The Industrial Revolution of Galaxies Cluster Finding in the Dark Energy Survey

Ricardo Ogando
ogando@linea.gov.br

Observatório Nacional
Centro Brasileiro de Pesquisas Físicas
Laboratório Nacional de Computação Científica
FINEP FAPERJ CNPq

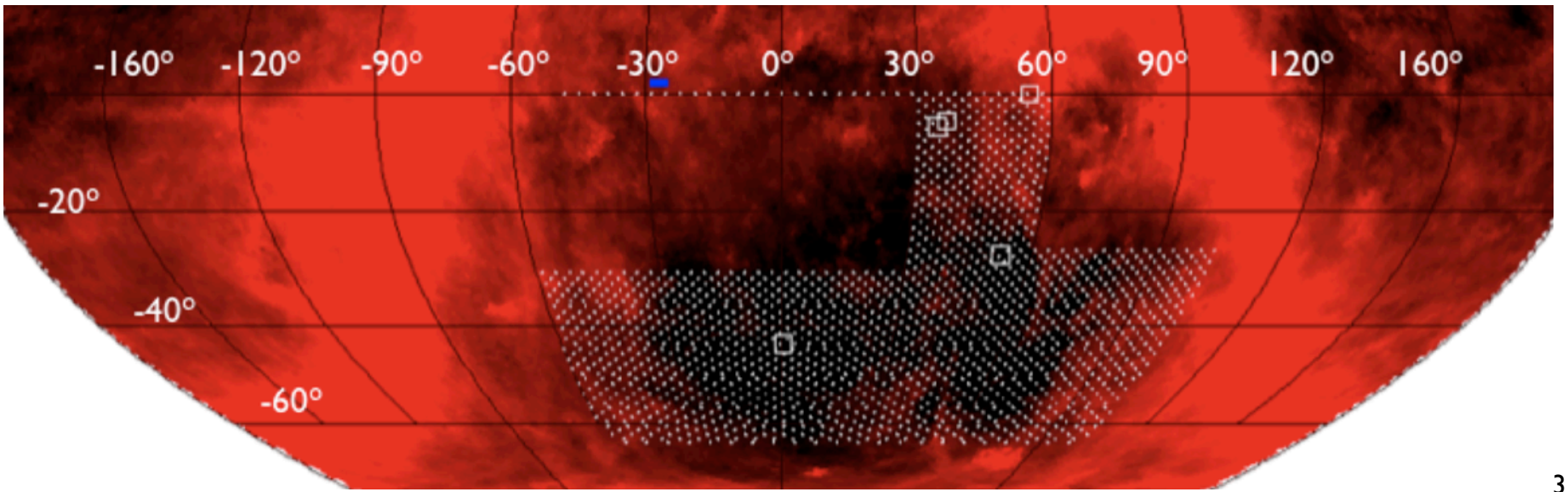
Outline

- Introduction
- Cluster Finders Industrialization
- Data handling
- Parallelization
- Results
- Scientific collaboration network
- Fine tuning and Cluster finding challenge
- Lessons learned

2

Introduction

- Dark Energy Survey 5000 square degrees
- 400 million galaxies spread on the cosmic web
- Four probes to cosmology: SN, BAO, WL and **Clusters**
- Several methods to find clusters of galaxies
- Understand their selection functions → better cosmology.



3

Cluster Finders Industrialization

- Run Cluster Finders (CFs) on **mock** catalogs
- Match **clusters** to **halos** in simulations
  - Calculate **completeness** and **purity**
- Have several CFs in a controlled environment
  - Keep track of data provenance, code version, and configurations
- Faster processing response time on data releases

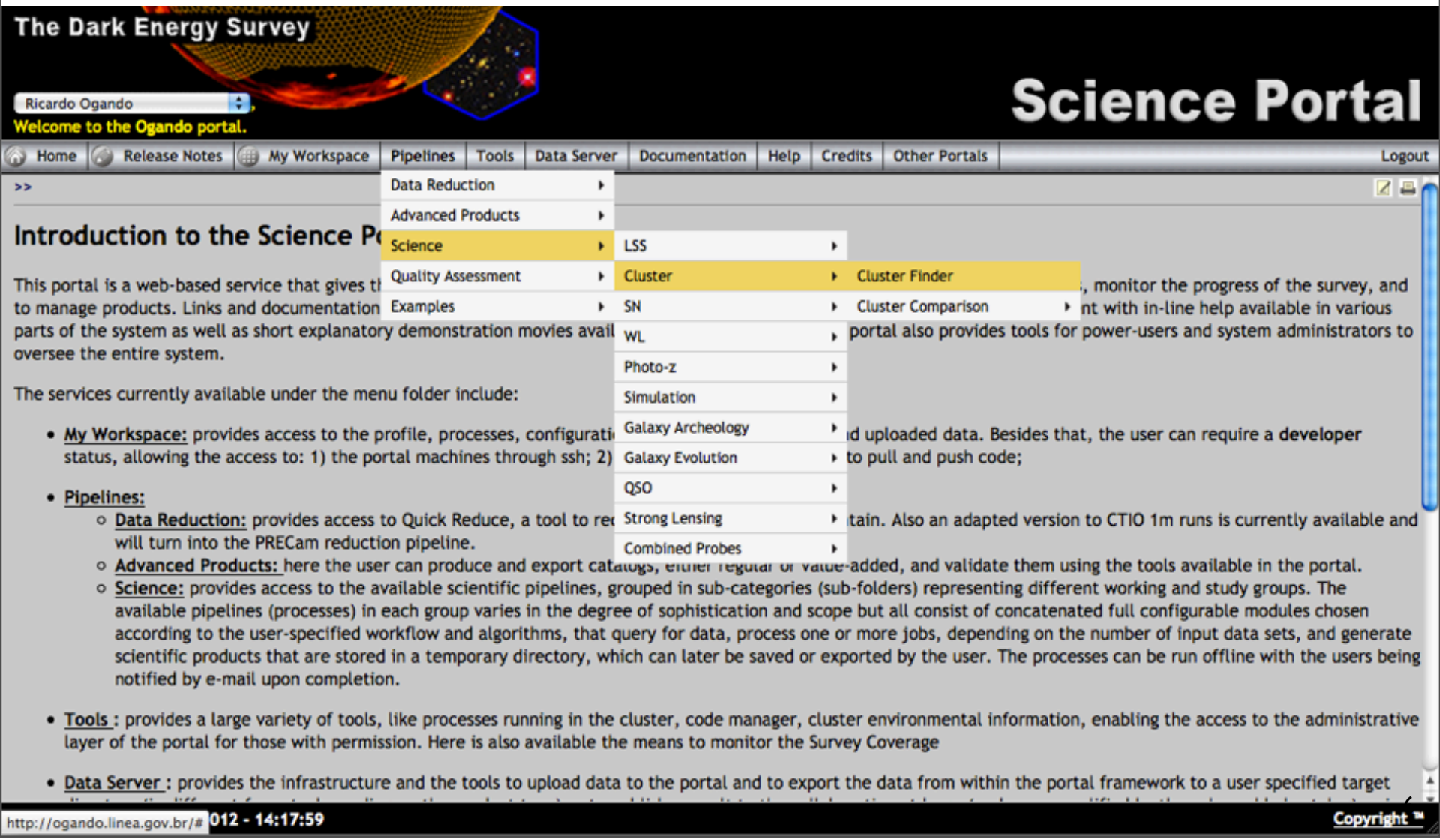
4

Cluster Finders Industrialization

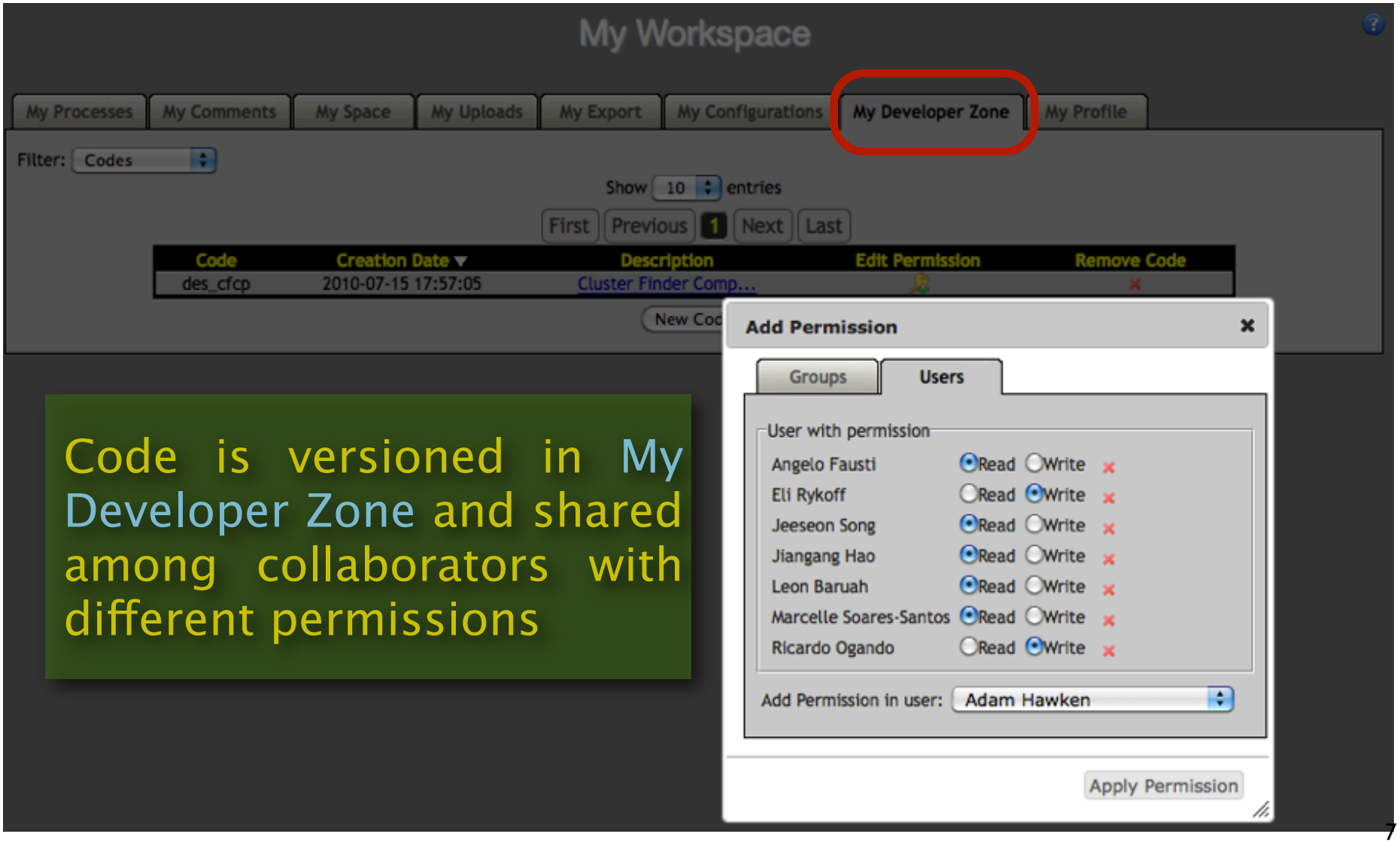
- **Matched Filter** (L. Olsen), **WAZP** (C. Benoist), VTRS (W. Barkhouse), GMBCG (J. Hao), Lambda (E. Rykoff), C4+ (L. Baruah), and zVT (M. Soares-Santos)
- **Framework** to compare results allowing:
  - **Fine tuning**. Finding the best set of CF input parameters, e.g. M* and threshold.
  - **Cluster Finding challenge**. Pros and cons of each CF method. Assess their selection functions.
- **Compliance**: CFs produce common output: RA, Dec, z, mass proxy rank

5

The Cluster Finder pipeline



Code versioning and share

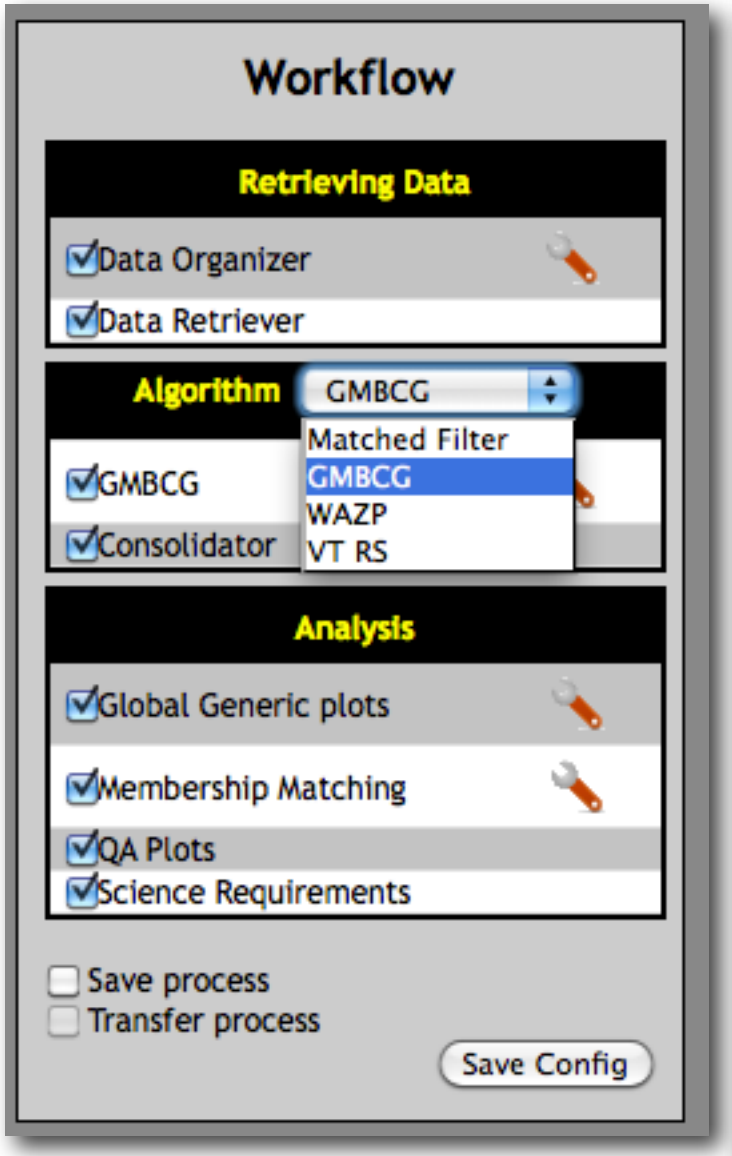


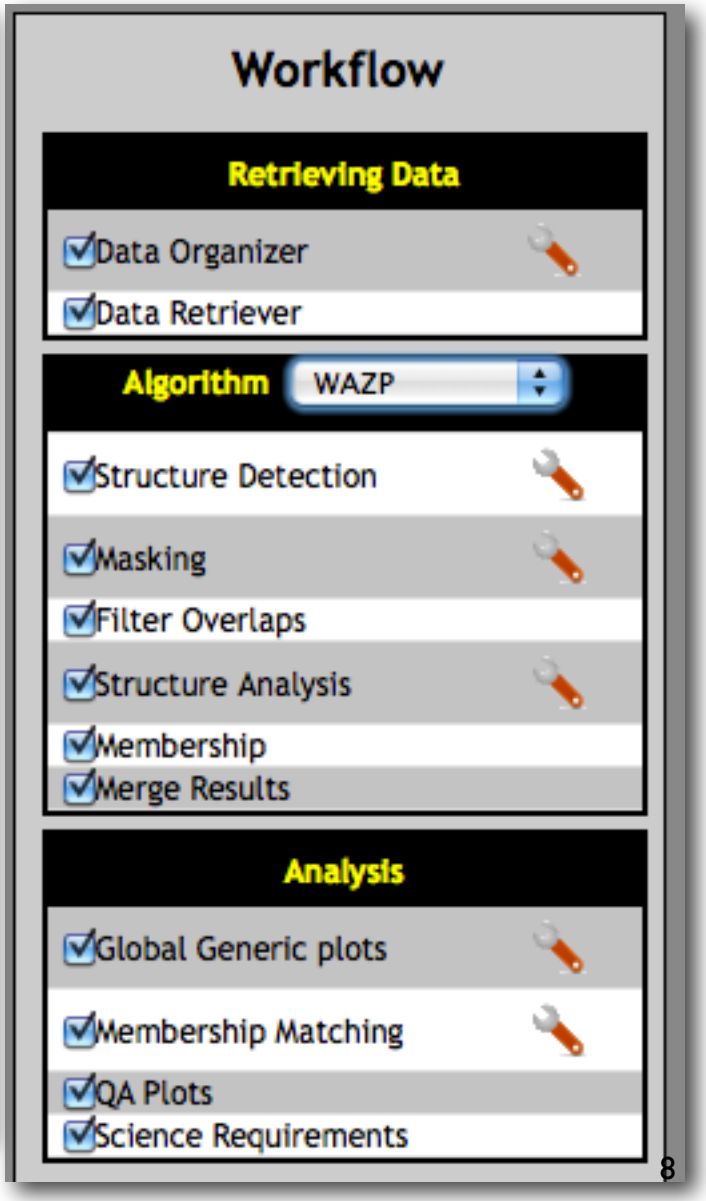
Code is versioned in My Developer Zone and shared among collaborators with different permissions

