The NASA Astrophysics Data System: Metadata Enrichment and Indexing

Edwin Henneken & Steve McDonald

ADS Users Group Meeting
January 17, 2017

Overview

- Enriching the collection
    - Linking to external data archives and full text (external & local)
    - Indexing SIMBAD objects, VizieR Catalogs, NED
    - Graphics
    - Usage data (“Co-Reads”)
    - References and Citations
    - Policy (arXiv vs. published)
    - Associated records

Enriching the collection

THE ASTRONOMICAL JOURNAL, 124:266–293, 2002 July

DETAILED STRUCTURAL DECOMPOSITION OF GALAXY IMAGES[1]

CHIEN Y. PENG,[2] LUIS C. HO,[3] CHRIS D. IMPEY,[2] AND HANS-WALTER RIX[4]
Received 2001 August 10; accepted 2002 March 27

ABSTRACT

We present a two-dimensional fitting algorithm (GALFIT) designed to extract structural components from galaxy images, with emphasis on closely modeling light profiles of spatially well-resolved, nearby galaxies observed with the *Hubble Space Telescope*. Our algorithm improves on previous techniques in two areas: by being able to simultaneously fit a galaxy with an arbitrary number of components and with optimization in computation speed, suited for working on large galaxy images. We use two-dimensional models such as the "Nuker" law, the Sérsic (de Vaucouleurs) profile, an exponential disk, and Gaussian or Moffat functions. The azimuthal shapes are generalized ellipses that can fit disky and boxy components. Some potential applications of our program include: standard modeling of global galaxy profiles; extracting bars, stellar disks, double nuclei, and compact nuclear sources; and measuring absolute dust extinction or surface brightness fluctuations after removing the galaxy model. When examined in detail, we find that even simple looking galaxies generally require at least three components to be modeled accurately, rather than the one or two components more often employed. Many galaxies with complex isophotes, ellipticity changes, and position angle twists can be modeled accurately in two dimensions. We illustrate this by way of 11 case studies, which include regular and barred spiral galaxies, highly disky lenticular galaxies, and elliptical galaxies displaying various levels of complexities. A useful extension of this algorithm is to accurately extract nuclear point sources in galaxies. We compare two-dimensional and one-dimensional extraction techniques on simulated images of galaxies having nuclear slopes with different degrees of cuspiness, and we then illustrate the application of the program to several examples of nearby galaxies with weak nuclei.

Key words: galaxies: bulges — galaxies: fundamental parameters — galaxies: nuclei — galaxies: structure — techniques: image processing — techniques: photometric

1. INTRODUCTION

Galaxies span a wide range of morphology and luminosity, and a very useful way to quantify them is to fit their light distribution with parametric functions. The de Vaucouleurs $R^{1/4}$ and exponential disk functions became standard functions to use after de Vaucouleurs (1948) showed many elliptical galaxies to have $R^{1/4}$ light distributions, while Freeman (1970) found later-type galaxies to be well described by a de Vaucouleurs bulge plus an exponential disk. Since then the empirical techniques of galaxy fitting and decomposition have led to a number of notable advances in understanding galaxy formation and evolution. These include investigations into the Tully-Fisher relationship (Tully & Fischer 1977), the fundamental plane of spheroids (Faber et al. 1987; Dressler et al. 1987; Djorgovski & Davis 1987; Bender, Burstein, & Faber 1992), the morphological transformation of galaxies in cluster environments (e.g., Dressler 1980; van Dokkum & Franx 2001), the bimodality of galaxy nuclear cusps (Lauer et al. 1995; Faber et al. 1997) and its implications for the formation of massive black holes (Ravindranath, Ho, & Filippenko 2002), and the cosmic evolution of galaxy morphology (e.g., Lilly et al. 1998; Marleau & Simard 1998).

There are two general types of galaxy fitting: one-dimensional fitting of surface brightness profiles (e.g., Kormendy 1977; Burstein 1979; Boroson 1981; Kent 1985; Baggett, Baggett, & Anderson 1998) and two-dimensional fitting of galaxy images (e.g., Shaw & Gilmore 1989; Byun & Freeman 1995; de Jong 1996; Simard 1998; Wadadekar, Robbason, & Kembhavi 1999; Khosroshahi, Wadadekar, & Kembhavi 2000), each with its own tradeoffs and benefits.

In one dimension an important consideration is how to first obtain a radial surface brightness profile from a two-dimensional image, for which there is no universally agreed upon procedure. A common practice is to use isophote fitting, which is a powerful technique when performed on well-resolved galaxies because it averages over elliptical annuli to increase the signal-to-noise ratio (S/N) at a given radius. However, as many galaxies have isophote twists and changing ellipticity as a function of radius, the galaxy profile is extracted along a radial arc that is ill defined. An alternative approach is to use a direct one-dimensional slice across an image. Burstein (1979) argues that only cuts along the major axis should be used in bulge-to-disk (B/D) decompositions. Meanwhile, Ferrarese et al. (1994) point out that galaxies with power-law central profiles may have different profiles along the major and minor axis.

Fitting profiles in one dimension is frequently used because it suffices for certain goals and is simple to implement. But many studies now resort to two-dimensional techniques. For B/D decompositions a number of authors (e.g., Byun & Freeman 1995; Wadadekar et al. 1999) have

[1] Based on observations with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc., under NASA contract NAS 5-26555.

[2] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721; cyp, cimpey@as.arizona.edu.
[3] The Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101; lho@ociw.edu.
[4] Max-Planck-Institut für Astronomie, Königstuhl 17, Heidelberg, D-69117, Germany.

