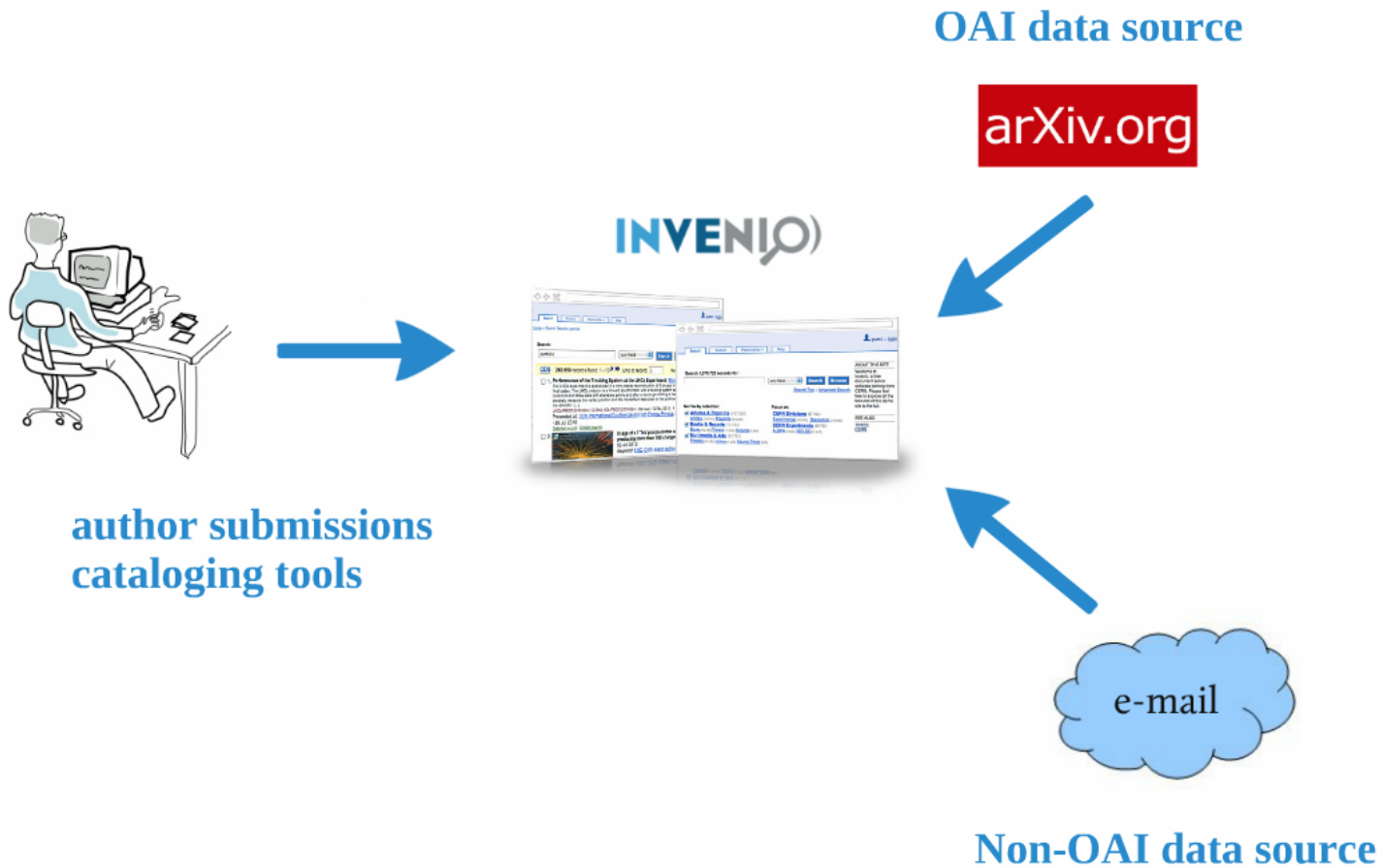


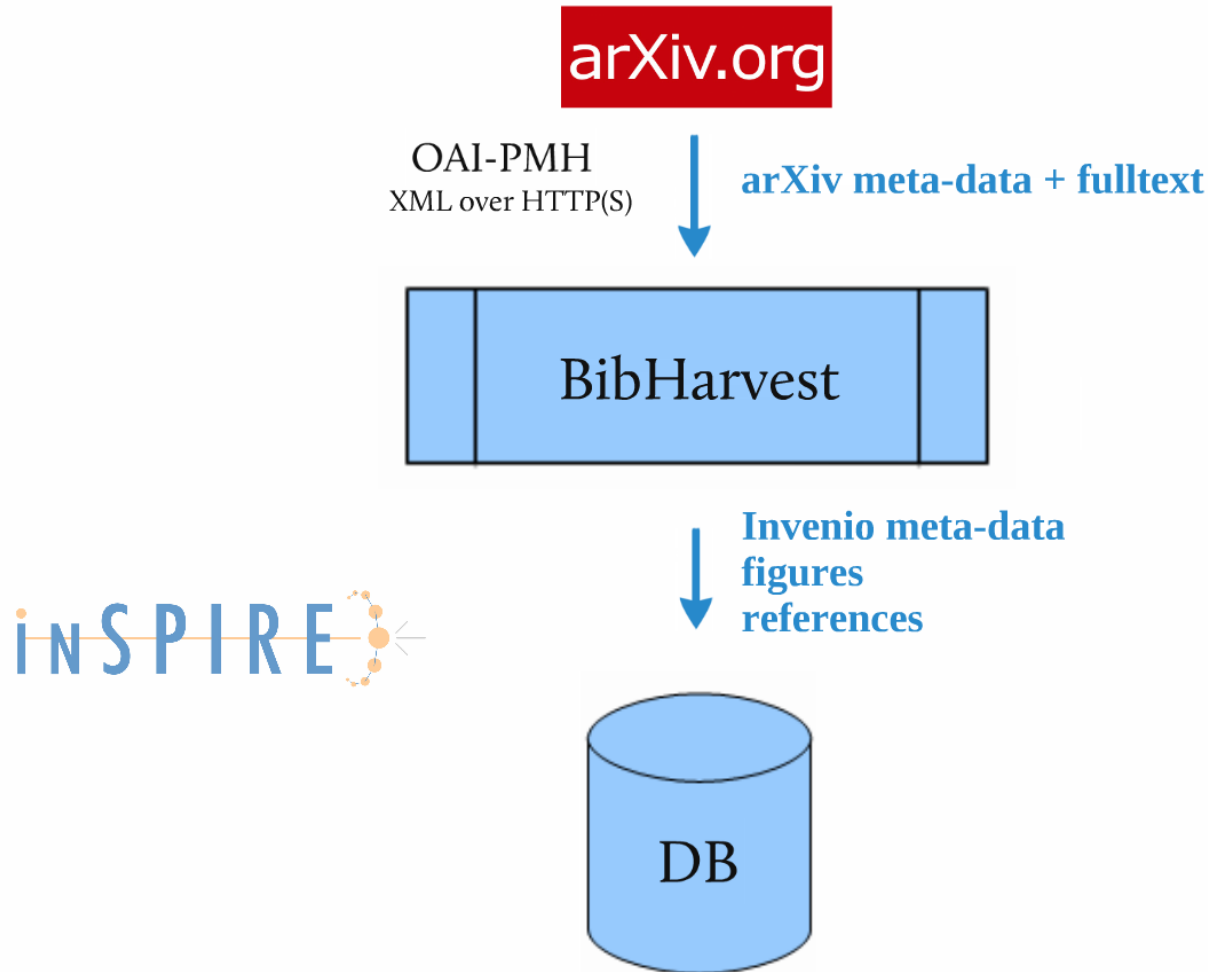
OAI Harvesting using BibHarvest

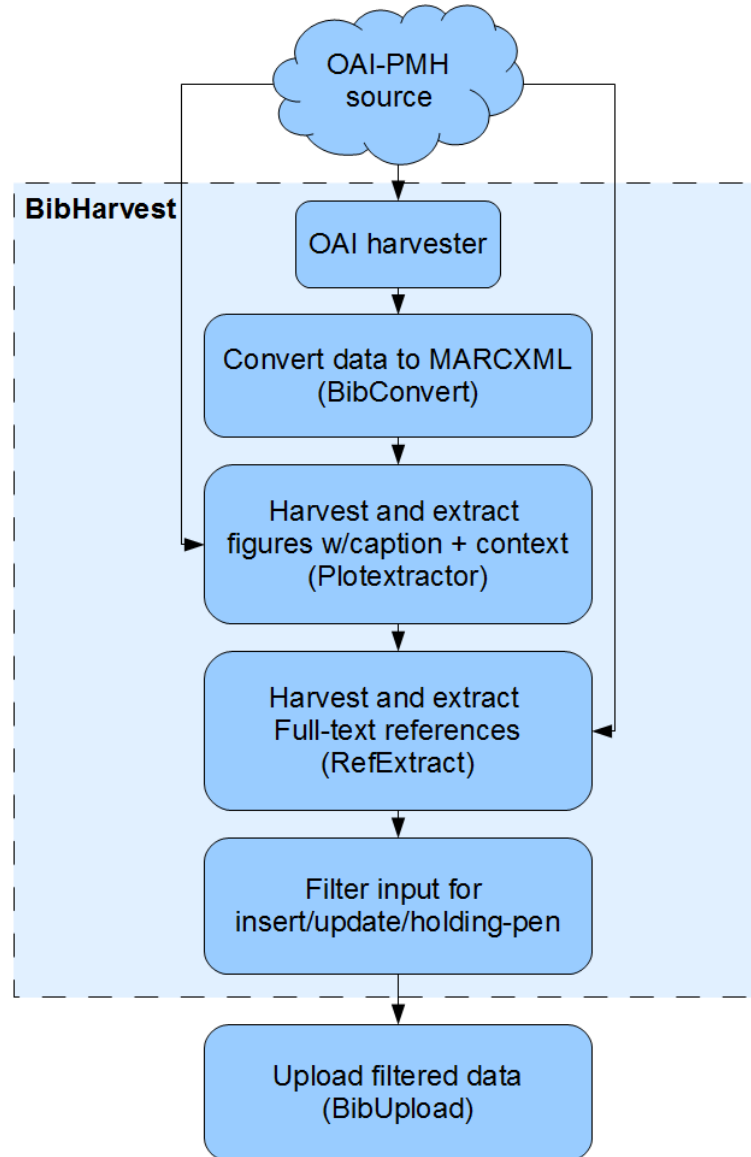
Invenio Workshop 2012

Ways of ingesting content in Invenio



Meta-data harvesting using OAI-PMH





find eprint arxiv:1007.5048

Search

[Search Tips](#)

[Advanced Search](#)

[find j "Phys.Rev.Lett.,105"](#) :: [more](#)

Sort by:

latest first

desc

- or rank by -