7

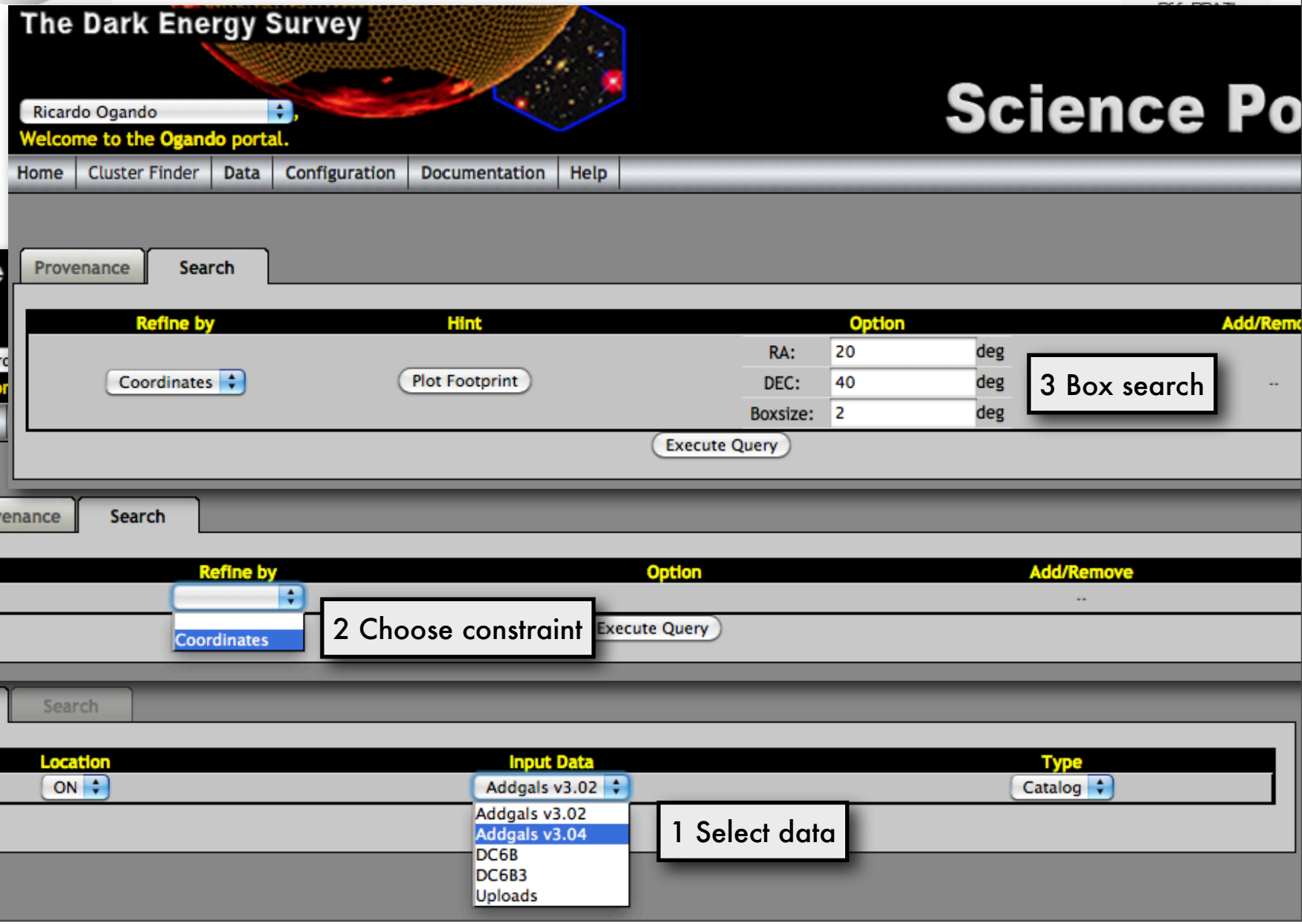
Workflow

- CF codes wrapped in **python** modules
- Modules concatenated in **XML** workflows





Data search


The Dark Energy Survey
Ricardo Ogando
Welcome to the Ogando portal.
Science Po
Home
Cluster Finder
Data
Configuration
Documentation
Help
Provenance
Search
Refine by
Hint
Option
Add/Remo
Coordinates
Plot Footprint
RA: 20 deg
DEC: 40 deg
Boxsize: 2 deg
3 Box search
Execute Query
Provenance
Search
Refine by
Option
Add/Remove
Coordinates
2 Choose constraint
Execute Query
Provenance
Search
Location
Input Data
Type
ON
Addgals v3.02
Catalog
Addgals v3.02
Addgals v3.04
DC6B
DC6B3
Uploads
1 Select data

Data handling

- ADDGALS: mock catalogs (200sqdeg)
- Data Challenges: mock observations (100sqdeg)
- Blind Cosmology Challenge: unknown cosmology mock catalog (5000sqdeg)
- Other data: CFHT, S82, and BOSS
- Ingested to PostgreSQL database.
- Q3C indexing for large astronomical catalogs

10

Provenance | Search | Selection

Found just one page.

File Name	Upload Date	File Type	File Class	Uploaded by
redmap_v2.13	2011-06-28 12:45:04.847170	application/fits-table	science	Jeeseon Song
lambda220	2011-06-17 22:49:18.354475	application/fits-table	science	Jeeseon Song
testcat	2011-06-06 14:07:04.167049	application/fits-table	science	Jeeseon Song
DES_truth_2n3n4.fit	2011-06-05 17:18:04.953233	application/fits-table	science	Jeeseon Song
DES_truth_2n3n4.fit	2011-06-05 17:18:04.952156	application/fits-table	science	Jeeseon Song
DES_truth_2n3n4.fit	2011-06-05 17:18:04.954176	application/fits-table	science	Jeeseon Song
Lindsey_total	2011-04-23 22:29:51.838825	application/fits-table	science	Jeeseon Song
Lindsey_new	2011-04-21 00:34:59.220673	application/fits-table	science	Jeeseon Song

Provenance | Search | Selection

File Name	Uploaded by
C4 v2.11 final catalog	Leon Baruah
C4 v2.11 test catalog	Leon Baruah

Select All Un

CFOs can also upload their cluster catalogs

Provenance | Search

Found just one page.

File Name	Upload Date	File Type	File Class	Uploaded by
rykoff_v2.13_matched_bgerke	2011-06-02 12:38:02.801418	application/fits-table	science	Brian Gerke
hao_gauss_v211_testing	2011-03-16 17:01:01.627407	application/fits-table	science	Brian Gerke
rykoff_v2.13_clusters	2011-02-17 20:32:08.078426	application/fits-table	science	Brian Gerke
baruah_v2.11_matched	2011-02-17 17:34:44.430108	application/fits-table	science	Brian Gerke
rykoff_v2.11_matched	2011-02-16 00:44:01.064198	application/fits-table	science	Brian Gerke

Provenance | Search | Selection

Found just one page.

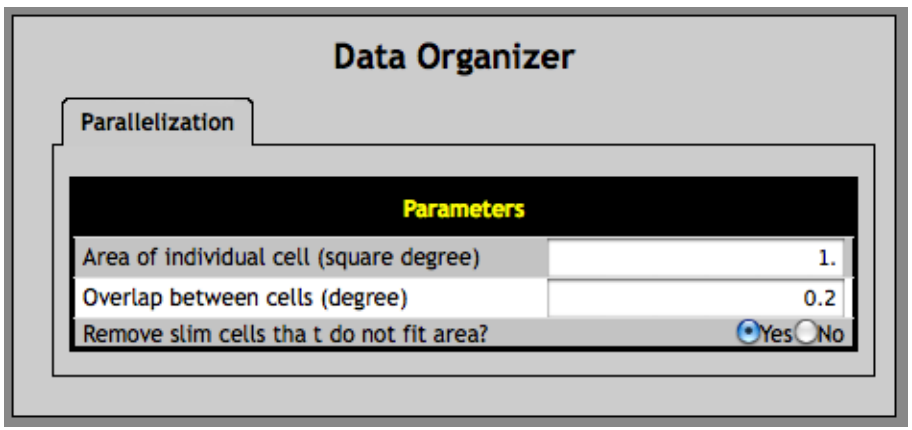
File Name	Upload Date	File Type	File Class	Uploaded by
S82_redMaPPer_v3.4_nord2	2011-09-10 19:02:37.614415	application/fits-table	science	Eli Rykoff
S82_redMaPPer_v3.4_nord	2011-09-10 17:20:22.655624	application/fits-table	science	Eli Rykoff
S82_redMaPPer_v3.4_like	2011-09-10 17:03:26.664036	application/fits-table	science	Eli Rykoff
Modprop Halos	2011-09-09 12:32:44.750945	application/fits-table	science	Eli Rykoff
Constant Halos	2011-09-09 12:30:54.903871	application/fits-table	science	Eli Rykoff
Bowl Halos	2011-09-09 12:30:11.745120	application/fits-table	science	Eli Rykoff
Bowl-gr Halos	2011-09-09 12:29:50.924805	application/fits-table	science	Eli Rykoff
Baseline Halos	2011-09-08 13:29:26.451147	application/fits-table	science	Eli Rykoff

Select All | Unselect All | Select All Pages | Unselect All Pages | There are 8 items returned.

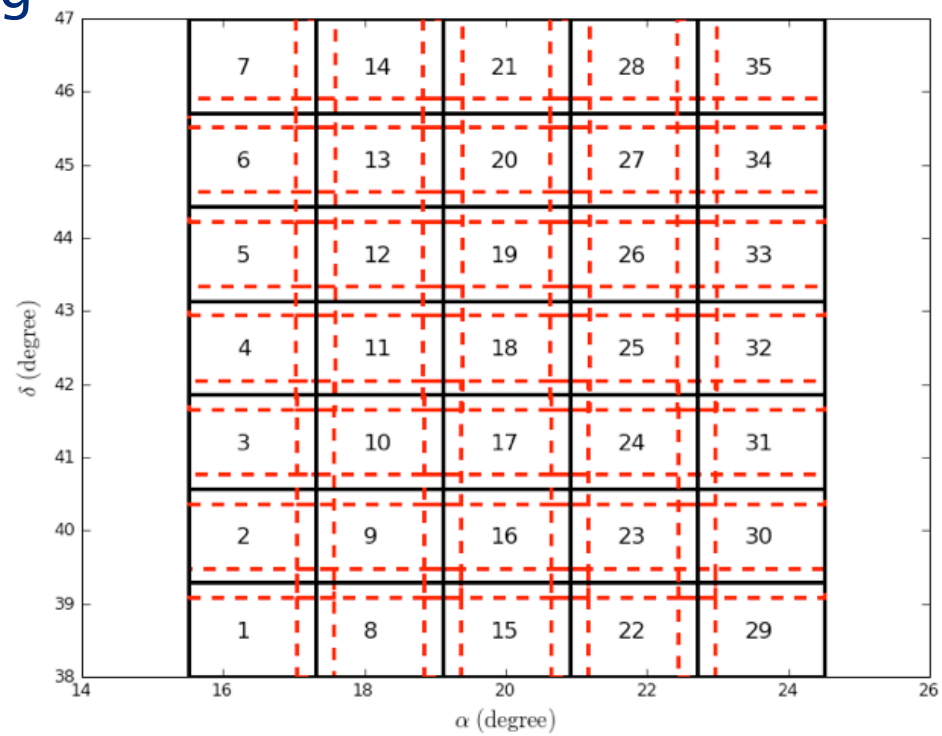
Confirm

Parallelization

- Important for large area analysis
- (almost) Embarrassingly parallel
- Resolve overlaps by bisection before stitching
- Some CF needs post-processing



60 sqdeg/35 cells



cell and overlap size configurable

12

Cluster Finder (WAZP)

View Processes | Results | Comments

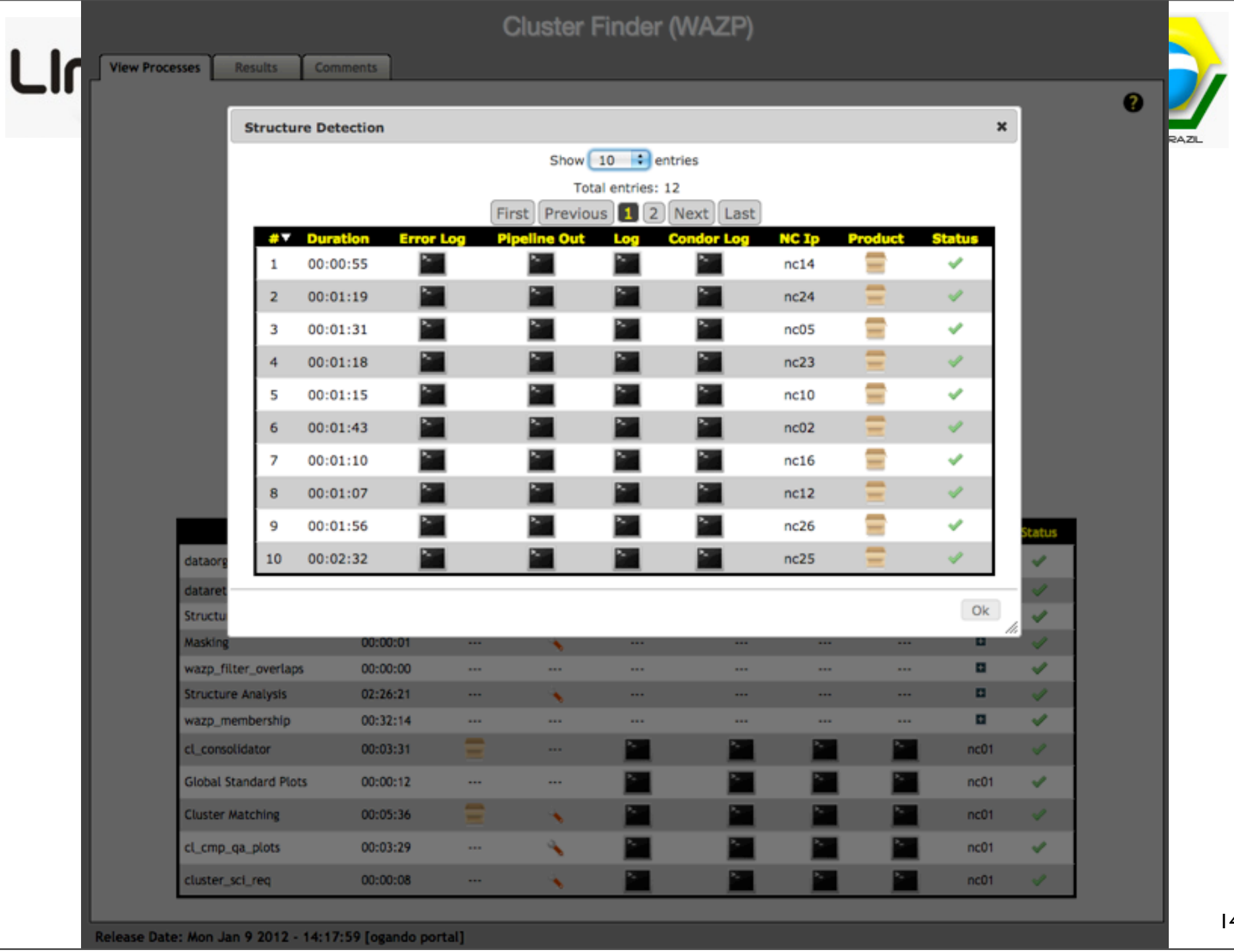
Pipeline: Cluster Finder	
User	Ricardo Ogando
Process	10007936
Start	2012-02-16 02:17:10
End	2012-02-16 08:50:15
Duration	06:33:05
Ingest Time	00:00:02
Expiration	2012-02-21 08:50:15
Jobs	78
Automatically Saved	Yes
Algorithm	WAZP
Configuration	View
Code Version	View