Detailed Structural Decomposition of Galaxy Images

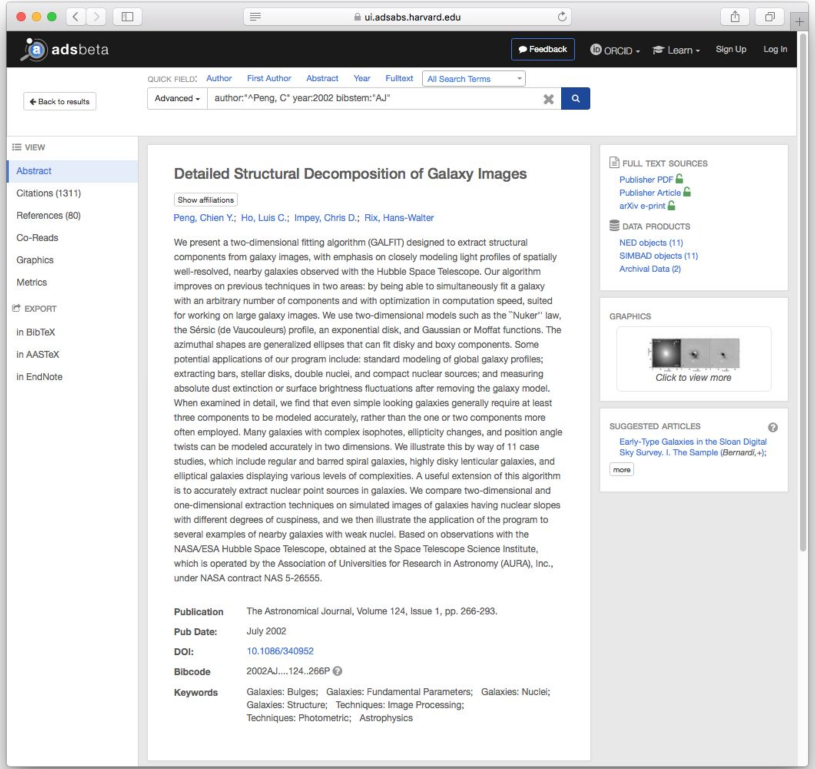
Show affiliations

Peng, Chien Y.; Ho, Luis C.; Impey, Chris D.; Rix, Hans-Walter

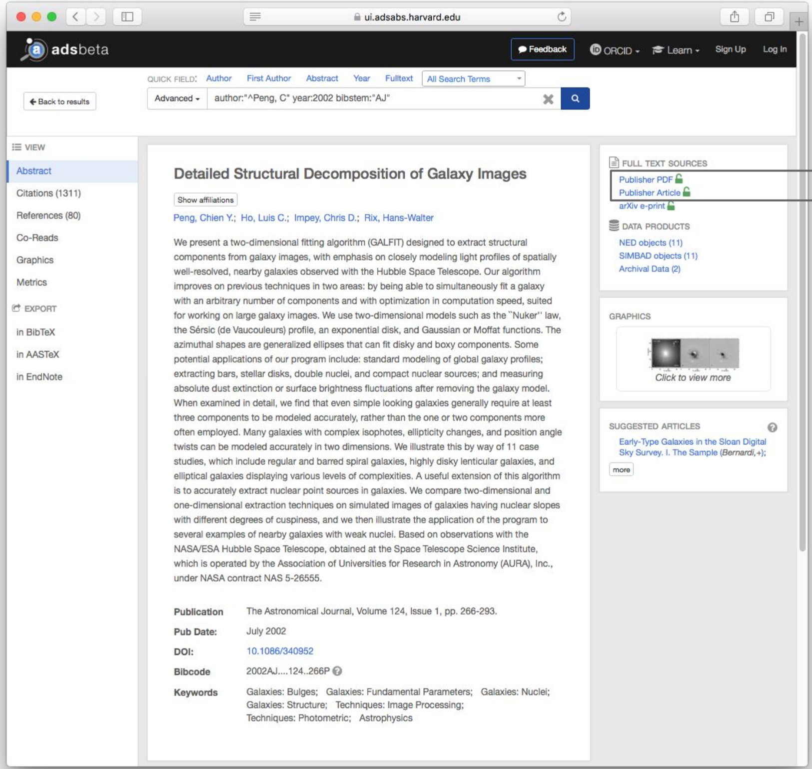
We present a two-dimensional fitting algorithm (GALFIT) designed to extract structural components from galaxy images, with emphasis on closely modeling light profiles of spatially well-resolved, nearby galaxies observed with the Hubble Space Telescope. Our algorithm improves on previous techniques in two areas: by being able to simultaneously fit a galaxy with an arbitrary number of components and with optimization in computation speed, suited for working on large galaxy images. We use two-dimensional models such as the `Nuker'' law, the Sérsic (de Vaucouleurs) profile, an exponential disk, and Gaussian or Moffat functions. The azimuthal shapes are generalized ellipses that can fit disky and boxy components. Some potential applications of our program include: standard modeling of global galaxy profiles; extracting bars, stellar disks, double nuclei, and compact nuclear sources; and measuring absolute dust extinction or surface brightness fluctuations after removing the galaxy model. When examined in detail, we find that even simple looking galaxies generally require at least three components to be modeled accurately, rather than the one or two components more often employed. Many galaxies with complex isophotes, ellipticity changes, and position angle twists can be modeled accurately in two dimensions. We illustrate this by way of 11 case studies, which include regular and barred spiral galaxies, highly disky lenticular galaxies, and elliptical galaxies displaying various levels of complexities. A useful extension of this algorithm is to accurately extract nuclear point sources in galaxies. We compare two-dimensional and one-dimensional extraction techniques on simulated images of galaxies having nuclear slopes with different degrees of cuspiness, and we then illustrate the application of the program to several examples of nearby galaxies with weak nuclei. Based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc., under NASA contract NAS 5-26555.

Publication	The Astronomical Journal, Volume 124, Issue 1, pp. 266-293.
Pub Date:	July 2002
DOI:	10.1086/340952
Bibcode	2002AJ....124..266P
Keywords	Galaxies: Bulges; Galaxies: Fundamental Parameters; Galaxies: Nuclei; Galaxies: Structure; Techniques: Image Processing; Techniques: Photometric; Astrophysics