Display results:

25 results

single list

Output format:

HTML brief

HEP

1 records found

1. **Diffractional W and Z Production at the Fermilab Tevatron.**

CDF Collaboration (T. Aaltonen (Helsinki Inst. of Phys.) *et al.*). FERMILAB-PUB-10-255-E. Jul 2010. 10 pp.

Published in **Phys.Rev. D82 (2010) 112004**

e-Print: **arXiv:1007.5048 [hep-ex]**

[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [HarvMac](#) | [EndNote](#)

[Abstract](#) and [Postscript](#) and [PDF](#) from arXiv.org

[Journal Server](#)

[Fermilab Library Server \(fulltext available\)](#)

[Detailed record](#) - [Similar records](#) - [Cited by 4 records](#)

Diffractive W and Z Production at the Fermilab Tevatron.

CDF Collaboration (T. Aaltonen (Helsinki Inst. of Phys.) *et al.*) [Show all 514 authors.](#)

FERMILAB-PUB-10-255-E.

Jul 2010

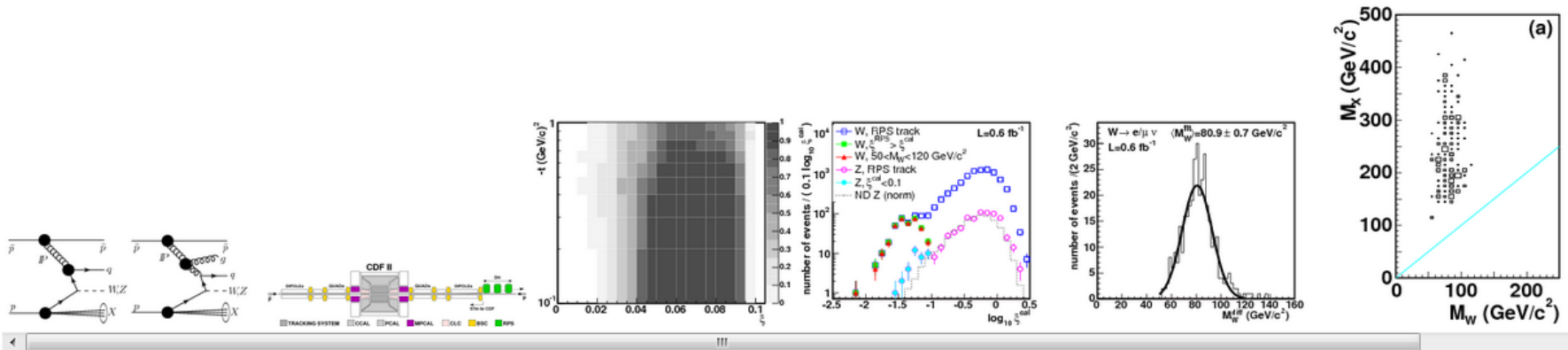
10 pp.