Input Data	
Input data	Addgals v3.04
Type	Catalog
RA (deg)	20
DEC (deg)	42.5
Selected Region (deg)	9
Objects	4803125
Area (sq deg)	54.67
Density (objects/sq arc min)	24.4

Module	Duration	Products	Config	Error Log	Pipeline Out	Log	Condor Log	NC	Status
dataorganizer	00:02:05	---						nc01	
dataretriever	00:01:35	---	---	---	---	---	---		
Structure Detection	00:01:24	---		---	---	---	---		
Masking	00:00:01	---		---	---	---	---		
wazp_filter_overlaps	00:00:00	---	---	---	---	---	---		
Structure Analysis	02:26:21	---		---	---	---	---		
wazp_membership	00:32:14	---	---	---	---	---	---		
cl_consolidator	00:03:31		---					nc01	
Global Standard Plots	00:00:12	---	---					nc01	
Cluster Matching	00:05:36							nc01	
cl_cmp_qa_plots	00:03:29	---						nc01	
cluster_sci_req	00:00:08	---						nc01	

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

13


Cluster Finder (WAZP)
View Processes
Results
Comments
Structure Detection
Show 10 entries
Total entries: 12
First Previous 1 2 Next Last
Duration Error Log Pipeline Out Log Condor Log NC Ip Product Status
1 00:00:55 nc14
2 00:01:19 nc24
3 00:01:31 nc05
4 00:01:18 nc23
5 00:01:15 nc10
6 00:01:43 nc02
7 00:01:10 nc16
8 00:01:07 nc12
9 00:01:56 nc26
10 00:02:32 nc25
Ok
Masking 00:00:01
wazp_filter_overlaps 00:00:00
Structure Analysis 02:26:21
wazp_membership 00:32:14
cl_consolidator 00:03:31 nc01
Global Standard Plots 00:00:12 nc01
Cluster Matching 00:05:36 nc01
cl_cmp_qa_plots 00:03:29 nc01
cluster_sci_req 00:00:08 nc01
Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

14

Cluster Finder (WAZP)

View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Quick Navigation

Input data

Data	Addgals v3.04
Type	Catalog
RA center (degree)	20
Dec center (degree)	42.5
Boxsize	9
Number of objects	4803125
Number of cells	12
Area of individual cell (square degree)	6.00
Overlap between cells (degree)	0.20
RA min (degree)	15.500
RA max (degree)	24.500
Dec min (degree)	38.000
Dec max (degree)	47.000
Total area (square degree)	54.67
Redshift type	Gaussian error photometric redshift
Magnitude type	ADDGALS OBSERVED magnitude

Data Organizer

Total area of the field (sq.deg.)	59.63
Mean area of tiles including overlap	6.00
Mean area of tiles without overlap	4.20
Ideal number of tiles	14.20
Effective Nr. of tiles	12
Min/Max areas of tiles accepted range	0.25 / 30.00
Min/Max areas of tiles including overlaps	5.62 / 7.02
Min/Max areas of tiles without overlaps	4.70 / 5.24

Algorithm

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

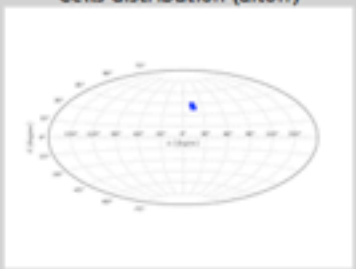
Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Quick Navigation

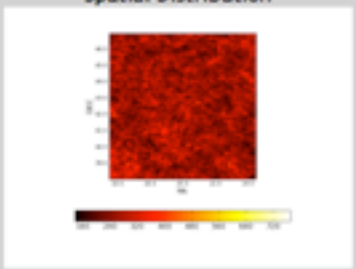
- Maps
- Footprint
- Irregular tiling

Maps ?

Cells distribution (aitoff)

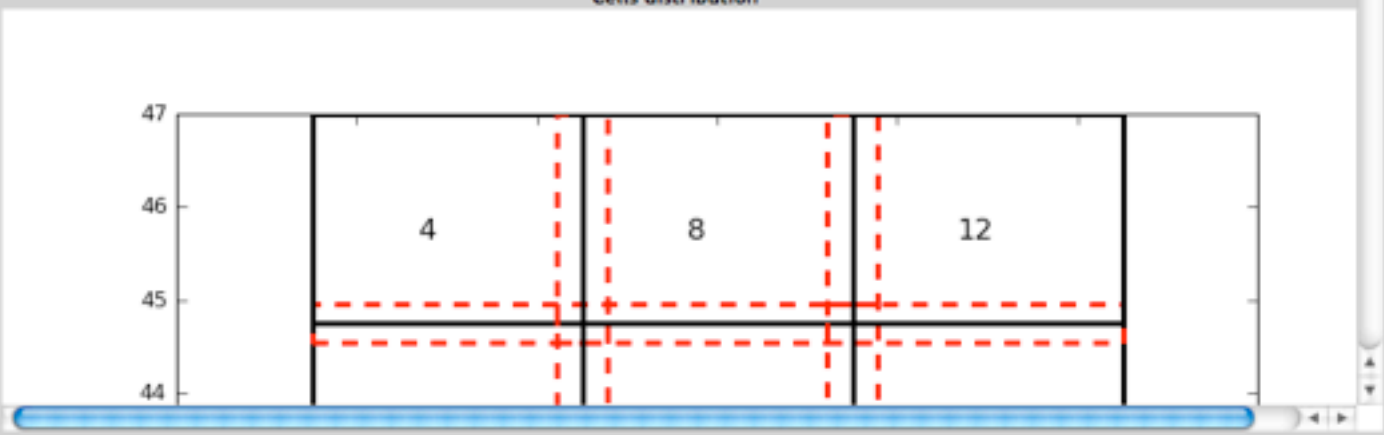


Spatial Distribution



Footprint ?

Cells distribution



Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

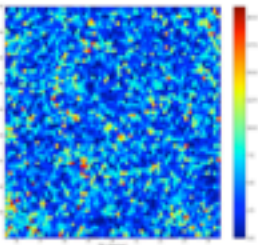
Cluster Finder (WAZP)

View Processes | **Results** | Comments

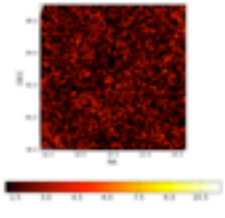
Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Consolidated catalog ?

Consolidated Clusters heatmap



Consolidated Clusters heatmap



Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Clusters/Halos properties | Clusters properties in redshift shells

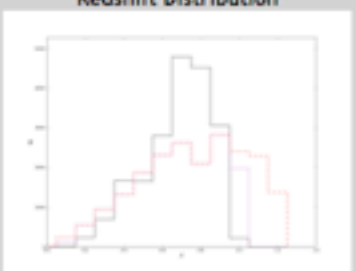
Quick Navigation

- Clusters properties
- Halos properties

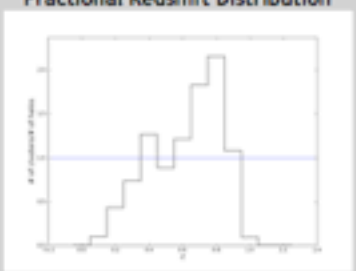
Clusters/Halos properties ?

Clusters properties ?

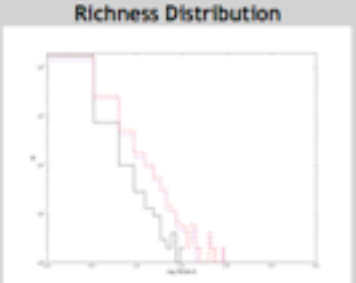
Redshift Distribution



Fractional Redshift Distribution



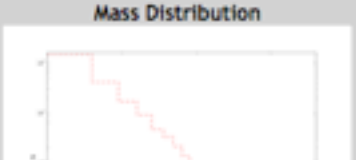
Richness Distribution



Richness vs Redshift

Halos properties ?

Mass Distribution



M200 vs NGALS

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

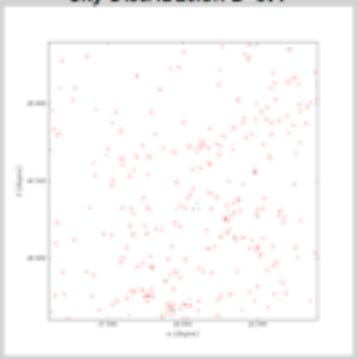
View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

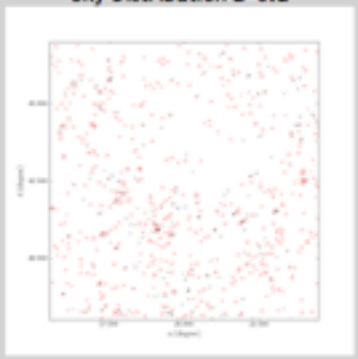
Clusters/Halos properties | Clusters properties in redshift shells

Spatial distribution ?

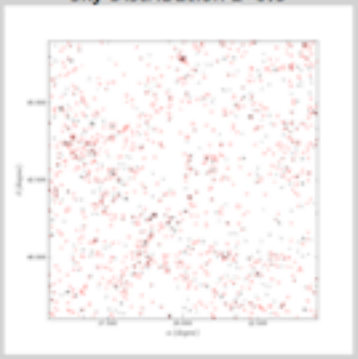
Sky Distribution z=0.1



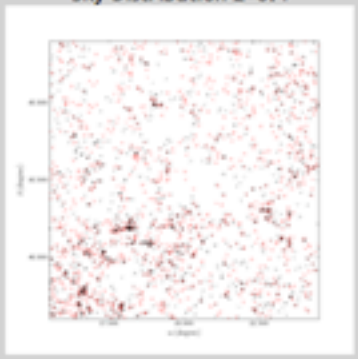
Sky Distribution z=0.2



Sky Distribution z=0.3



Sky Distribution z=0.4



Sky Distribution z=0.5

Sky Distribution z=0.6

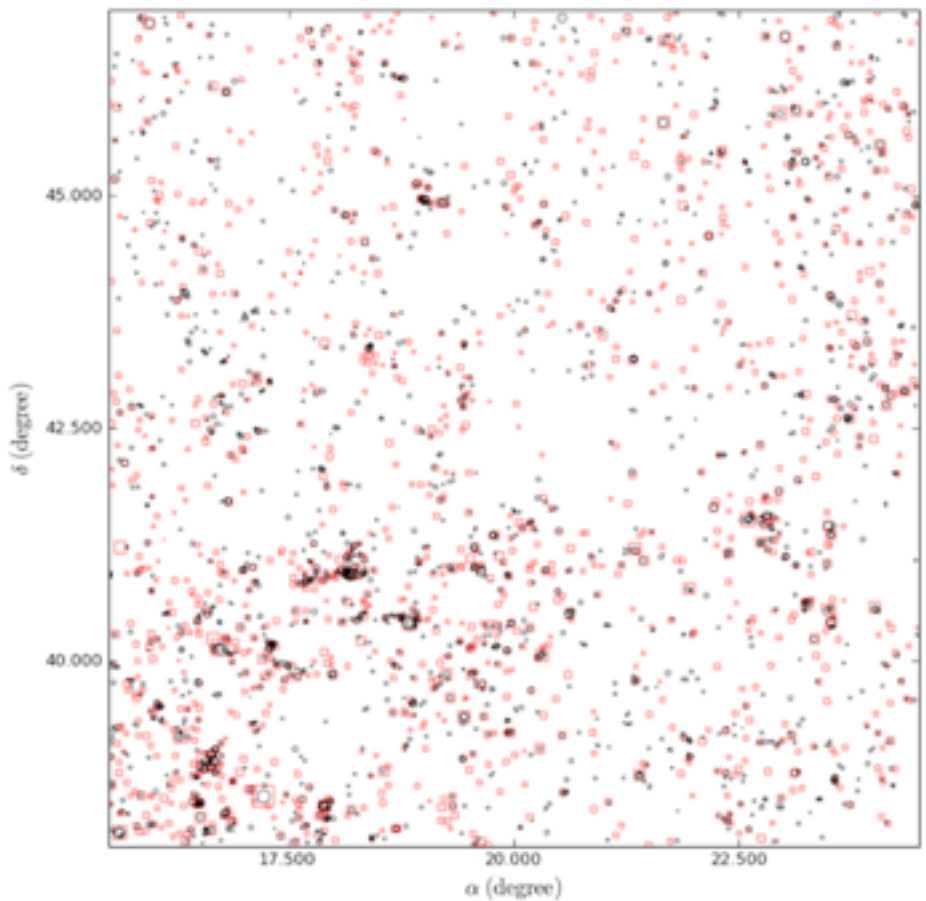
Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Clusters/Halos properties | Cluster



Spatial distribution of clusters (black circles) and halos (red squares, M200>1e+13 (M☉)) in the redshift bin 0.4, scaled by their respective richness, from 0 to 50 and from 0 to 217, respectively.

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

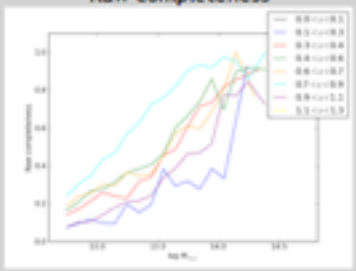
Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Quick Navigation

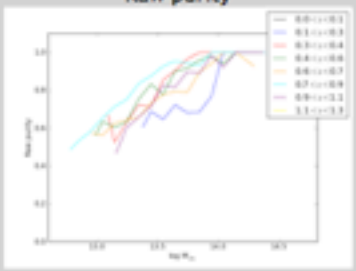
- Completeness/Purity
- Centering statistics
- Redshift statistics
- Mobs-Mtrue Unique
- Mobs-Mtrue Non-unique
- Cluster Comparison panels
- Mass-rank fit

Completeness/Purity ?