Enriching the collection

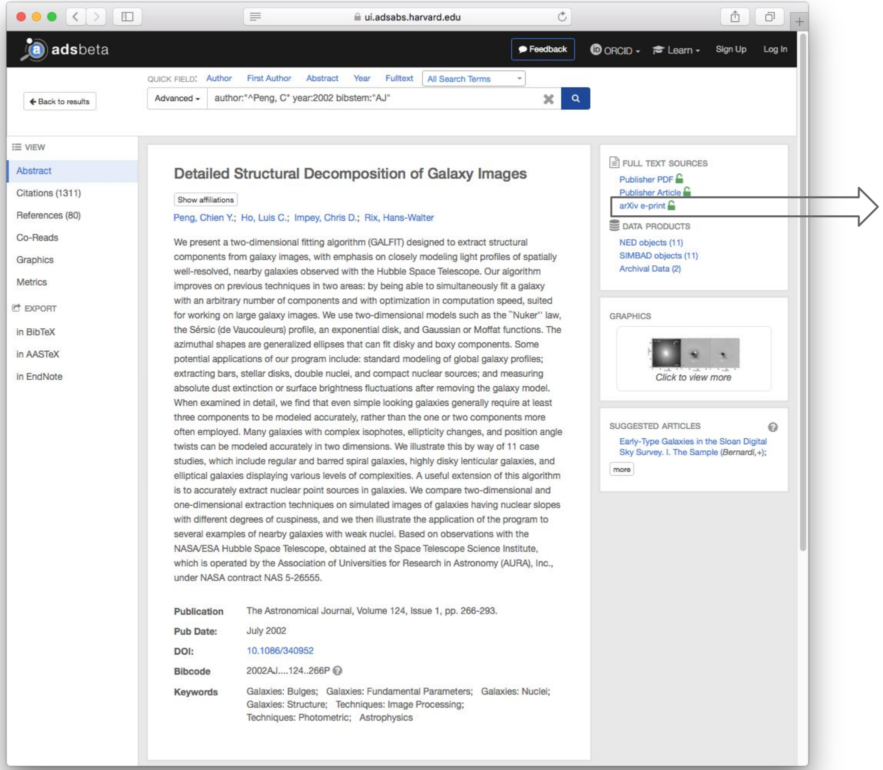


Enriching the collection - Linking: Full Text



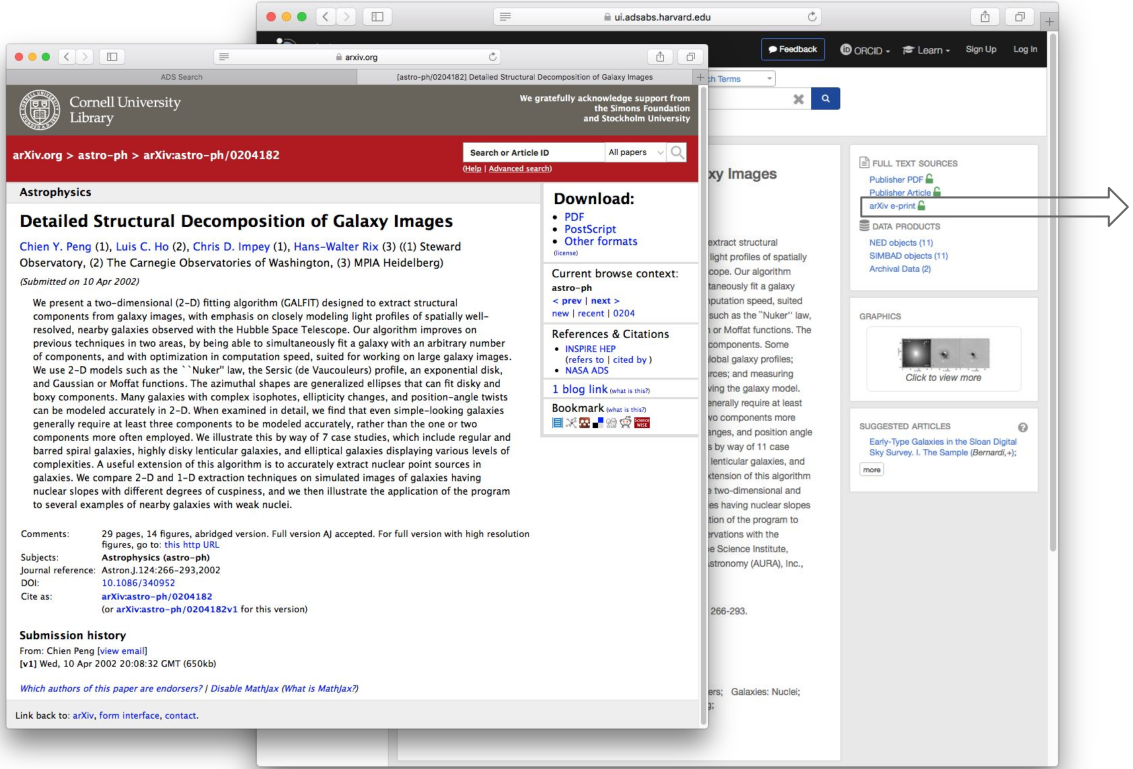
Publisher Full Text

Enriching the collection - Linking: Full Text



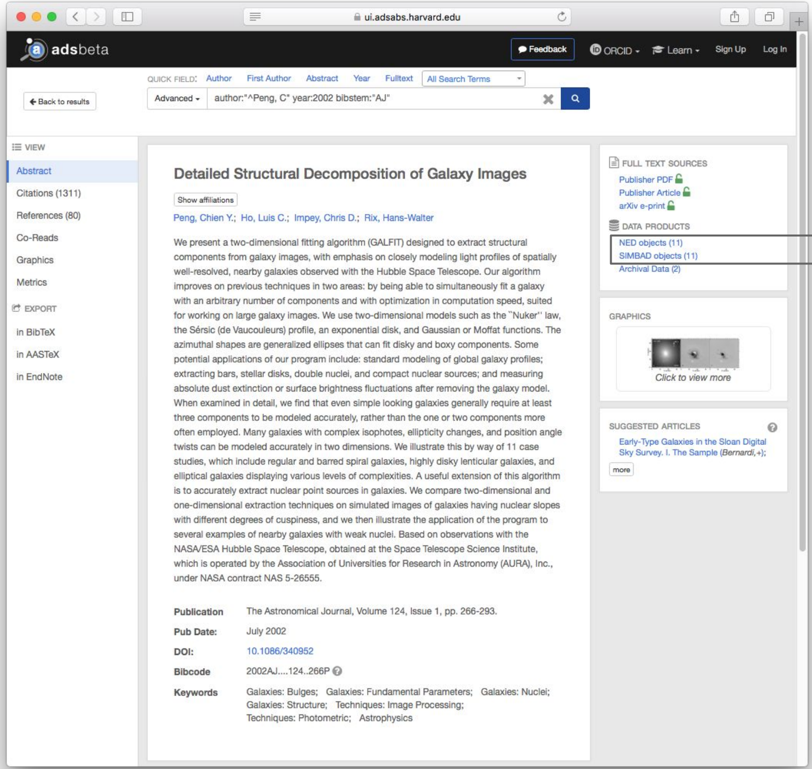
arXiv match

Enriching the collection - Linking: Full Text