Phys.Rev. D82 (2010) 112004
e-Print: [arXiv:1007.5048 \[hep-ex\]](#)

Abstract: We report on a measurement of the fraction of events with a W or Z boson which are produced diffractively in antiproton-proton collisions at a center of mass energy of 1.96 TeV, using data from 0.6 inverse femtobarns of integrated luminosity collected with the CDF-II detector equipped with a Roman-pot spectrometer that detects the antiproton (\bar{p}) from $\bar{p}p \rightarrow \bar{p}p[X+W/Z]$. We find that $(1.00 \pm 0.11)\%$ of Ws and $(0.88 \pm 0.22)\%$ of Zs are produced diffractively in a region of antiproton or proton fractional momentum loss (ξ) of $0.03 < \xi < 0.10$ and 4-momentum transferred squared t of $-1 < t < 0$ $(\text{GeV}/c)^2$, where we account for the events in which the proton scatters diffractively while the antiproton dissociates, $\bar{p}p \rightarrow [X+W/Z]p$, by doubling the measured proton dissociation fraction. We also report on searches for W and Z production in double Pomeron exchange, $p+\bar{p} \rightarrow p+[X+W/Z]+\bar{p}$, and on exclusive Z production, $\bar{p}p \rightarrow \bar{p}p+Z+p$. No signal is seen above background for these processes, and comparisons are made with expectations.

Keyword(s): INSPIRE: [W: hadroproduction](#) | [Z0: hadroproduction](#) | [pomeron: exchange](#) | [diffraction](#) | [Batavia TEVATRON Coll](#) | [CDF](#) | [anti-p p: inclusive reaction](#) | [anti-p p: exclusive reaction](#) | [intermediate boson: leptonic decay](#) | [W: mass](#) | [experimental results](#) | [exchange: \(2pomeron\)](#) | [anti-p p \$\rightarrow\$ W anti-p anything](#) | [anti-p p \$\rightarrow\$ Z0 anti-p anything](#) | [anti-p p \$\rightarrow\$ Z0 anti-p](#) | [1960 GeV-cms](#)

Author supplied: [diffraction](#)



DEMO

How to setup OAI harvesting

Your Account

You are logged in as admin. You may want to a) [logout](#); b) edit your [account settings](#).

You have **0** new messages out of **0** messages.

You have 0 personal baskets and are subscribed to 0 group baskets and **0** other users public baskets.

You own the following alerts::

You have made 334 queries. A [detailed list](#) is available with a possibility to (a) view search results and (b) subscribe to an automatic email alerting service for these queries.

You can consult the list of [your groups](#) you are administering or are a member of.

You can consult the list of [your submissions](#) and inquire about their status.

You can consult the list of [your approvals](#) with the documents you approved or refereed.

You can consult the list of [your tickets](#).

You are enabled to the following roles: *superadmin*.

Here are some interesting web admin links for you:

[Configure BibFormat](#)

[Configure BibKnowledge](#)

[Configure BibIndex](#)

[Configure BibRank](#)

[Configure OAI Harvest](#)

[Configure OAI Repository](#)

[Configure WebAccess](#)

[Configure WebComment](#)

[Your Account](#)

[Your Messages](#)

[Your Baskets](#)

[Your Alert Searches](#)

[Your Searches](#)

[Your Groups](#)

[Your Submissions](#)

[Your Approvals](#)

[Your Tickets](#)

[Your Administrative Activities](#)

URL:

admin/bibharvest/oaiharvestadmin.py

OAI Harvest Admin Interface

Overview of sources [\[?\]](#)

[add new OAI source](#)