Raw completeness

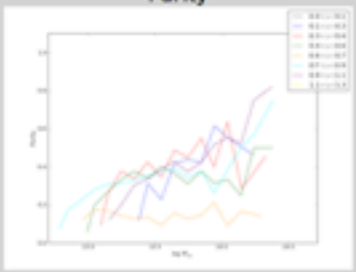


Raw purity

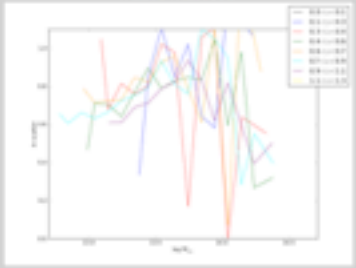


Completeness

Purity



Mass scatter



Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

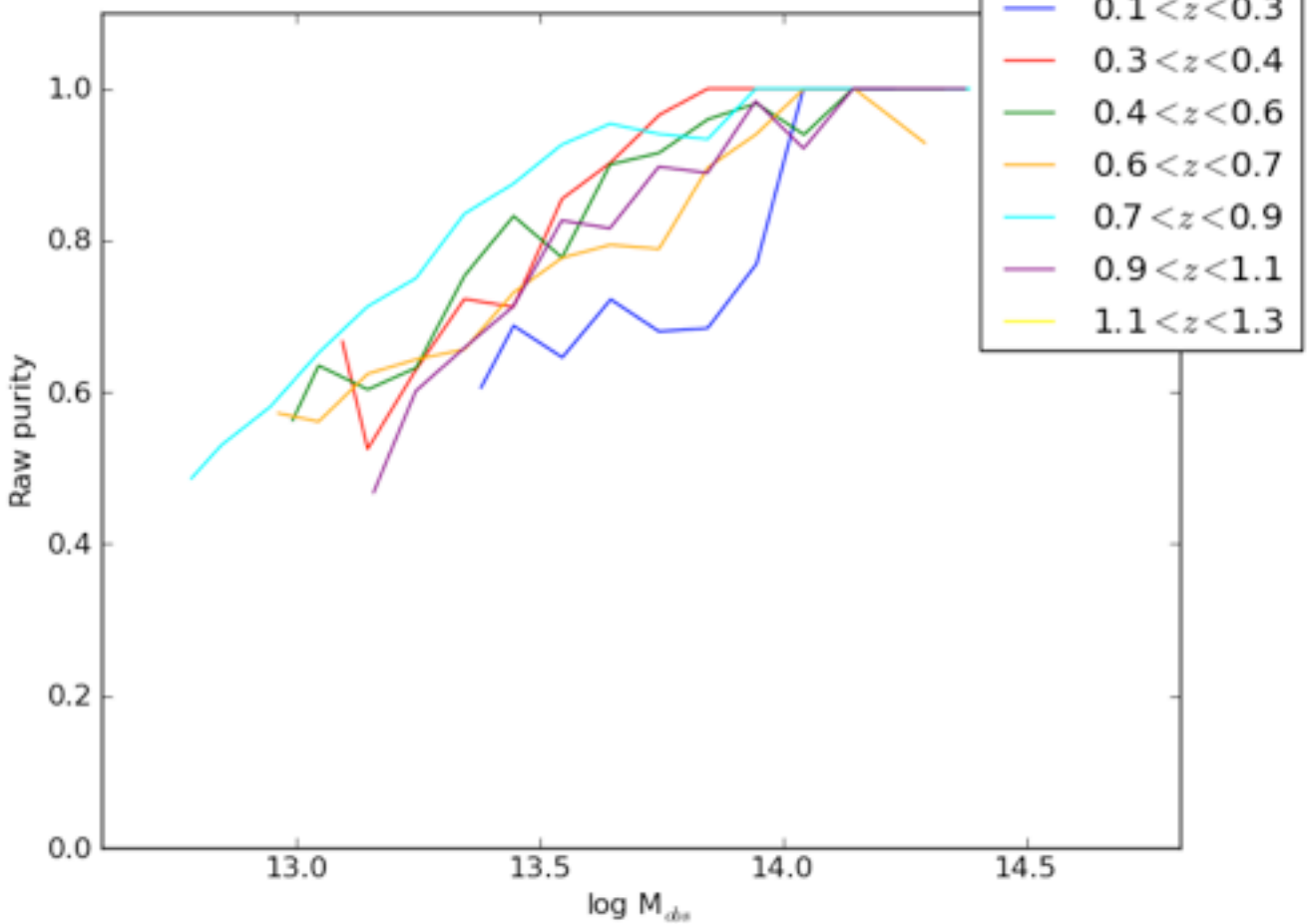
Cluster Finder (WAZP)

View Processes | Results | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Quick Navig

Completeness/Purity
Centering statistics
Redshift statistics
Mobs-Mtrue Unique
Mobs-Mtrue Non-unic
Cluster Comparison
Mass-rank fit



Raw purity

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

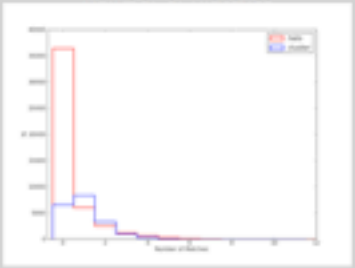
Cluster Finder (WAZP)

View Processes | **Results** | Comments

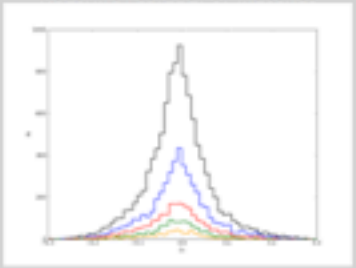
Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Matching

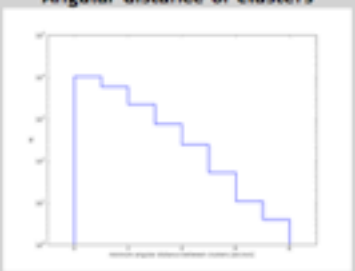
Number of matches



Redshift residual of matches



Angular distance of clusters



Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

Unmatched Clusters | Oddly Matched Clusters

Quick Navigation

- Cluster Rank: 8
  - Cluster
  - Nearby Halos
- Cluster Rank: 18
  - Cluster
  - Nearby Halos
- Cluster Rank: 23
  - Cluster
  - Nearby Halos
- Cluster Rank: 28
  - Cluster
  - Nearby Halos
- Cluster Rank: 35
  - Cluster
  - Nearby Halos

Unmatched Clusters ?

Cluster Rank: 8 ?

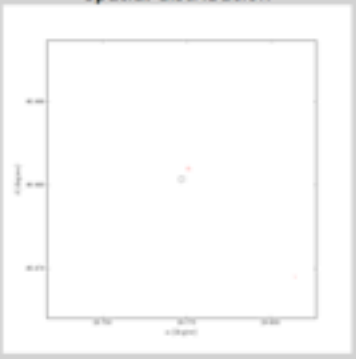
Cluster

Rank	RA	Dec	z	Ngals
8	19.773	40.480	0.932	64

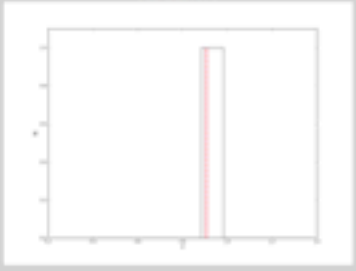
Nearby Halos

Rank	RA	Dec	z	Ngals	Mass
215	19.776	40.481	0.905	27	2.0E+14
56811	19.807	40.473	0.906	2	1.2E+13

Spatial distribution



Redshift



Cluster Rank: 18 ?

Cluster

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | Science Requirements | Matching metrics

The Sarah Hansen Chronicles

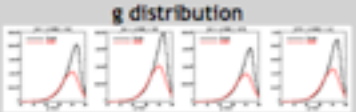
Quick Navigation

- Number counts (!mere bogus display!)
- Colors
- Galaxy recovery
- Red sequence
- Richness

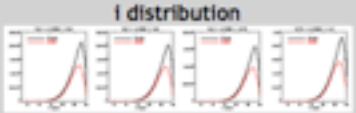
The Sarah Hansen Chronicles ?

Number counts (!mere bogus display!) ?

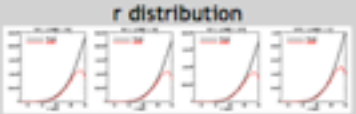
g distribution



i distribution



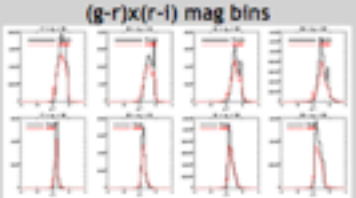
r distribution



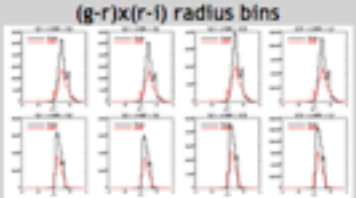
z distribution

Colors ?

(g-r)x(r-i) mag bins

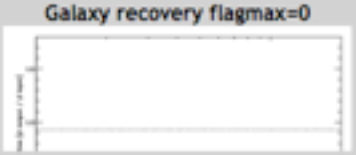


(g-r)x(r-i) radius bins

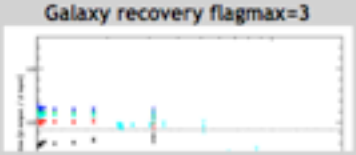


Galaxy recovery ?

Galaxy recovery flagmax=0



Galaxy recovery flagmax=3



Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Finder (WAZP)

View Processes | **Results** | Comments

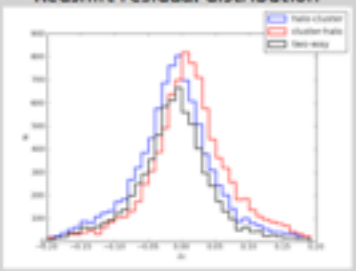
Summary | Data Organizer | Consolidator | Generic Plots | Cluster Comparison | QA | View unmatched | CWG tests | **Science Requirements** | Matching metrics

Quick Navigation

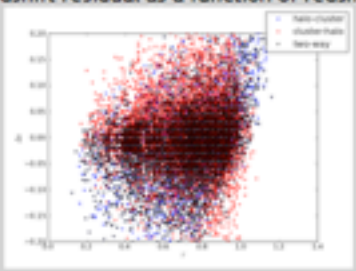
- R24: redshift statistics
- R32: mass richness scatter

R24: redshift statistics

Redshift residual distribution



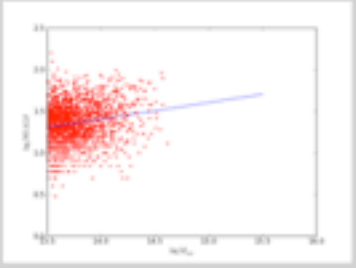
Redshift residual as a function of redshift



Science Requirement 24 status		✔
Uncertainty in the bias of cluster photo-z σ(zbias) < 0.03	0.0007	✔
Uncertainty in dispersion of cluster photo-z σ(σz) < 0.0016	0.00006084	✔

