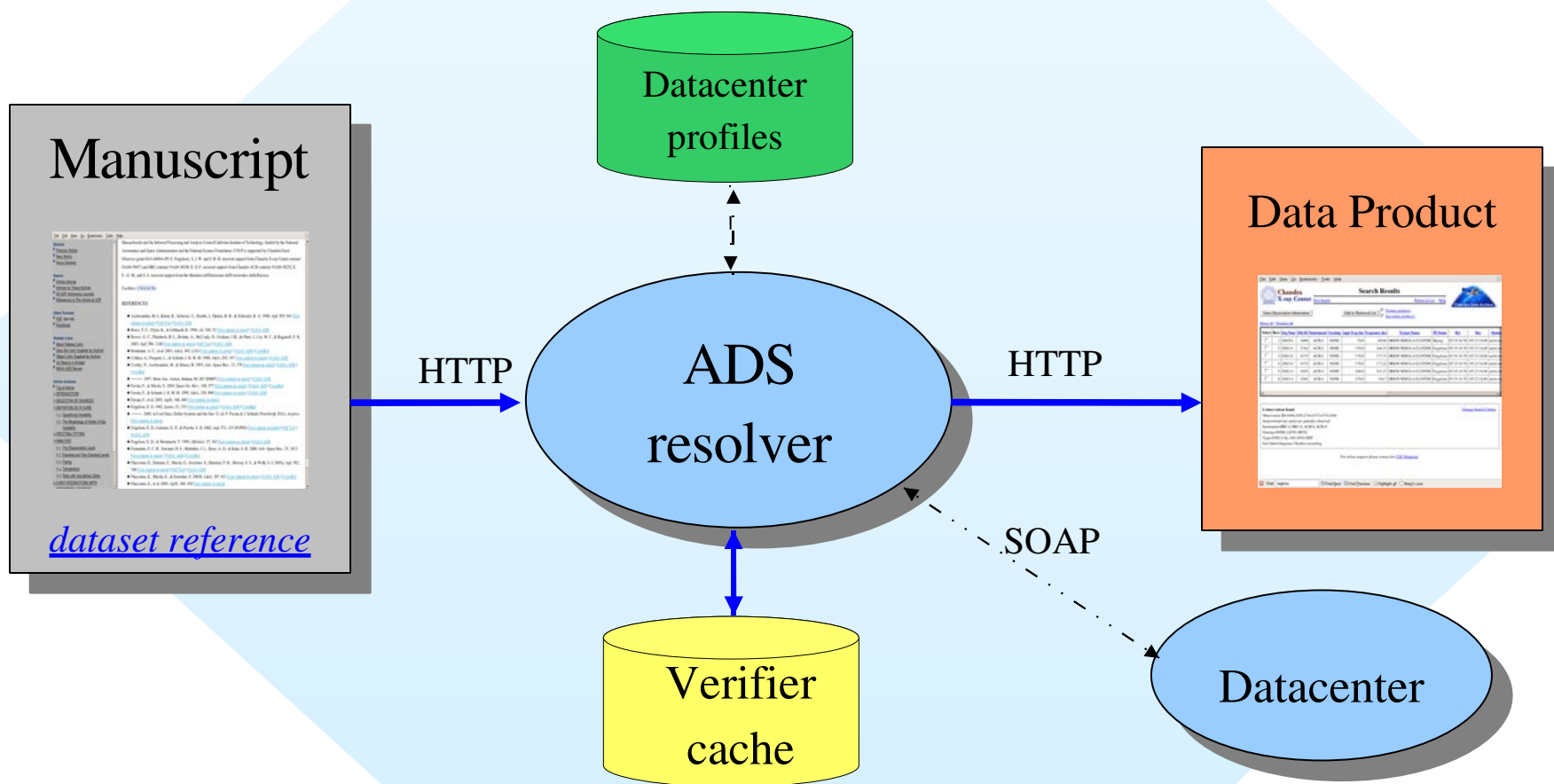




The Smithsonian/NASA Astrophysics Data System

Closing the Loop: Linking Datasets to Publications and Back

Dataset Link Resolution





Exchange of Identifiers

- Helps ADS maintain links to data centers and journal articles
- Helps data centers maintain bibliometric data, links to ADS
- Allows “grouping” of datasets by bibliography
- Facilitates corrections and updates



Proposed Linking Scheme

- ADS keeps track of all dataset ids cited in the current literature
- Data centers harvest from ADS bibcode-dataset id pairs for the data collections they archive and merge data with local bibliographic holdings
- ADS harvests Data centers for authoritative bibcode-dataset mappings and creates links



The Smithsonian/NASA Astrophysics Data System

Closing the Loop: Linking Datasets to Publications and Back

Harvesting Example

<http://vo.ads.harvard.edu/dv/HarvestDatasets.cgi?set=nasa.cda>

File Edit View Go Bookmarks Tools Help

Published Dataset Identifiers for Chandra Data Archive (at CfA/SAO) since
Thu Jan 1 00:00:00 1970 UTC