1 OAI sources currently present in the database

name	baseURL	metadatatprefix	frequency	bibconvertfile	postprocess	actions
arxiv	http://export.arxiv.org/oai2	arXiv	daily	/opt/cds-invenio/etc/bibconvert/config/oaiarXiv2inspire.xsl	c-r-p-t-f-u	edit / delete / test / history / harvest

Harvesting status [\[?\]](#)

Next oaiharvest task

- scheduled time: 2011-06-02 05:04:34

- current status: WAITING

[View Holding Pen](#)

name	last update
arxiv	2011-06-01 05:53:41

Edit OAI Source

Edit OAI source [?](#)

Source name

Base URL



Remote OAI-PMH repository

Sets *(Leave all unchecked for non-selective harvesting)*

- cs (Computer Science)
- math (Mathematics)
- nlin (Nonlinear Sciences)
- physics (Physics)
- physics:astro-ph (Astrophysics)
- physics:cond-mat (Condensed Matter)
- physics:gr-qc (General Relativity and Quantum Cosmology)
- physics:hep-ex (High Energy Physics - Experiment)
- physics:hep-lat (High Energy Physics - Lattice)
- physics:hep-ph (High Energy Physics - Phenomenology)
- physics:hep-th (High Energy Physics - Theory)
- physics:math-ph (Mathematical Physics)
- physics:nucl-ex (Nuclear Experiment)
- physics:nucl-th (Nuclear Theory)
- physics:physics (Physics (Other))
- physics:quant-ph (Quantum Physics)
- q-bio (Quantitative Biology)
- q-fin (Quantitative Finance)
- stat (Statistics)

Metadata prefix

Frequency

Postprocess *(Leave all unchecked for no post-processing)*

- convert (c)
- extract references (r)
- extract plots (p)
- attach full-text (t)
- filter (f)
- upload (u)



Active workflow components

BibConvert configuration file (if needed by postprocess) Eg: oaidc2marcxml.xsl, oaimarc2marcxml.xsl

BibFilter program (if needed by postprocess)

Modify

Automatic harvesting daemon

```
$ oaiharvest -r arxiv -s 24h --fixed-time
```

ID	PROC [PRI]	USER	RUNTIME	SLEEP	STATUS	PROGRESS
7	inveniogc	admin	2011-06-02 01:00:00	20h	RUNNING	Postponed 1 time(s)
3	bibreformat	admin	2011-06-02 01:02:23	5m	WAITING	
4	bibrank	admin	2011-06-02 01:02:37	5m	WAITING	
1	bibindex	admin	2011-06-02 01:02:53	5m	WAITING	
2	webcoll	admin	2011-06-02 01:03:08	5m	WAITING	
10696	dbdump	admin	2011-06-02 01:54:48	21h	WAITING	
10828	oaiharvest	jlavik	2011-06-02 05:04:34	21h	WAITING	
5	bibrank	admin	2011-06-08 21:47:11	7d	WAITING	

Automatic Mode [A Manual] [1/2/3 Display] [P Purge] [l/L Log] [O Opts] [Q Quit] 2011-06-02 01:01:46

New developments

OAI Harvest Admin Interface

Overview of sources [\[?\]](#)

[\[add new OAI source\]](#)

3 OAI sources currently present in the database

name	base URL	metadata prefix	frequency	last run	post-processes	actions	comments
arxiv	http://export.arxiv.org/oai2	arXiv	daily	Never harvested	c-p-u	edit / delete / test / history / harvest	testing
test	asdasd	test	daily	2011-08-16 13:24:47		edit / delete / test / history / harvest	
journal	http://export.arxiv.org/oai2	arXiv	weekly	2011-08-17 14:56:23	c	edit / delete / test / history / harvest	Silly comment

Harvesting status [\[?\]](#)

No oaiharvest task currently scheduled.

[View Holding Pen](#)

Edit OAI Source

Edit OAI source [?]