R32: mass richness scatter

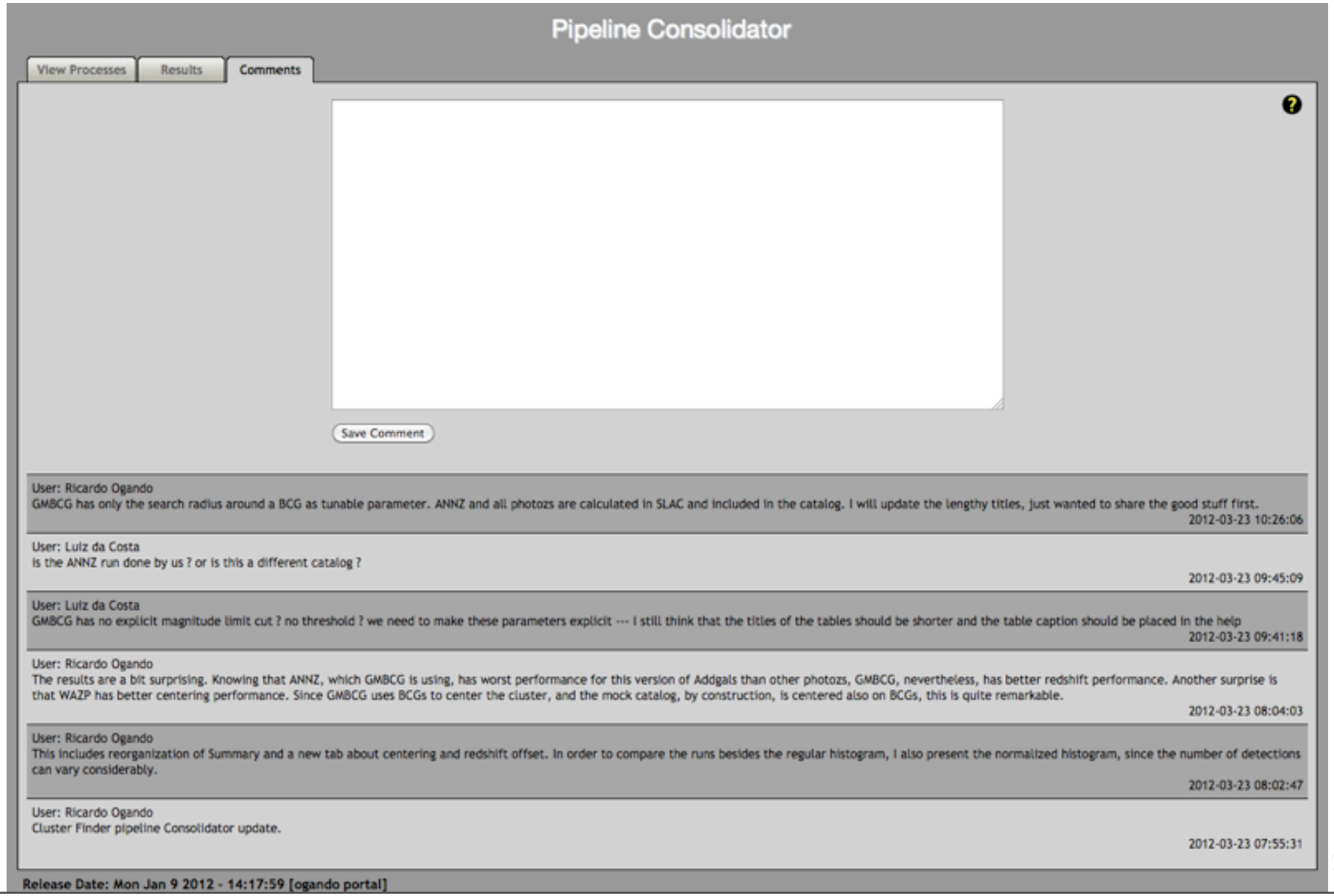
Mass Richness



Science Requirement 32: Mass Richness scatter	0.237

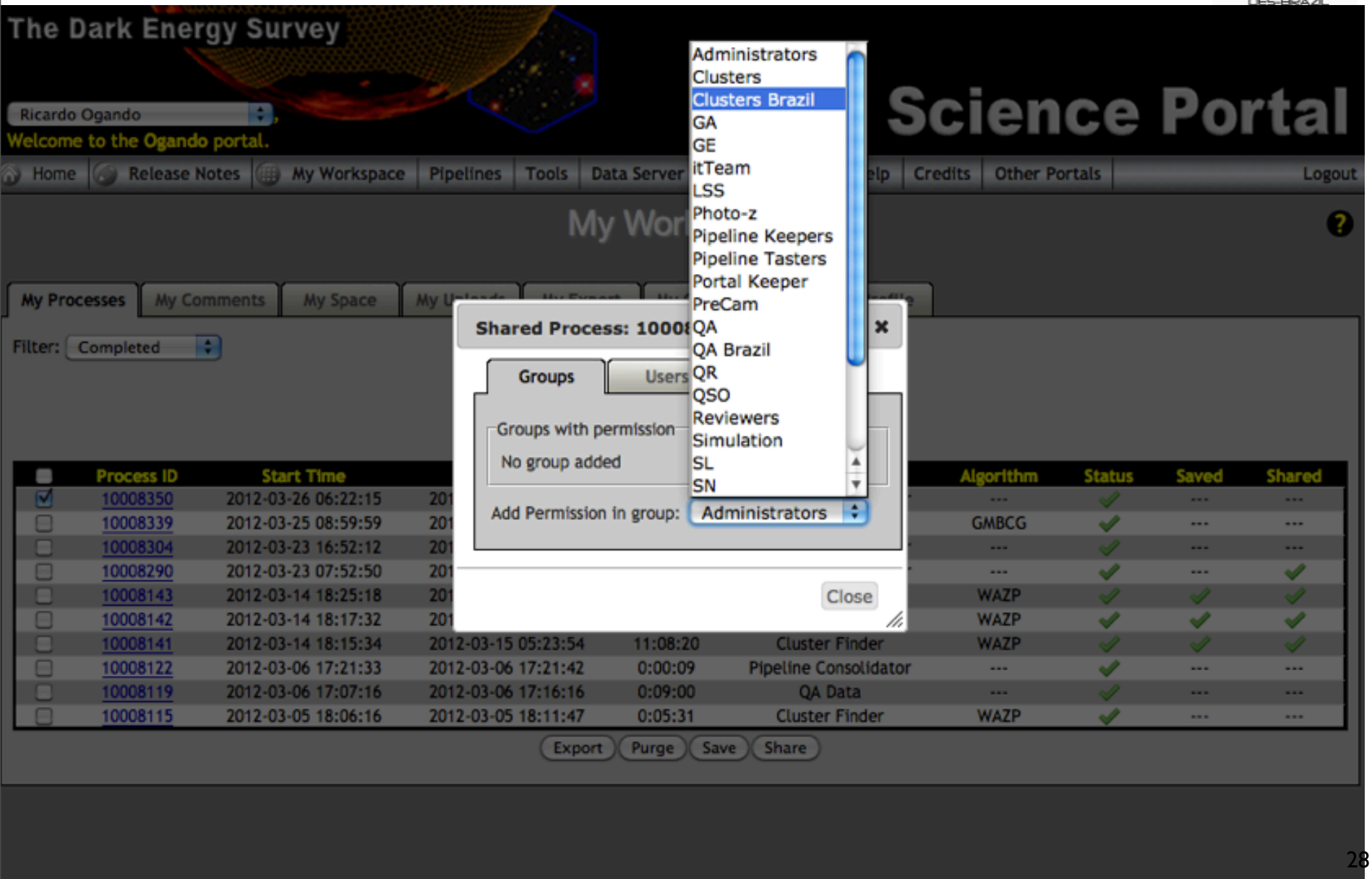
Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Comments


Pipeline Consolidator
View Processes
Results
Comments
Save Comment
User: Ricardo Ogando
GMBCG has only the search radius around a BCG as tunable parameter. ANNZ and all photozs are calculated in SLAC and included in the catalog. I will update the lengthy titles, just wanted to share the good stuff first.
2012-03-23 10:26:06
User: Luiz da Costa
is the ANNZ run done by us ? or is this a different catalog ?
2012-03-23 09:45:09
User: Luiz da Costa
GMBCG has no explicit magnitude limit cut ? no threshold ? we need to make these parameters explicit --- I still think that the titles of the tables should be shorter and the table caption should be placed in the help
2012-03-23 09:41:18
User: Ricardo Ogando
The results are a bit surprising. Knowing that ANNZ, which GMBCG is using, has worst performance for this version of Addgals than other photozs, GMBCG, nevertheless, has better redshift performance. Another surprise is that WAZP has better centering performance. Since GMBCG uses BCGs to center the cluster, and the mock catalog, by construction, is centered also on BCGs, this is quite remarkable.
2012-03-23 08:04:03
User: Ricardo Ogando
This includes reorganization of Summary and a new tab about centering and redshift offset. In order to compare the runs besides the regular histogram, I also present the normalized histogram, since the number of detections can vary considerably.
2012-03-23 08:02:47
User: Ricardo Ogando
Cluster Finder pipeline Consolidator update.
2012-03-23 07:55:31
Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

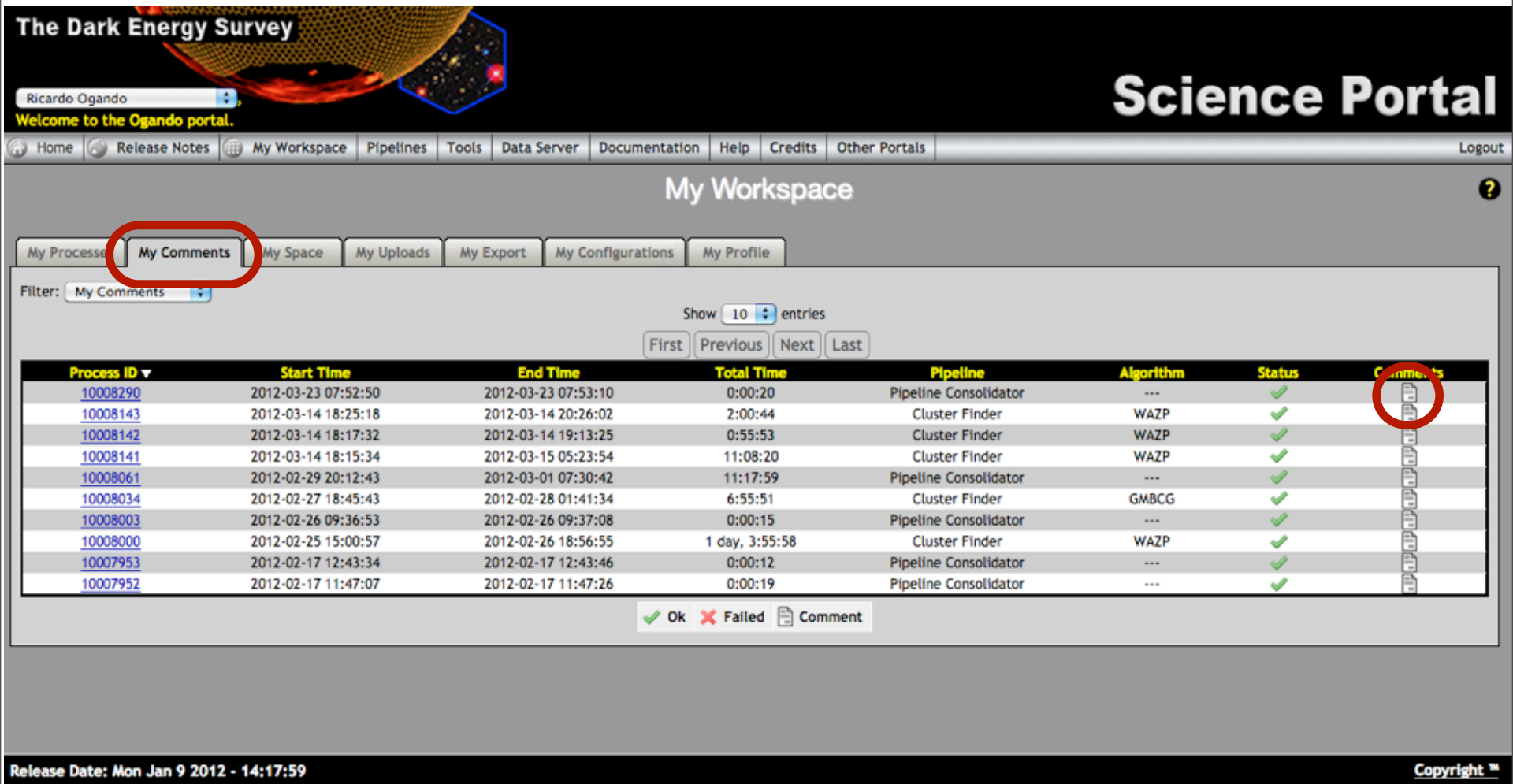
27

Jobs can be **shared** with groups or individuals, which can comment the results



28

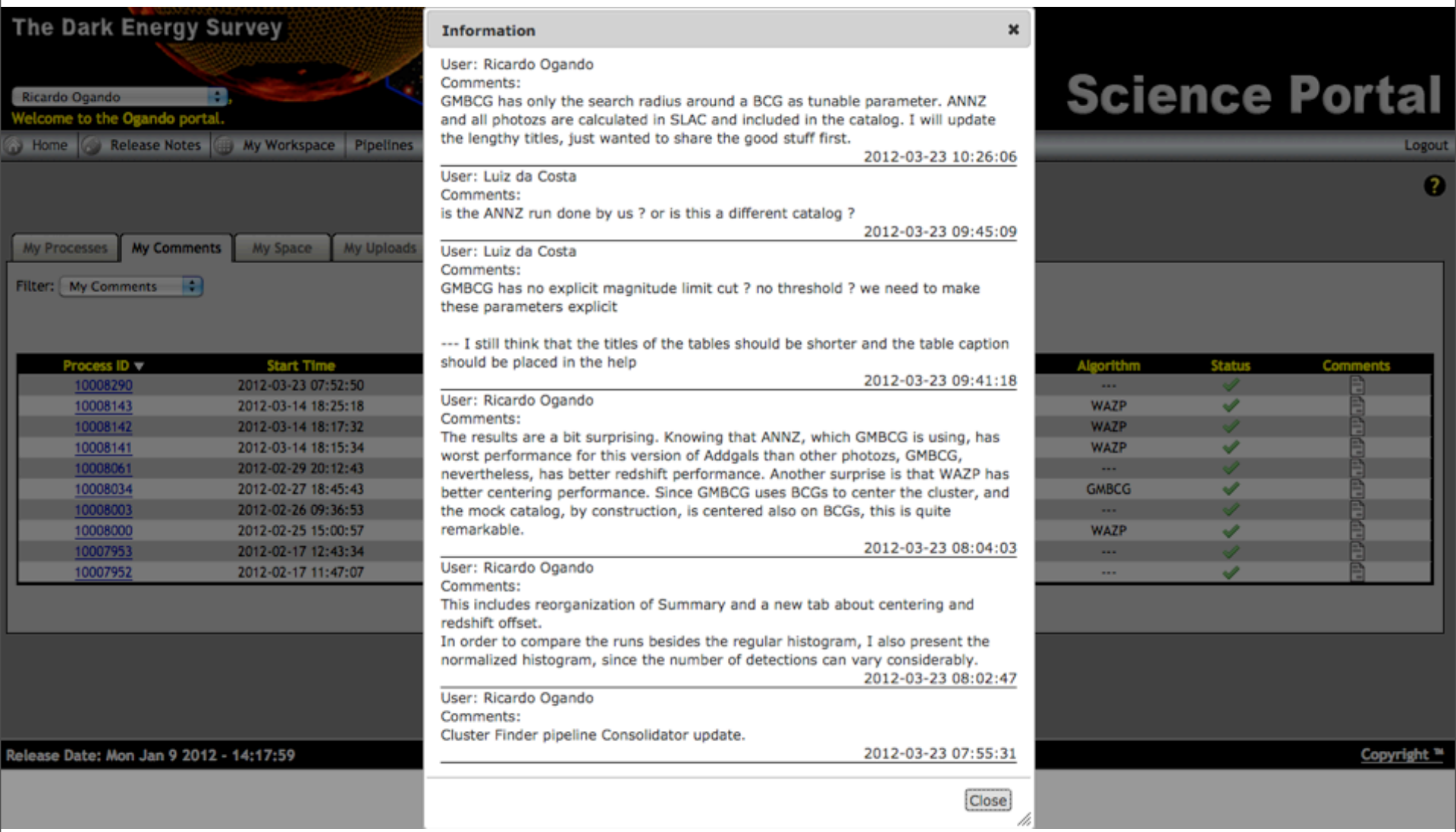
Comments



Browse comments through **My Workspace**

29

Comments


The Dark Energy Survey
Ricardo Ogando
Welcome to the Ogando portal.
Home
Release Notes
My Workspace
Pipelines
Science Portal
Logout
My Processes
My Comments
My Space
My Uploads
Filter: My Comments
Process ID
Start Time
Algorithm
Status
Comments
10008290 2012-03-23 07:52:50 ---
10008143 2012-03-14 18:25:18 WAZP
10008142 2012-03-14 18:17:32 WAZP
10008141 2012-03-14 18:15:34 WAZP
10008061 2012-02-29 20:12:43 ---
10008034 2012-02-27 18:45:43 GMBCG
10008003 2012-02-26 09:36:53 ---
10008000 2012-02-25 15:00:57 WAZP
10007953 2012-02-17 12:43:34 ---
10007952 2012-02-17 11:47:07 ---
Release Date: Mon Jan 9 2012 - 14:17:59
Copyright
Information
User: Ricardo Ogando
Comments:
GMBCG has only the search radius around a BCG as tunable parameter. ANNZ and all photozs are calculated in SLAC and included in the catalog. I will update the lengthy titles, just wanted to share the good stuff first.
2012-03-23 10:26:06
User: Luiz da Costa
Comments:
is the ANNZ run done by us ? or is this a different catalog ?
2012-03-23 09:45:09
User: Luiz da Costa
Comments:
GMBCG has no explicit magnitude limit cut ? no threshold ? we need to make these parameters explicit
--- I still think that the titles of the tables should be shorter and the table caption should be placed in the help
2012-03-23 09:41:18
User: Ricardo Ogando
Comments:
The results are a bit surprising. Knowing that ANNZ, which GMBCG is using, has worst performance for this version of Addgals than other photozs, GMBCG, nevertheless, has better redshift performance. Another surprise is that WAZP has better centering performance. Since GMBCG uses BCGs to center the cluster, and the mock catalog, by construction, is centered also on BCGs, this is quite remarkable.
2012-03-23 08:04:03
User: Ricardo Ogando
Comments:
This includes reorganization of Summary and a new tab about centering and redshift offset.
In order to compare the runs besides the regular histogram, I also present the normalized histogram, since the number of detections can vary considerably.
2012-03-23 08:02:47
User: Ricardo Ogando
Comments:
Cluster Finder pipeline Consolidator update.
2012-03-23 07:55:31
Close

30

Pipeline Consolidator

- Access saved processes
- Compare side by side
  - Published metrics
  - Common output from halo-cluster matching code
- Cases:
  - Fine Tuning: WAZP
  - Cluster Challenge: WAZP VS GMBCG

31

WAZP Fine Tuning

- WAZP (Christophe Benoist/OCA)
  - Identification of galaxy over-densities in Ra, Dec, zphot space.
  - The underlying algorithm uses 2D and 1D density field reconstruction based on wavelet transforms.
- Tunable parameters:
  - Threshold
  - Magnitude limit
  - Parallelization (size of cells in sky partition)

32

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

Quick Navigation

- Data and Algorithm
- Data and Algorithm Properties
- Cluster Properties
- Halo Properties
- Matching
- Mass scatter
- Limits of Raw Properties
- Limits of Cosmological Properties

Data and Algorithm

Process ID	Data	Algorithm	Code version	Redshift type
10007934	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007935	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007936	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007937	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift

Data and Algorithm Properties

Process ID	Magnitude limit	Total area (square degree)	Cell area (sqdeg)	Number of cells	Threshold
10007934	24.0	54.67	3.00	35	5.00
10007935	23.0	54.67	3.00	35	5.00
10007936	24.0	54.67	6.00	12	5.00
10007937	24.0	54.67	3.00	35	3.00

Cluster Properties

Process ID	Mean density	Mean redshift	Mean richness
10007934	320.2	0.73	20
10007935	164.8	0.59	17
10007936	360.2	0.73	20
10007937	681.2	0.72	17

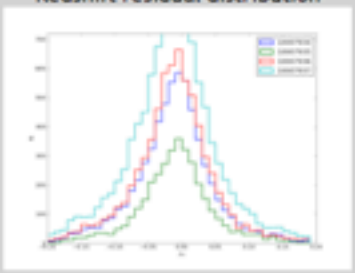
elease Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