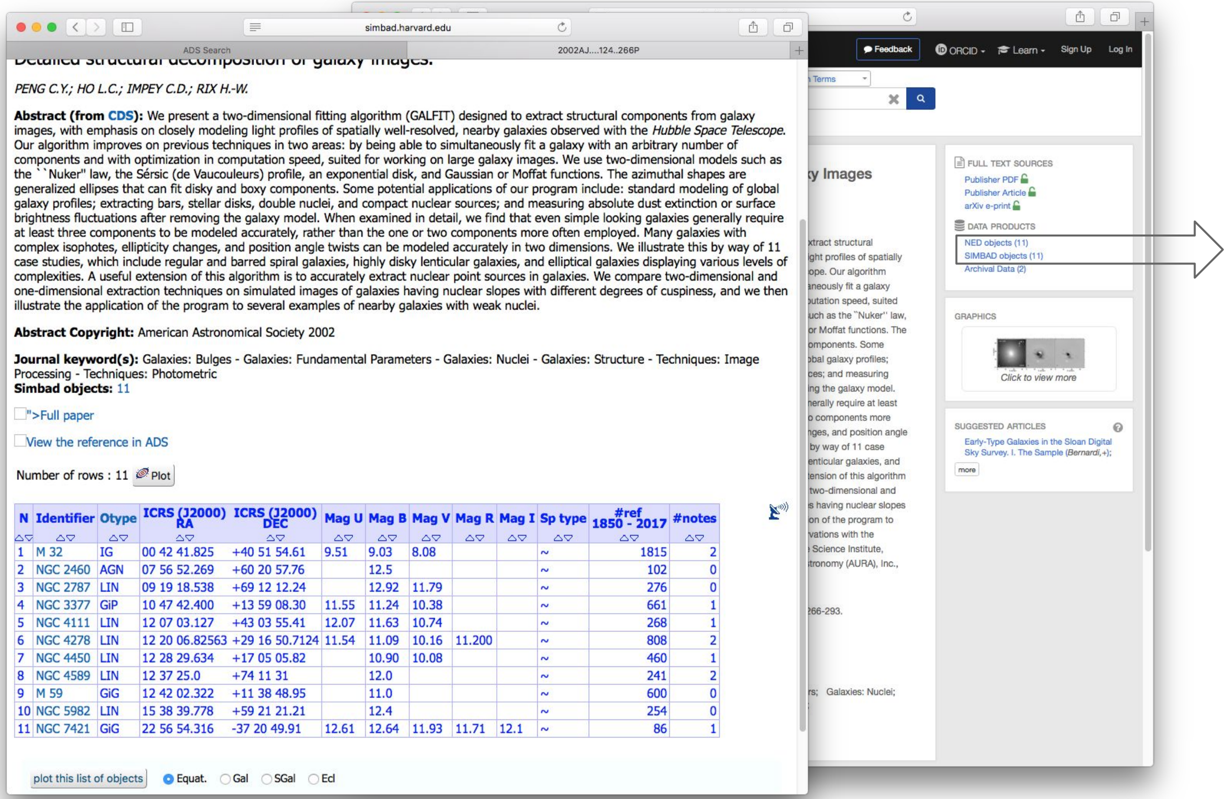
arXiv match

Enriching the collection - Linking: External Data



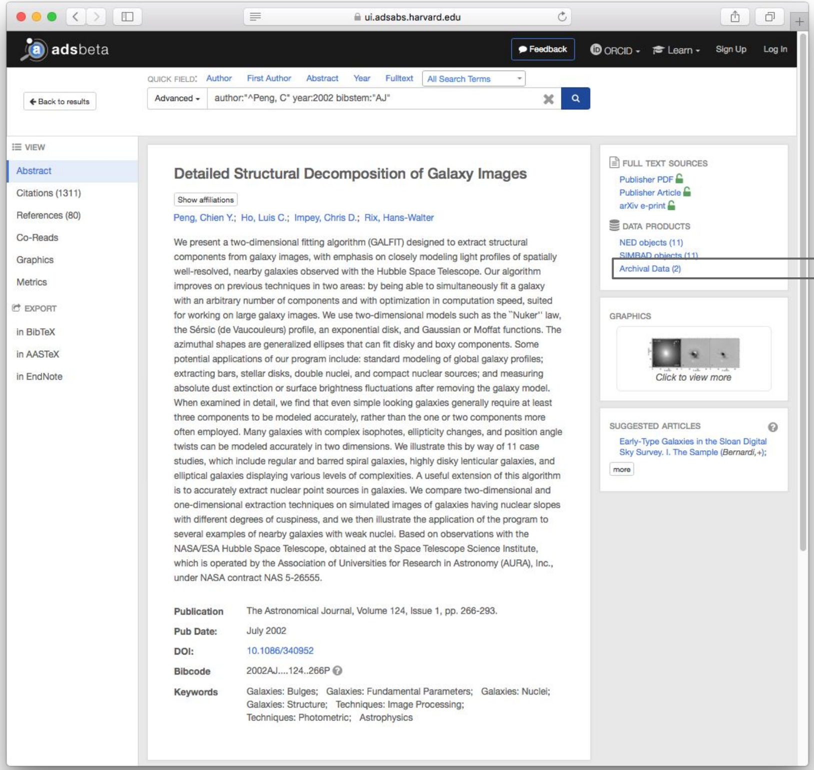
Astronomical Databases
(SIMBAD, NED)

Enriching the collection - Linking: External Data



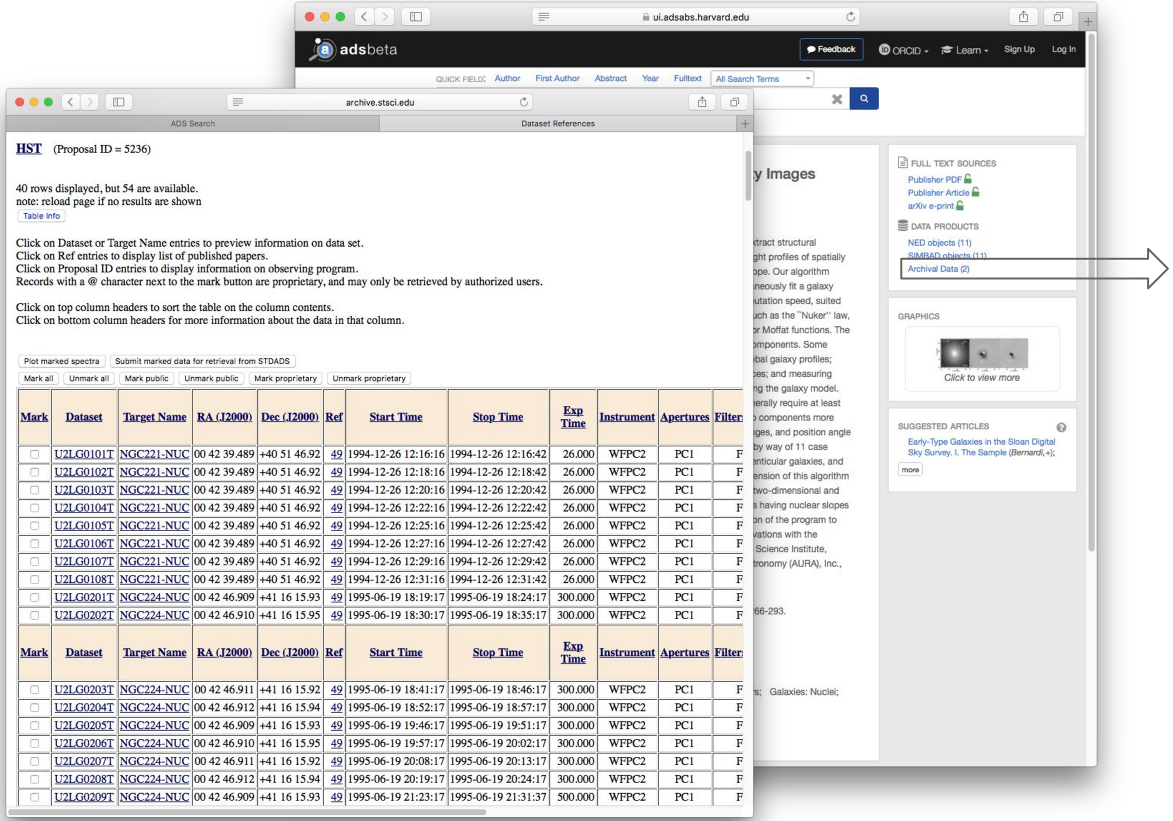
Astronomical Databases
(SIMBAD, NED)