Dataset ID	Bibcode	Status	Link
ADS/Sa.CXO#obs/433	2005ApJ...631..733Y	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=433
ADS/Sa.CXO#obs/4759	2005ApJ...631..733Y	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=4759
ADS/Sa.CXO#obs/4760	2005ApJ...631..733Y	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=4760
ADS/Sa.CXO#obs/4761	2005ApJ...631..733Y	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=4761
ADS/Sa.CXO#obs/4762	2005ApJ...631..733Y	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=4762
ADS/Sa.CXO#obs/2548	2005ApJ...635L..81A	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=2548
ADS/Sa.CXO#obs/6204	2005ApJ...635L..81A	VERIFIED	http://cda.harvard.edu/chaser/ocatList.do?obsid=6204
ADS/Sa.CXO#defset/COUP	2005ApJS...160..319G	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html
ADS/Sa.CXO#defset/COUP	2005ApJS...160..353G	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html
ADS/Sa.CXO#defset/COUP	2005ApJS...160..379F	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html
ADS/Sa.CXO#defset/COUP	2005ApJS...160..390P	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html
ADS/Sa.CXO#defset/COUP	2005ApJS...160..401P	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html
ADS/Sa.CXO#defset/COUP	2005ApJS...160..423W	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html

```
enterName="Chandra Data Archive (at CfA/SAO)" since="Thu Jan 1 00:00:00 1970 UTC">
.631..733Y" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/433" url="http://cda.harvard.edu/
.631..733Y" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/4759" url="http://cda.harvard.edu/
.631..733Y" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/4760" url="http://cda.harvard.edu/
.631..733Y" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/4761" url="http://cda.harvard.edu/
.631..733Y" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/4762" url="http://cda.harvard.edu/
.635L..81A" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/2548" url="http://cda.harvard.edu/
.635L..81A" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/6204" url="http://cda.harvard.edu/
.160..319G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..353G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..379F" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..390P" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..401P" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..423W" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..450P" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..503T" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..511K" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..530G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..557R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.160..582P" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/COUP" url="http://cxc.harvard.edu/
.161...98" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#defset/Xbootes" url="http://cxc.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5015" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5016" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5017" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5018" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5019" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5020" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5021" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5022" url="http://cda.harvard.edu/
.161...211" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/5164" url="http://cda.harvard.edu/
.639..185G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/00587" url="http://cda.harvard.edu/
.639..185G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/00588" url="http://cda.harvard.edu/
.639..185G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/00630" url="http://cda.harvard.edu/
.639..185G" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/05123" url="http://cda.harvard.edu/
.642..260R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/05130" url="http://cda.harvard.edu/
.642..260R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/04840" url="http://cda.harvard.edu/
.645..293Z" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/04641" url="http://cda.harvard.edu/
.645..293Z" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/05362" url="http://cda.harvard.edu/
.645..293Z" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/05363" url="http://cda.harvard.edu/
.645..293Z" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/06099" url="http://cda.harvard.edu/
.647L.107R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/04889" url="http://cda.harvard.edu/
.647L.107R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/05733" url="http://cda.harvard.edu/
.647L.131R" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/00587" url="http://cda.harvard.edu/
.648...111" status="VERIFIED" desc="Dataset ID ADS/Sa.CXO#obs/00370" url="http://cda.harvard.edu/
```



Harvesting Example (cont.)

<http://vo.ads.harvard.edu/dv/HarvestDatasets.cgi?set=nasa.cda&since=2006-10-01>

File Edit View Go Bookmarks Tools Help

Published Dataset Identifiers for Chandra Data Archive (at CfA/SAO)
since Sun Oct 1 00:00:00 2006 UTC

Dataset ID	Bibcode	Status	Link
ADS/Sa.CXO#defset/COUP	2006ApJ...649..914S	VERIFIED	http://cxc.harvard.edu/cda/DefSet/COUP.html

File Edit View Help

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="DataSets.xsl"?>
<records center="nasa.cda"
  centerName="Chandra Data Archive (at CfA/SAO)"
  since="Sun Oct 1 00:00:00 2006 UTC">
  <record bibcode="2006ApJ...649..914S"
    status="VERIFIED"
    desc="Dataset ID ADS/Sa.CXO#defset/COUP"
    url="http://cxc.harvard.edu/cda/DefSet/COUP.html"
    id="ADS/Sa.CXO#defset/COUP" />
</records>
```



Requirements: Journals

- Add data set identifier tags to their macros.
- Verify identifiers during copy-editing
- Provide links
- Return identifiers with article metadata to ADS



Requirements: Data Centers

- Tag all data products with dataset identifiers and inform users
- Provide dataset verification service
- Harvest ADS for published identifiers
- Provide harvesting interface for ADS



Requirements: ADS

- Maintain central dataset verifier and resolver
- Provide service to verify bibcodes
- Provide harvesting service for published datasets-bibcodes
- Harvest complete list of dataset-bibcode pairs from data centers



Conclusions

- Links to datasets are already in the literature
 - Need to raise awareness and publisher participation
- Benefits to Data Centers
 - Automated exchange of dataset-bibcode pairs improves coverage and accuracy
 - Methodology can be used for any id-id exchange
- ADS providing much of the infrastructure required



The Smithsonian/NASA Astrophysics Data System

Closing the Loop: Linking Datasets to Publications and Back

For more information...

<http://vo.ads.harvard.edu>