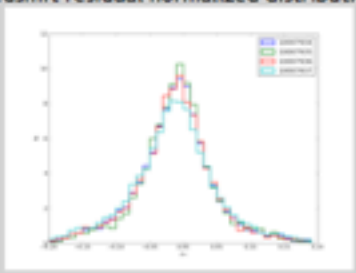
View Processes | **Results** | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

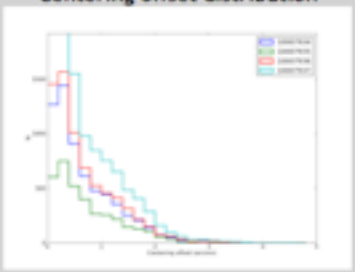
Redshift residual distribution



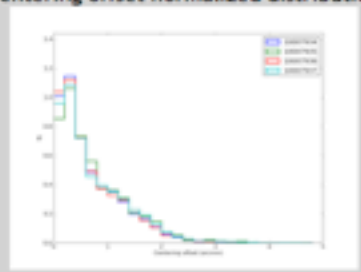
Redshift residual normalized distribution



Centering offset distribution



Centering offset normalized distribution

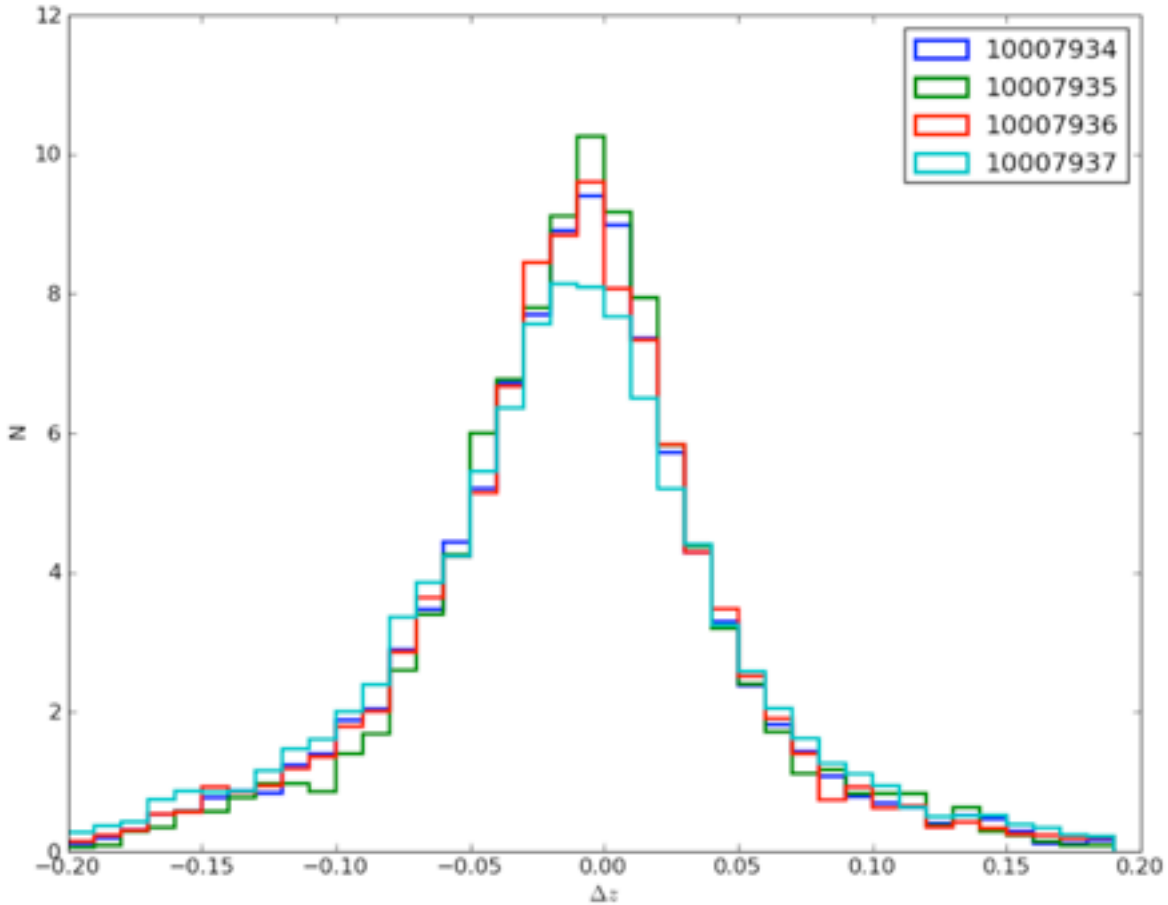


lease Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity



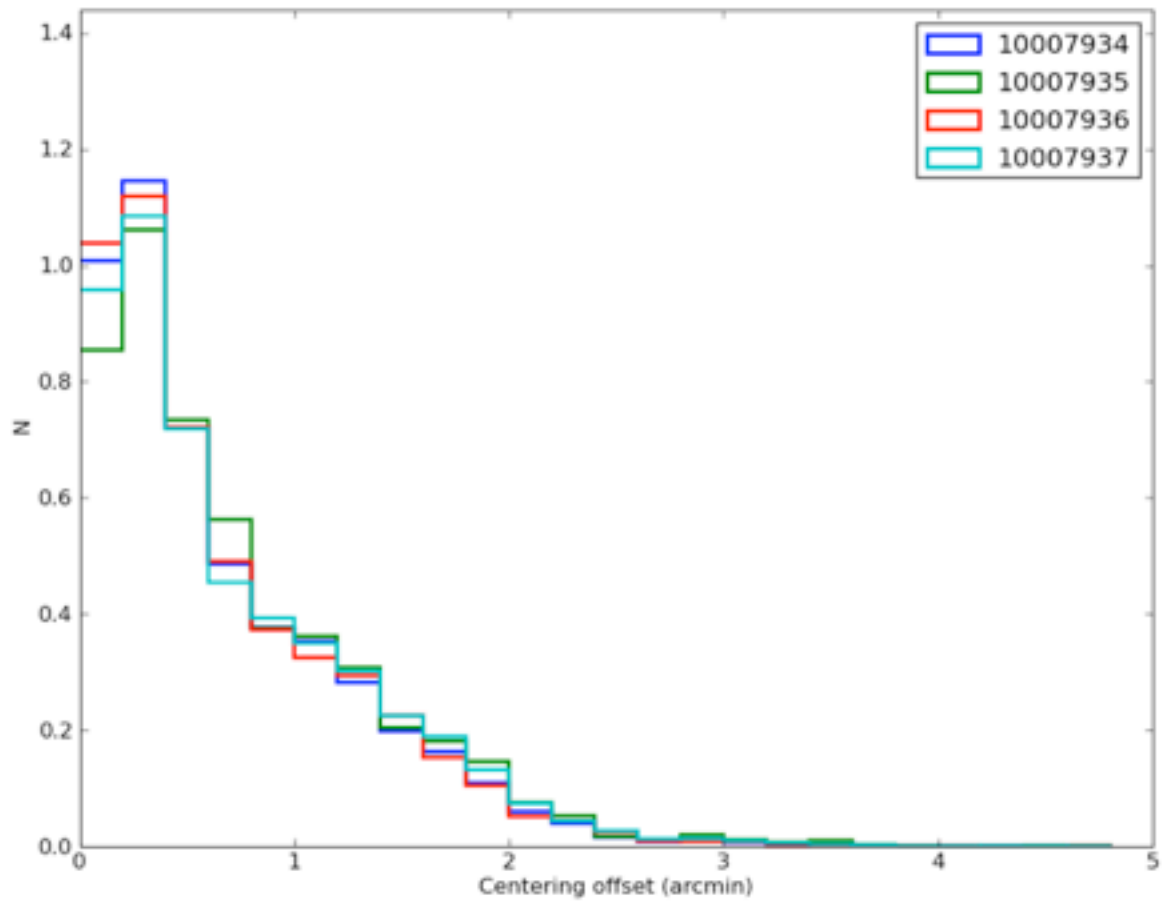
Normalized distribution of redshift residual of each process.

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity



Centering offset normalized distribution of each process.

lease Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

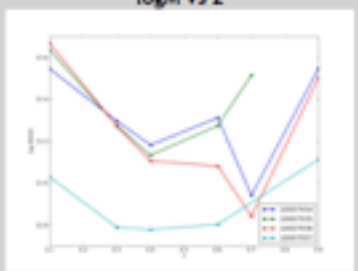
Raw | Cosmological

Quick Navigation

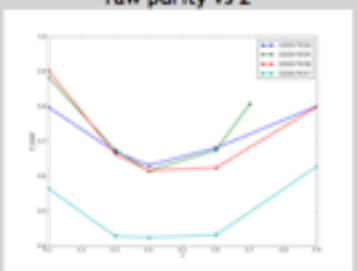
- Raw completeness 30%
- Raw completeness 60%
- Raw completeness 90%

Raw completeness 30%

logM vs z

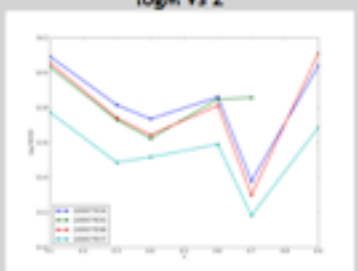


raw purity vs z

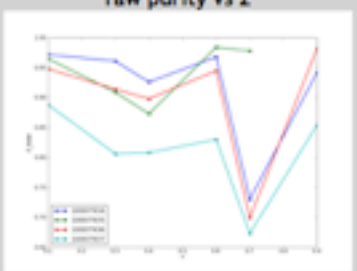


Raw completeness 60%

logM vs z



raw purity vs z



Raw completeness 90%

logM vs z

raw purity vs z

Raw

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

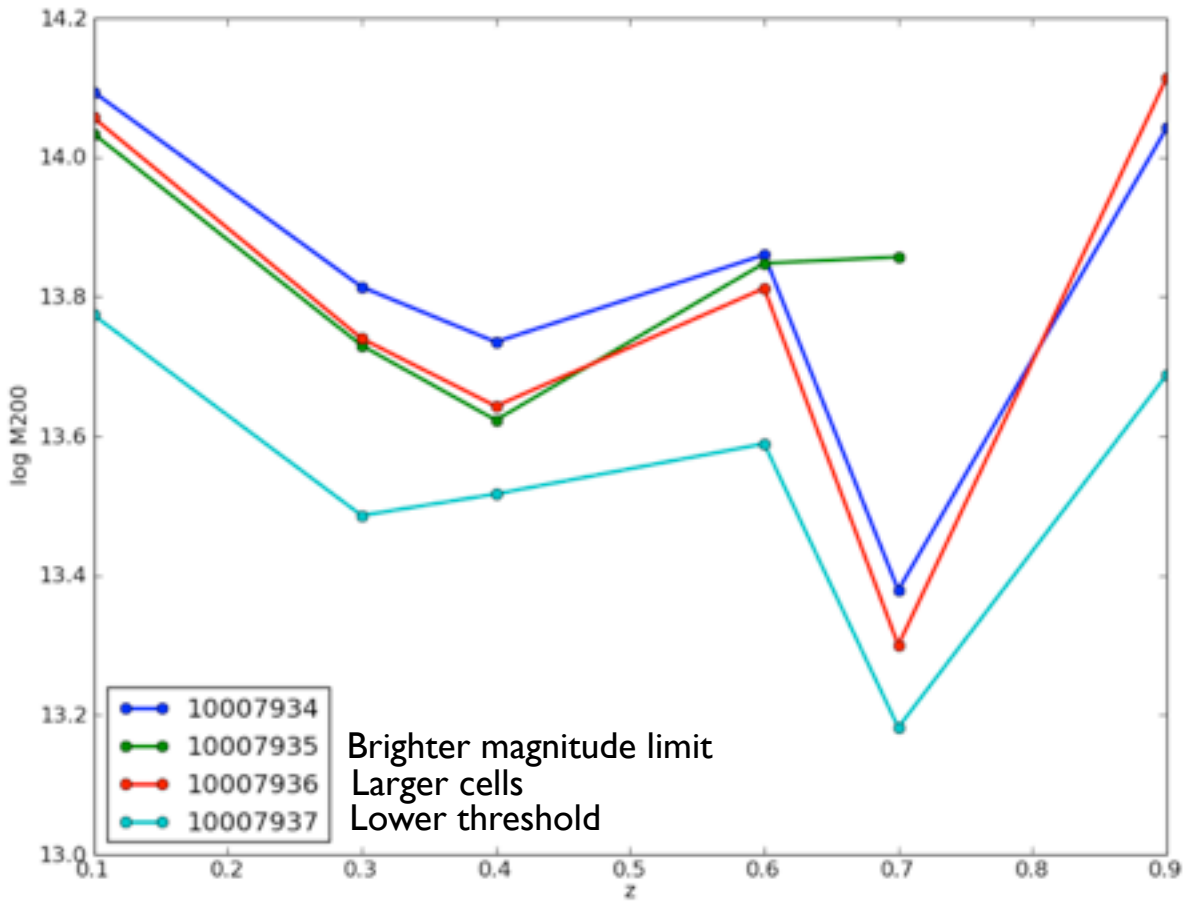
Raw | Cosmo

Quick Naviga

Raw completeness 30

Raw completeness 60

Raw completeness 90



Plot of log(Mass) vs redshift for different processes at raw completeness=0.6

Raw

lease Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Cluster Challenge

- WAZP VS GMBCG
- GMBCG searches for over-densities of red sequence galaxies + BCG. It uses photometric redshifts to assign the cluster's redshift.

39

Pipeline Consolidator

View Processes | **Results** | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

Quick Navigation

- Data and Algorithm
- Data and Algorithm Properties
- Cluster Properties
- Halo Properties
- Matching
- Mass scatter
- Limits of Raw Properties
- Limits of Cosmological Properties

Data and Algorithm

Process ID	Data	Algorithm	Code version	Redshift type
10008000	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10008034	Addgals v3.04	GMBCG	gmbcgp-41-1:+	ANNZ

Data and Algorithm Properties

Process ID	Magnitude limit	Total area (square degree)	Area of individual cell (square degree)	Number of cells	Threshold
10008000	24.0	202.10	6.00	48	3.00
10008034	---	202.10	50.00	4	---

Cluster Properties

Process ID	Cluster Mean density	Cluster Mean redshift	Cluster Mean richness
10008000	657.5	0.72	16
10008034	468.7	0.66	8

Halo Properties

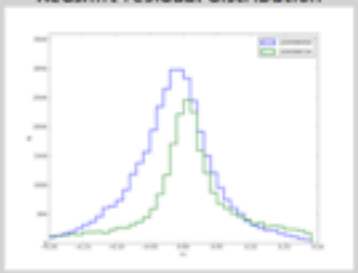
Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

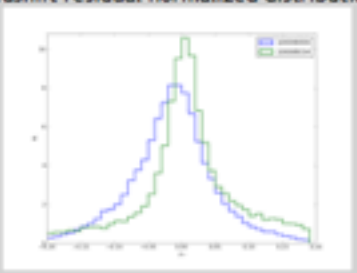
View Processes | **Results** | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

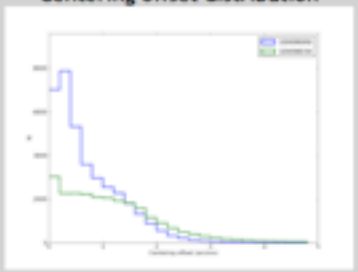
Redshift residual distribution



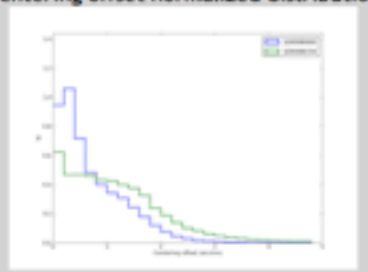
Redshift residual normalized distribution