Enriching the collection - Linking: External Data



Data Archives

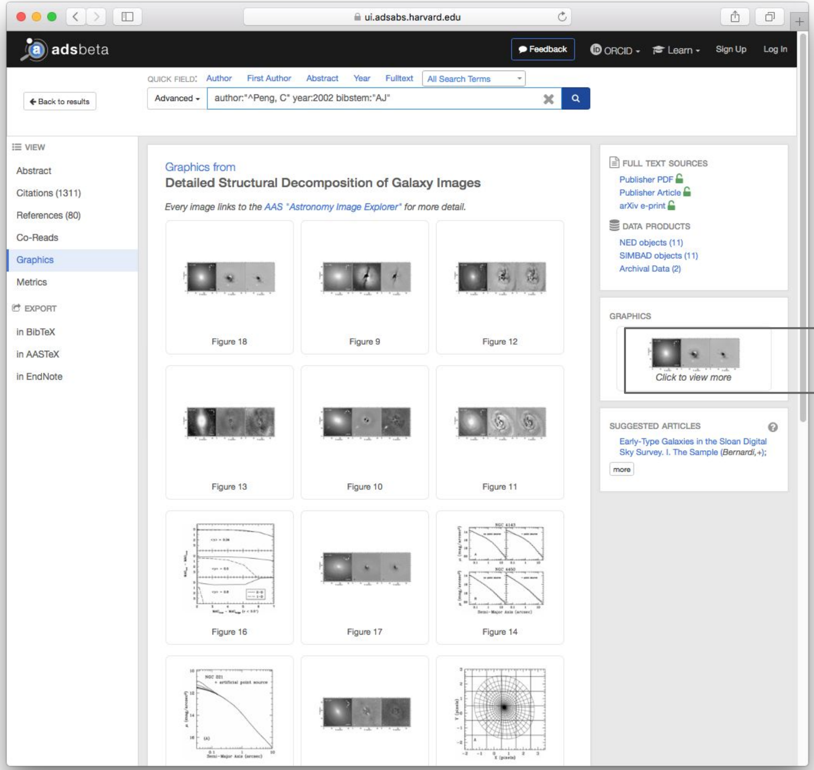
Enriching the collection - Linking: External Data



Mark	Dataset	Target Name	RA (J2000)	Dec (J2000)	Ref	Start Time	Stop Time	Exp Time	Instrument	Apertures	Filters
	U2LG0101T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:16:16	1994-12-26 12:16:42	26.000	WFPC2	PC1	F
	U2LG0102T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:18:16	1994-12-26 12:18:42	26.000	WFPC2	PC1	F
	U2LG0103T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:20:16	1994-12-26 12:20:42	26.000	WFPC2	PC1	F
	U2LG0104T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:22:16	1994-12-26 12:22:42	26.000	WFPC2	PC1	F
	U2LG0105T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:25:16	1994-12-26 12:25:42	26.000	WFPC2	PC1	F
	U2LG0106T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:27:16	1994-12-26 12:27:42	26.000	WFPC2	PC1	F
	U2LG0107T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:29:16	1994-12-26 12:29:42	26.000	WFPC2	PC1	F
	U2LG0108T	NGC221-NUC	00 42 39.489	+40 51 46.92	49	1994-12-26 12:31:16	1994-12-26 12:31:42	26.000	WFPC2	PC1	F
	U2LG0201T	NGC224-NUC	00 42 46.909	+41 16 15.93	49	1995-06-19 18:19:17	1995-06-19 18:24:17	300.000	WFPC2	PC1	F
	U2LG0202T	NGC224-NUC	00 42 46.910	+41 16 15.95	49	1995-06-19 18:30:17	1995-06-19 18:35:17	300.000	WFPC2	PC1	F
Mark	Dataset	Target Name	RA (J2000)	Dec (J2000)	Ref	Start Time	Stop Time	Exp Time	Instrument	Apertures	Filters
	U2LG0203T	NGC224-NUC	00 42 46.911	+41 16 15.92	49	1995-06-19 18:41:17	1995-06-19 18:46:17	300.000	WFPC2	PC1	F
	U2LG0204T	NGC224-NUC	00 42 46.912	+41 16 15.94	49	1995-06-19 18:52:17	1995-06-19 18:57:17	300.000	WFPC2	PC1	F
	U2LG0205T	NGC224-NUC	00 42 46.909	+41 16 15.93	49	1995-06-19 19:46:17	1995-06-19 19:51:17	300.000	WFPC2	PC1	F
	U2LG0206T	NGC224-NUC	00 42 46.910	+41 16 15.95	49	1995-06-19 19:57:17	1995-06-19 20:02:17	300.000	WFPC2	PC1	F
	U2LG0207T	NGC224-NUC	00 42 46.911	+41 16 15.92	49	1995-06-19 20:08:17	1995-06-19 20:13:17	300.000	WFPC2	PC1	F
	U2LG0208T	NGC224-NUC	00 42 46.912	+41 16 15.94	49	1995-06-19 20:19:17	1995-06-19 20:24:17	300.000	WFPC2	PC1	F
	U2LG0209T	NGC224-NUC	00 42 46.909	+41 16 15.93	49	1995-06-19 21:23:17	1995-06-19 21:31:37	500.000	WFPC2	PC1	F

Data Archives

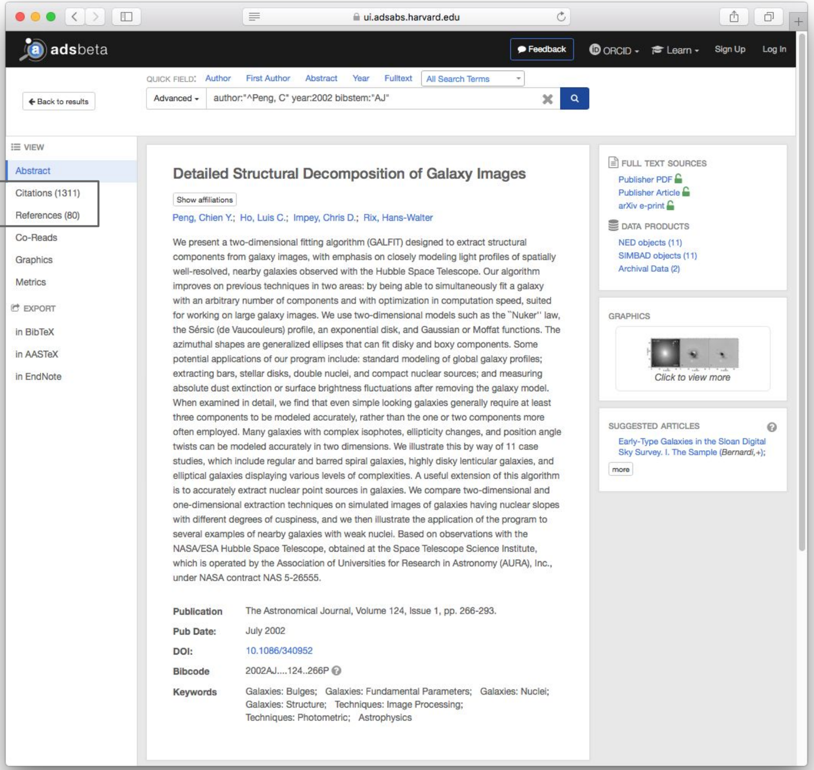
Enriching the collection - Article Graphics