OAI HARVESTING JOB OPTIONS

Source name	<input type="text" value="arxiv"/>	e.g. arxiv
Base URL	<input type="text" value="http://export.arxiv.org/oai2"/>	
Meta-data prefix	<input type="text" value="arXiv"/>	e.g. oai_dc
Harvesting frequency	<input type="text" value="daily"/>	
Comments (optional)	<input type="text" value="testing"/>	

AUTOMATIC POST-HARVEST PROCESSES

(Leave all unchecked for no post-processing. * arguments are mandatory)

convert (c)

Cfg File *

extract plots (p)

Extraction Source * PDF
 LaTeX

extract references (r)

extract authors (a)

attach full-text (t)

filter (f)

upload (u)

Priority

Name

Modify

SETS TO HARVEST

- cs (Computer Science)
- math (Mathematics)
- nlin (Nonlinear Sciences)
- physics (Physics)
- physics:astro-ph (Astrophysics)
- physics:cond-mat (Condensed Matter)
- physics:gr-qc (General Relativity and Quantum Cosmology)
- physics:hep-ex (High Energy Physics - Experiment)
- physics:hep-lat (High Energy Physics - Lattice)
- physics:hep-ph (High Energy Physics - Phenomenology)
- physics:hep-th (High Energy Physics - Theory)
- physics:math-ph (Mathematical Physics)
- physics:nucl-ex (Nuclear Experiment)
- physics:nucl-th (Nuclear Theory)
- physics:physics (Physics (Other))
- physics:quant-ph (Quantum Physics)
- q-bio (Quantitative Biology)
- q-fin (Quantitative Finance)
- stat (Statistics)

In the future

Holding Pen improvements

Integration of BibMatch as workflow component

Improve one-shot harvesting and testing

Edit OAI Source

Edit OAI source [x]

HARVESTING OPTIONS

Source name	<input type="text" value="arxiv"/>	e.g. arxiv
Base URL	<input type="text" value="http://export.arxiv.org/oai2"/>	
Meta-data prefix	<input type="text" value="arXiv"/>	e.g. oai_dc
Harvesting frequency	<input type="text" value="daily"/>	
Comments (optional)	<input type="text" value="testing"/>	

PRE-PROCESSING

(Leave all unchecked for no post-processing. * arguments are mandatory)

convert (c)
Cfg File *

download material (d)
Pdf Url
Tarball Url

additional workflow (f)
 match (m)

HOLDING PEN PROCESSES

extract plots (p)
Extraction Source * PDF
 LaTeX

extract references (r)
 extract authors (a)
 classify record
 manual keywording
 upload (u)
Priority
Name

Modify

SETS TO HARVEST

- cs (Computer Science)
- math (Mathematics)
- nlin (Nonlinear Sciences)
- physics (Physics)
- physics:astro-ph (Astrophysics)
- physics:cond-mat (Condensed Matter)
- physics:gr-qc (General Relativity and Quantum Cosmology)
- physics:hep-ex (High Energy Physics - Experiment)
- physics:hep-lat (High Energy Physics - Lattice)
- physics:hep-ph (High Energy Physics - Phenomenology)
- physics:hep-th (High Energy Physics - Theory)
- physics:math-ph (Mathematical Physics)
- physics:nucl-ex (Nuclear Experiment)
- physics:nucl-th (Nuclear Theory)
- physics:physics (Physics (Other))
- physics:quant-ph (Quantum Physics)
- q-bio (Quantitative Biology)
- q-fin (Quantitative Finance)
- stat (Statistics)



Ingestion Storage

Sort by TASK COMPLETED: **Harvested** **Converted** Refextracted Keyworded Matched STATUS: Running Pending Attention Completed Error

Source
Harvest date
Status

arXiv (hep-ex) - 2012-02-09 04:55:23

Filter by keywords

For selected:

Assign



<input type="checkbox"/>	Extreme Coronal Line Emitters: Tidal Disruption of Stars by Massive Black Holes in Galactic Nuclei? Completed: harvested, converted, matched	arXiv:YYYY.XXXX	RUNNING - Extracting References	5 minutes ago (2012-09-02 09:35:06)	jekyll
<input type="checkbox"/>	The X-ray reflector in NGC 4945: a time and space resolved portrait Completed: harvested, converted, matched, plotextracted	arXiv:YYYY.XXXX	WAITING - Next: Attach fulltext	25 minutes ago (2012-09-02 09:15:06)	
<input type="checkbox"/>	Cores in warm dark matter haloes: a Catch 22 problem Completed: harvested, converted, matched	arXiv:YYYY.XXXX	ATTENTION - Resolve match	25 minutes ago (2012-09-02 09:15:06)	romeo
<input type="checkbox"/>	Gravitational Fragmentation in Galaxy Mergers: A Stability Criteria Completed: harvested, converted, matched	arXiv:YYYY.XXXX	ERROR - Extracting plots	25 minutes ago (2012-09-02 09:15:06)	