Centering offset distribution



Centering offset normalized distribution

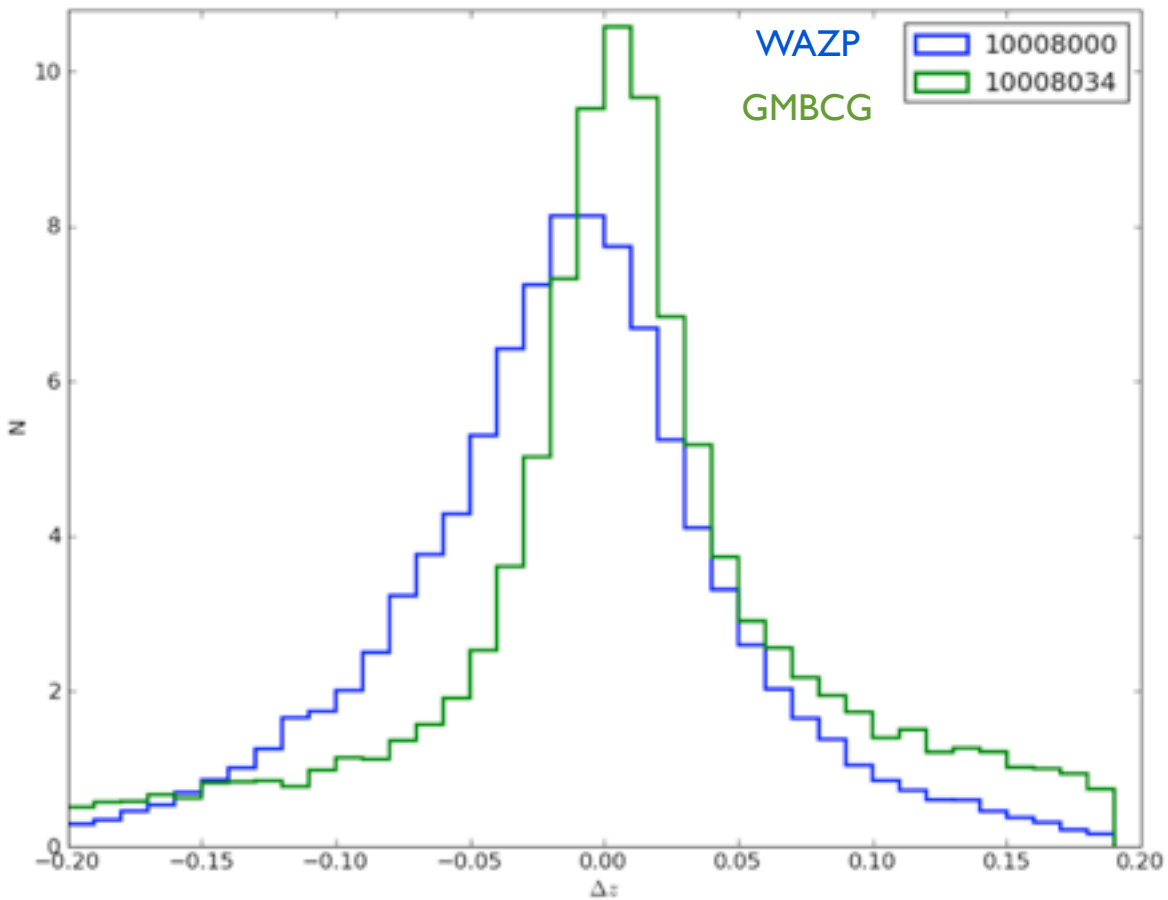


Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity



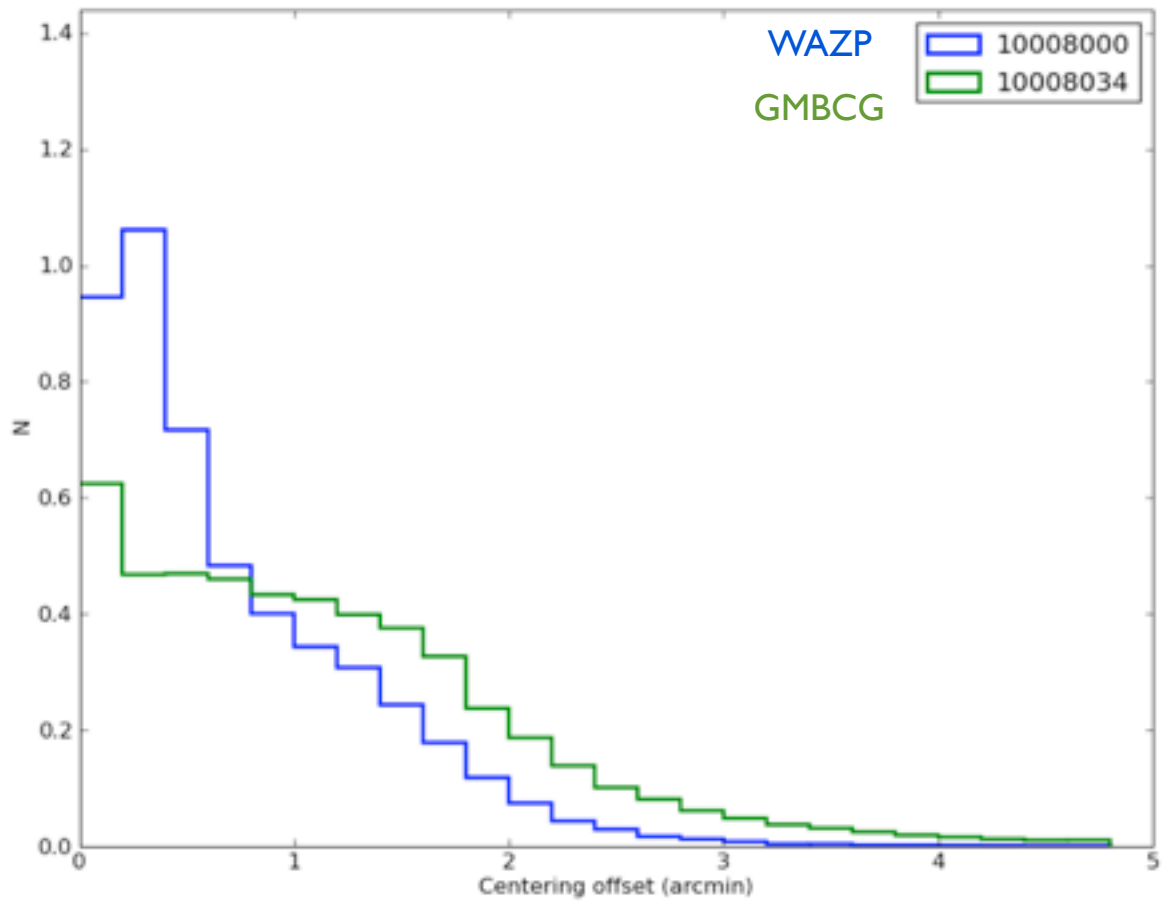
Normalized distribution of redshift residual of each process.

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity



Centering offset normalized distribution of each process.

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | **Results** | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

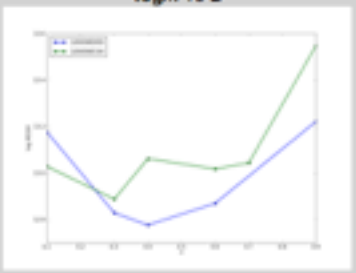
Raw | Cosmological

Quick Navigation

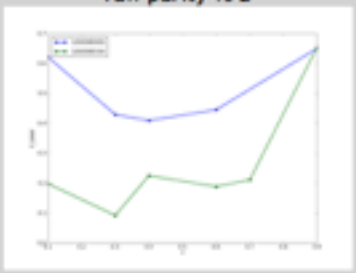
- Raw completeness 30%
- Raw completeness 60%
- Raw completeness 90%

Raw completeness 30%

logM vs z

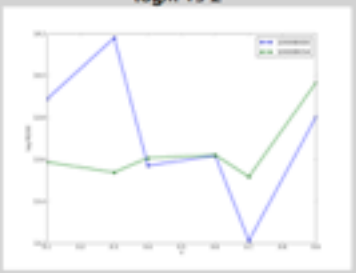


raw purity vs z



Raw completeness 60%

logM vs z



raw purity vs z

Raw completeness 90%

logM vs z

raw purity vs z

Raw

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

Pipeline Consolidator

View Processes | Results | Comments

Summary | Cluster Finder Consolidator Plots | Centering | Completeness | Purity

Raw | Cost

Quick Navi

Raw completeness

Raw completeness

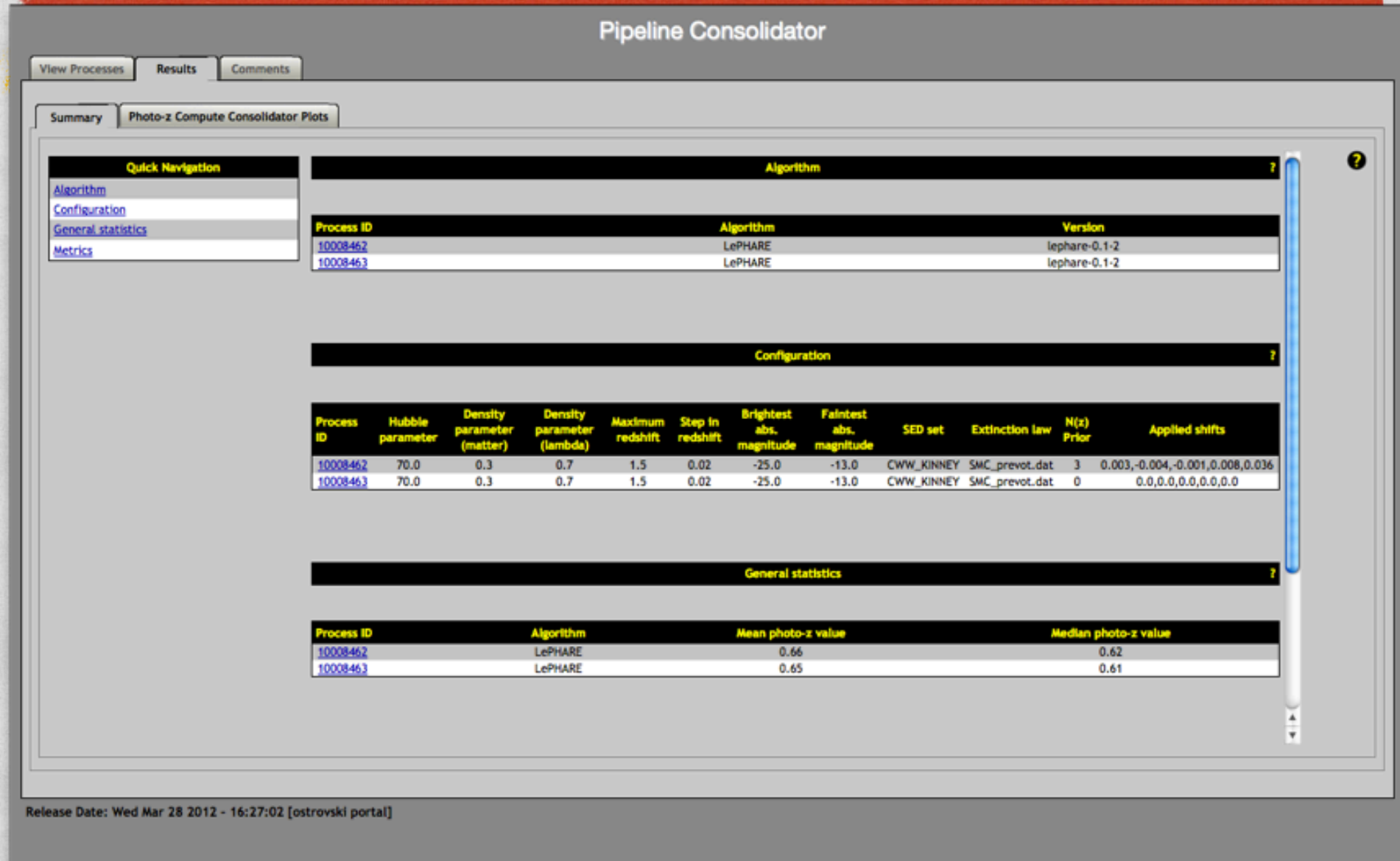
Raw completeness

Plot of log(Mass) vs redshift for different processes at raw completeness=0.6

Raw

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]

PHOTO-Z



Pipeline Consolidator

View Processes | Results | Comments

Summary | Photo-z Compute Consolidator Plots

Quick Navigation

- Algorithm
- Configuration
- General statistics
- Metrics

Algorithm

Process ID	Algorithm	Version
10008462	LePHARE	lephare-0.1-2
10008463	LePHARE	lephare-0.1-2

Configuration

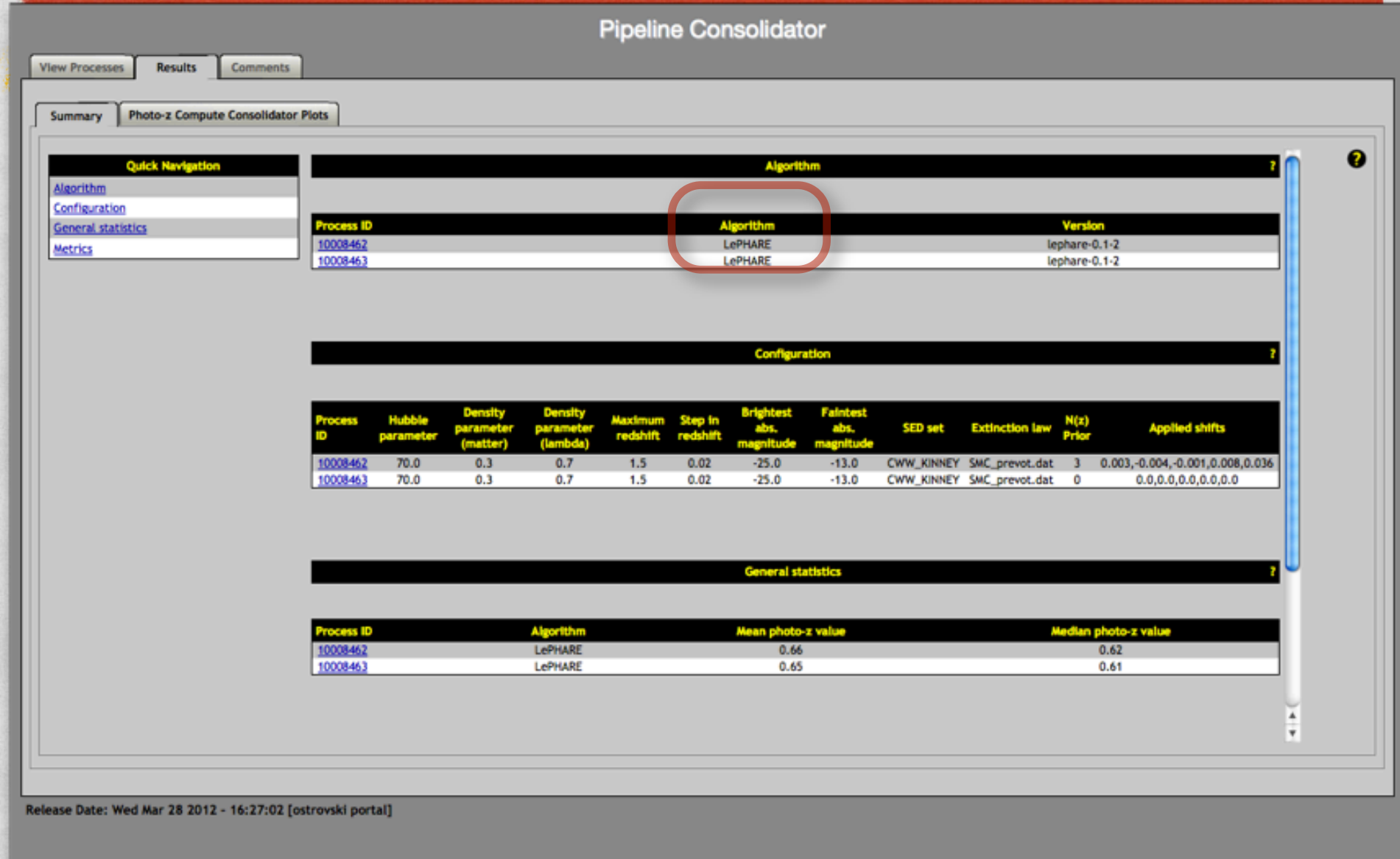
Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61

Release Date: Wed Mar 28 2012 - 16:27:02 [ostrovski portal]

PHOTO-Z



Pipeline Consolidator

View Processes | Results | Comments

Summary | Photo-z Compute Consolidator Plots

Quick Navigation

- Algorithm
- Configuration
- General statistics
- Metrics

Algorithm

Process ID	Algorithm	Version
10008462	LePHARE	lephare-0.1-2
10008463	LePHARE	lephare-0.1-2

Configuration

Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61

Release Date: Wed Mar 28 2012 - 16:27:02 [ostrovski portal]

PHOTO-Z

Pipeline Consolidator

View Processes | Results | Comments

Summary | Photo-z Compute Consolidator Plots

Quick Navigation

- Algorithm
- Configuration
- General statistics
- Metrics

Algorithm

Process ID	Algorithm	Version
10008462	LePHARE	lephare-0.1-2
10008463	LePHARE	lephare-0.1-2

Configuration

Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61

Release Date: Wed Mar 28 2012 - 16:27:02 [ostrovski portal]

Lessons learned

- CFs are in C, C++, Fortran, IDL, Perl, Python.
- Need special libraries.
- All wrapped by python modules
- Time to complete code wrapping depends on
  - have source code running
  - length of workflow,
  - length of configuration,
  - data needs (e.g. from FITS to ASCII),
  - language, libraries installation, and
  - **understanding** code.
- Some of those are first-timers' difficulties.