Thumbnail View Figures

Enriching the collection - Citation Network

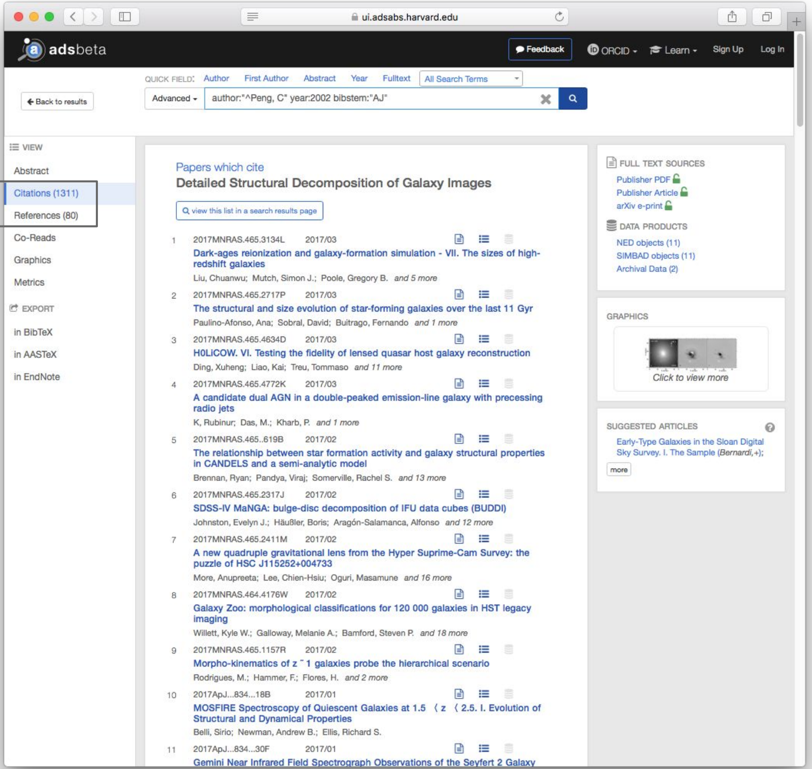
Citations/References



Enriching the collection - Citation Network

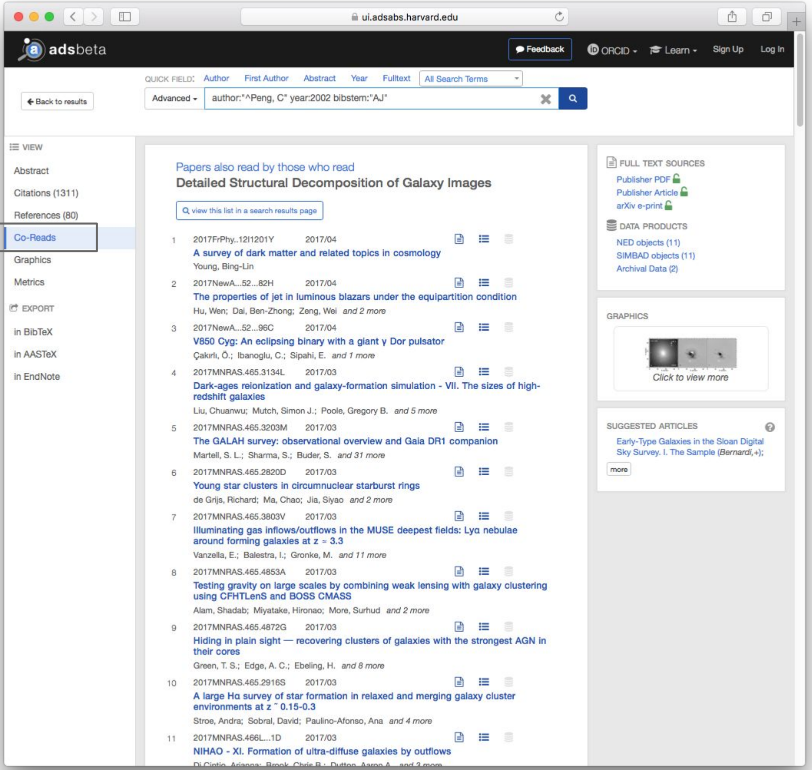
Enriching the collection - Citation Network

Citations/References



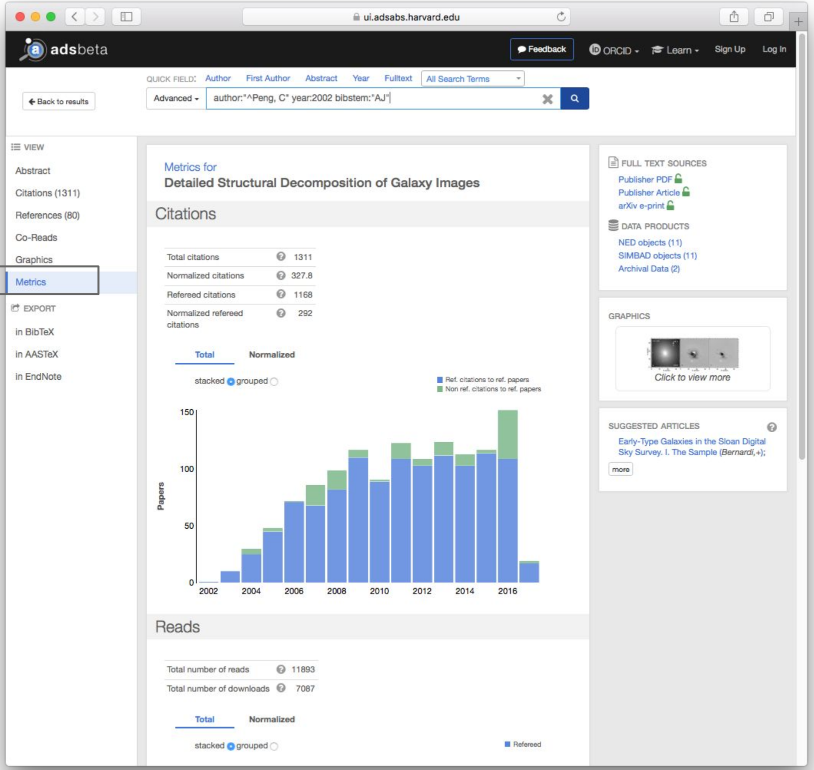
Enriching the collection - Usage Data (“Co-Reads”)

Co-Reads (ADS/arXiv)



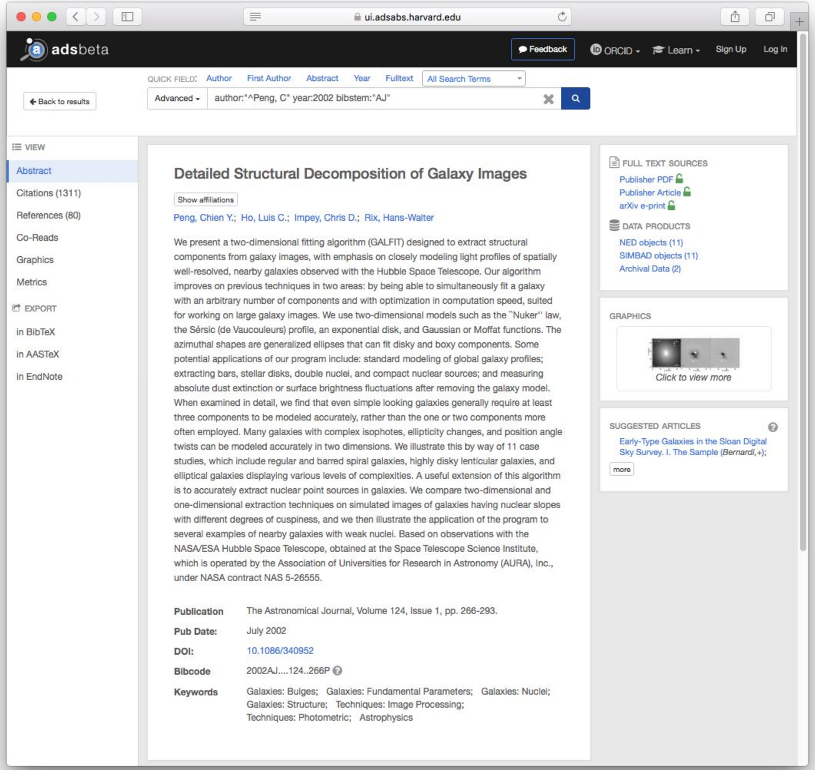
Enriching the collection - Basic Metrics (usage, citations)

Basic Metrics



Enriching the collection - Associated Information

- Errata
- Addenda
- Series
- arXiv matches
- ...



ETL: Extract, Transform And Load

Data Ingest To Production Servers

Ingest: Extract, Transform, Load

Bibliographic

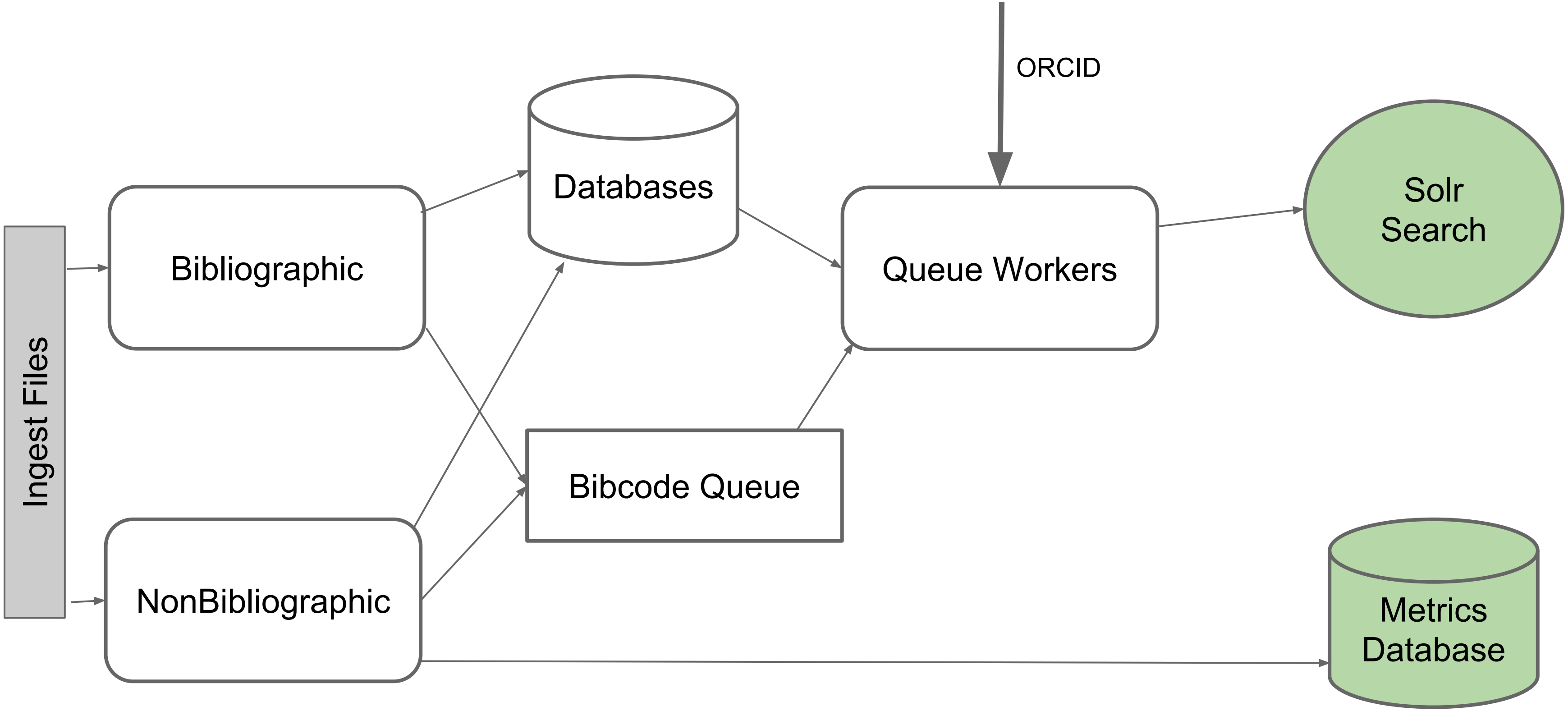
- From Publishers
- Variety Of Formats

Non-Bibliographic

- Refereed Status, Read Counts, Citation Network
- Bibliographic Groups, Data Links, SIMBAD Objects
- arXiv.org, ADS Usage

	Classic	**Bumblebee**
Source Content	Metadata,References,Abstract(bib) ASCII tables (nonbib)	Metadata, References, **Full-text** ADS Classic indexes (nonbib)
Back-end	Flat files, ASCII tables	**SQL and MongoDB**
Index	Author, Title, Abstract, Keywords; 30 GB in size	**~50 fields** including full-text, affiliations,**ORCIDs**; **400 GB** in size
Code	300K lines of home-grown C / PERL code, single threaded	Open source stack: **SOLR**, **SQL**, **RabbitMQ**, **Docker** containers
Performance	48 hours required to reindex all data, 12 hours to mirror	**3** hours to Metrics, 48 to Solr
Features	Custom tagged files format, ASCII index, limited phrase search	API, phrase and positional search, facets, scalable metrics, citations