47

Conclusions

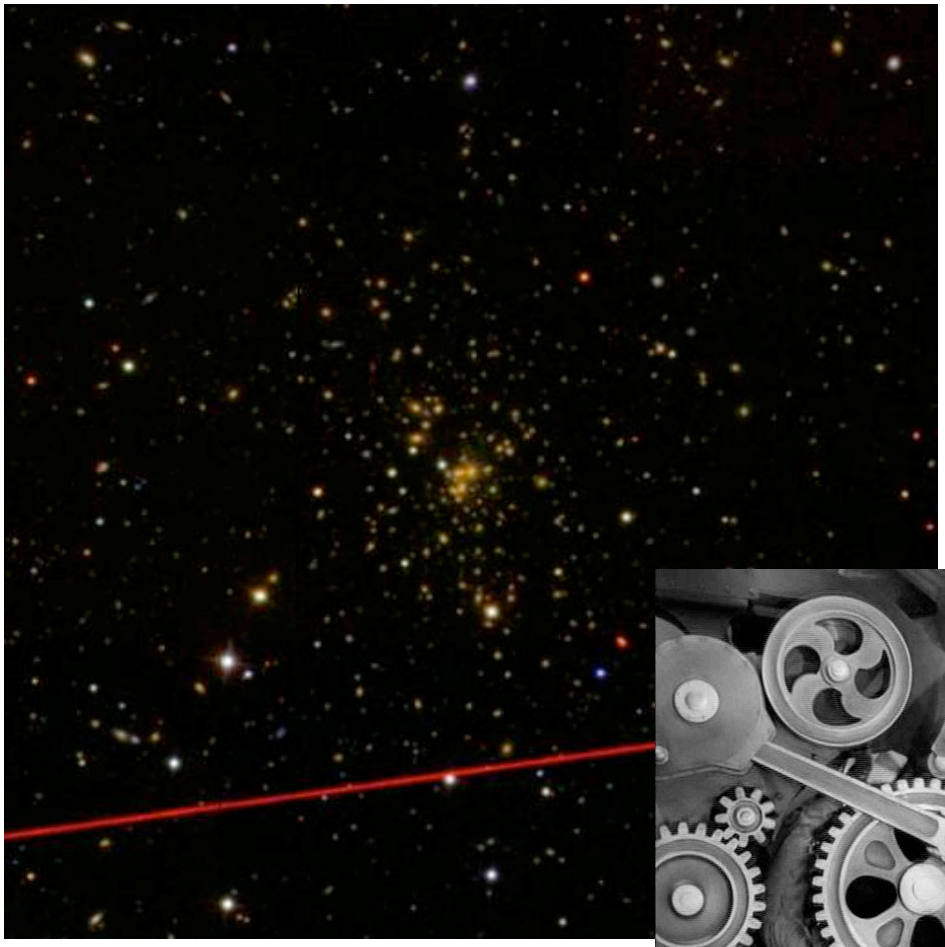
- Lots of people, codes, languages, cultures and schedules
  - Several Cluster Finders wrapped.
- Legacy. Production of cluster catalogs at the push of a button
- Pipeline Consolidator allowing easy comparison
- Future
  - close the loop with cosmology module
  - Run on real data

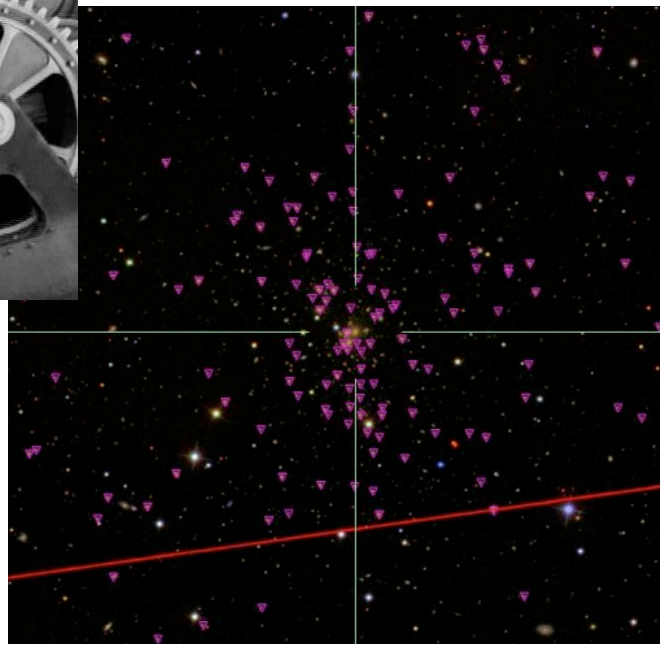
48

Thanks

49

Thanks

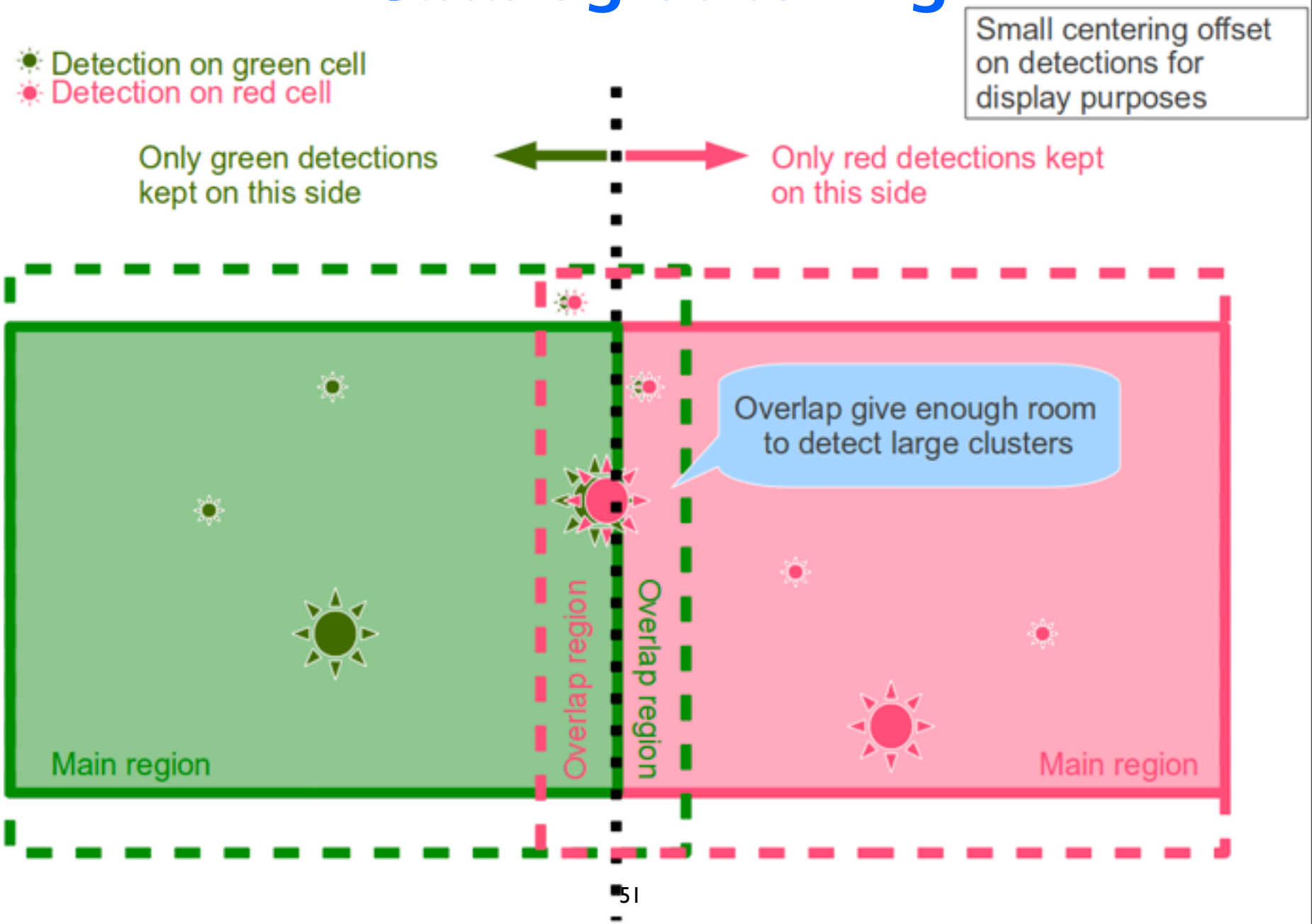




Extra Slides

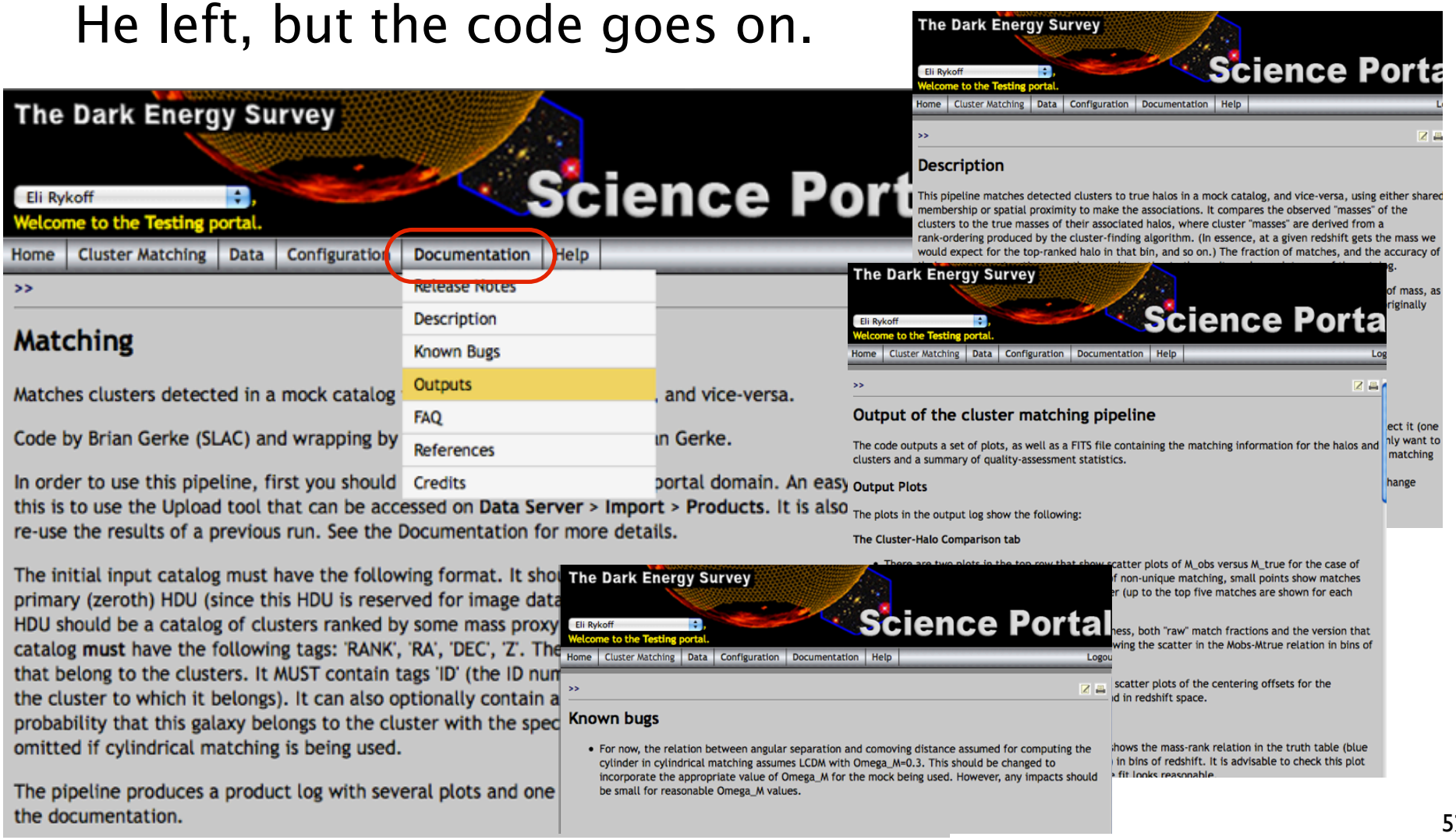
50

Catalog stitching



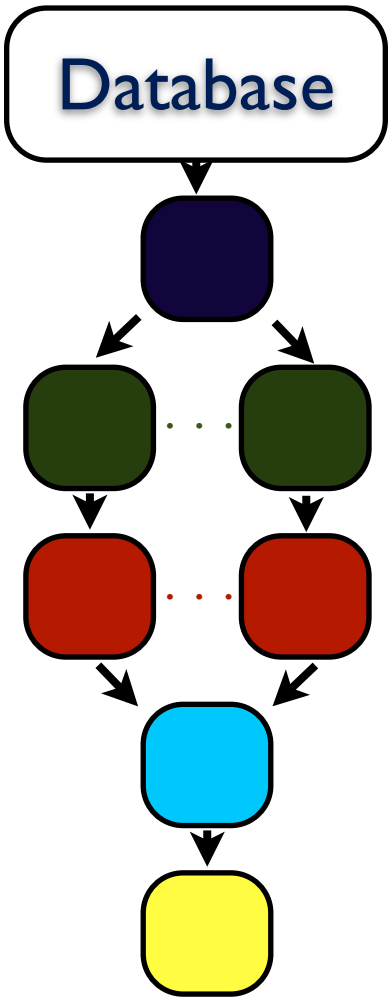
Legacy

Brian Gerke (SLAC) was leading this project.
He left, but the code goes on.



52

Parallelization in the portal



Data Organizer: multiple queries of configurable area size	1toN
Data Retriever: execute queries and produces FITS files	1to1
Cluster Finder	1to1
Consolidator: join catalogs	Nto1
Cluster Comparison, Plots...	1to1

53