Improved Ingest Pipeline



Improving Ingest

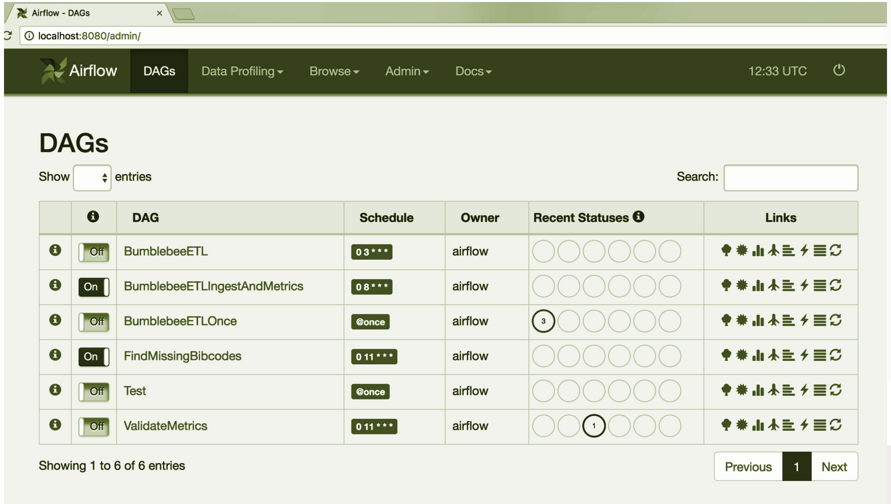
Performance: MongoDB To Postgres

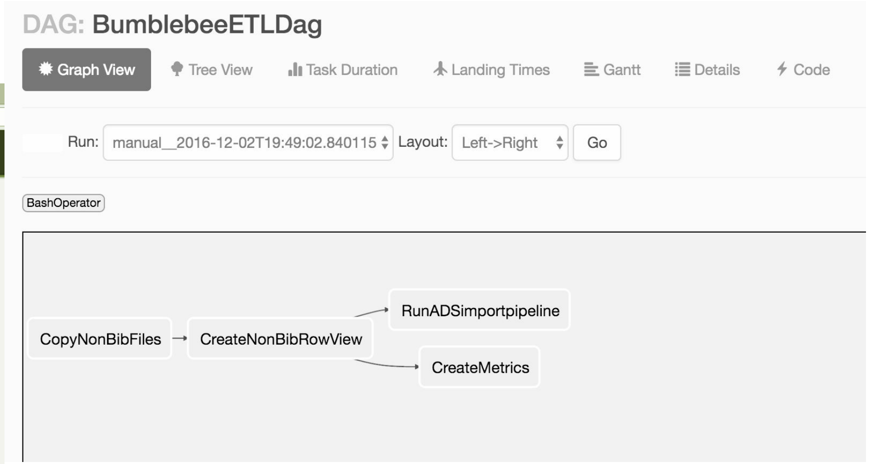
Reliability: Single Threaded

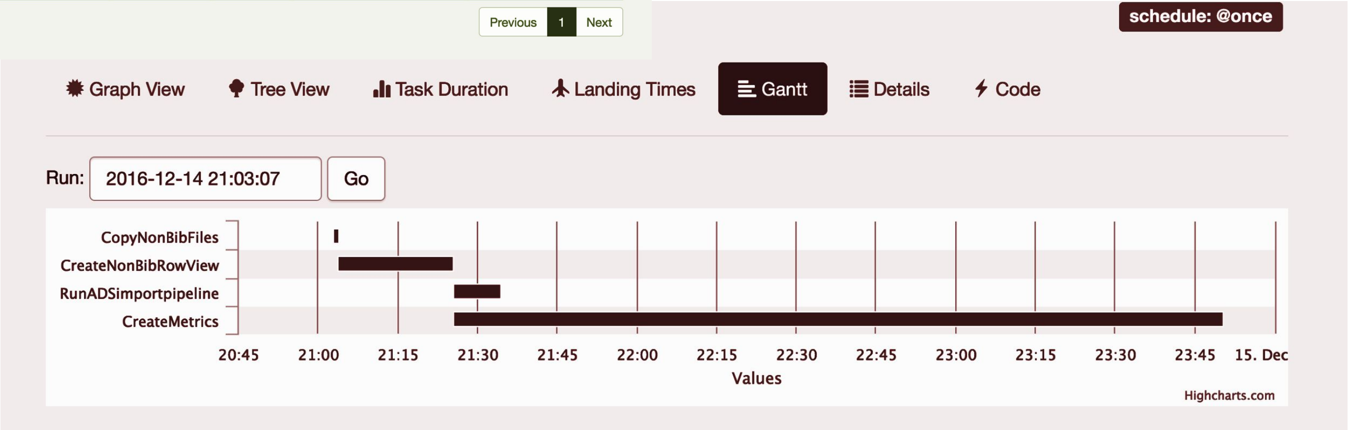
Deployability: Reuse Bumblebee Microservice Framework

Visibility: Manage With Airflow

Airflow


Airflow - DAGs
localhost:8080/admin/
Airflow
DAGs
Data Profiling
Browse
Admin
Docs
12:33 UTC
DAGs
Show entries
Search:
DAG
Schedule
Owner
Recent Statuses
Links
Off
BumblebeeETL
0 3 * * *
airflow
On
BumblebeeETLIngestAndMetrics
0 8 * * *
airflow
Off
BumblebeeETLOnce
@once
airflow
On
FindMissingBibcodes
0 11 * * *
airflow
Off
Test
@once
airflow
Off
ValidateMetrics
0 11 * * *
airflow
Showing 1 to 6 of 6 entries
Previous
1
Next


DAG: BumblebeeETLDag
Graph View
Tree View
Task Duration
Landing Times
Gantt
Details
Code
Run:
manual__2016-12-02T19:49:02.840115
Layout:
Left->Right
Go
BashOperator
CopyNonBibFiles
CreateNonBibRowView
RunADSimportpipeline
CreateMetrics


schedule: @once
Graph View
Tree View
Task Duration
Landing Times
Gantt
Details
Code
Run:
2016-12-14 21:03:07
Go
CopyNonBibFiles
CreateNonBibRowView
RunADSimportpipeline
CreateMetrics
20:45
21:00
21:15
21:30
21:45
22:00
22:15
22:30
22:45
23:00
23:15
23:30
23:45
15. Dec
Values
Highcharts.com

Thank you!
Q&A

Overview

- Extraction / Transform / Load
  - Content being indexed
  - Technology stack: ADS Classic vs. new architecture
  - Pipeline details