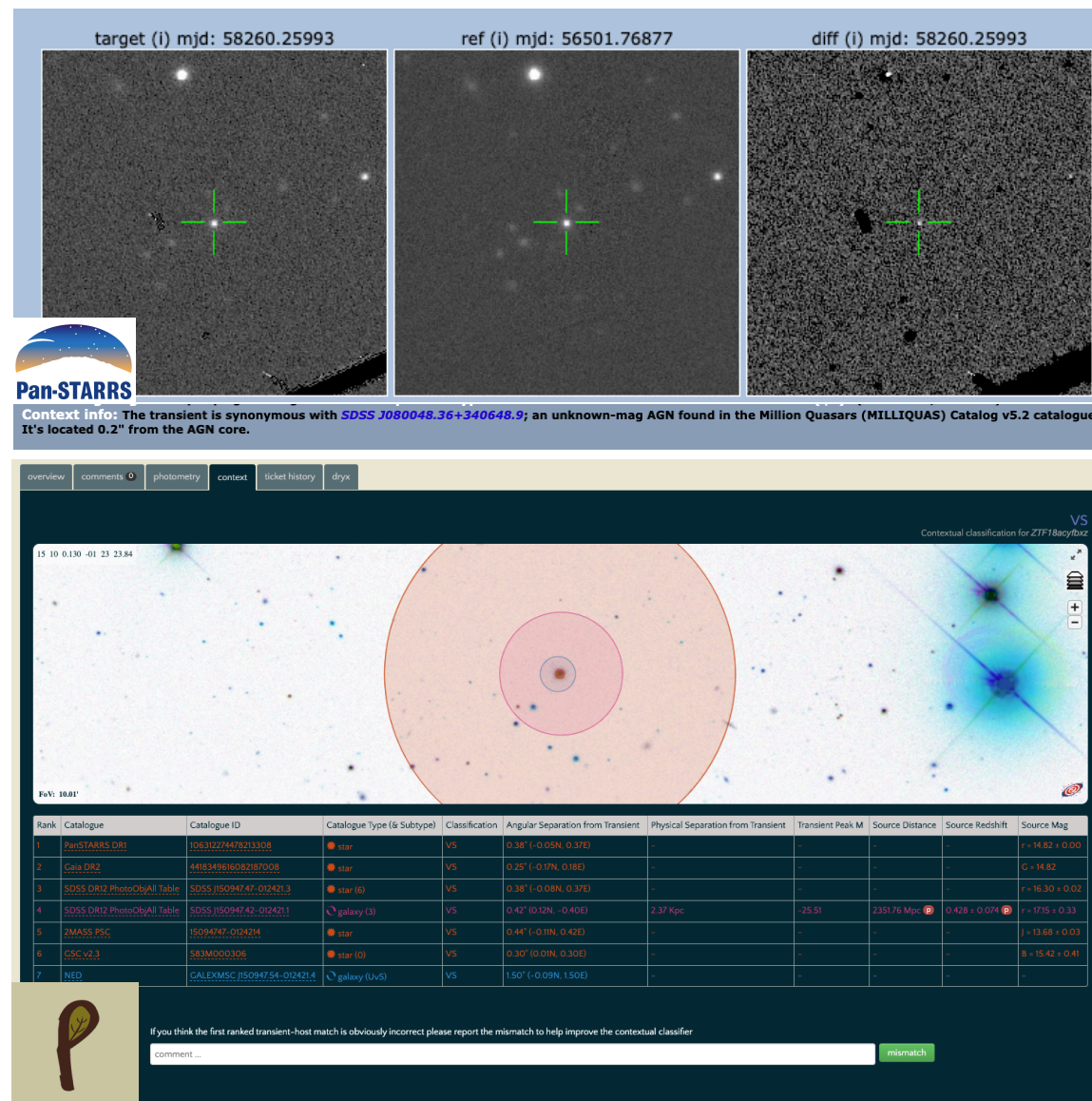# Sherlock - Contextual Transient Classifier

> *a boosted decision tree algorithm that mines a library of historical and on-going astronomical survey data in an attempt to predict the nature of a transient object based on the context derived from the resulting crossmatched associations.*

- Classified as SN at distance 12.05 arcsec.
-
- The transient is possibly associated with *LCRSB143138.2-063929*; a 15.55 mag galaxy found in the NED-D Galaxy Catalogue v13.1. It's located 2920.30 S, 5017.30 E (4.9 Kpc) from the galaxy centre. A host distance of 83.5 Mpc implies a transient M = -15.54.
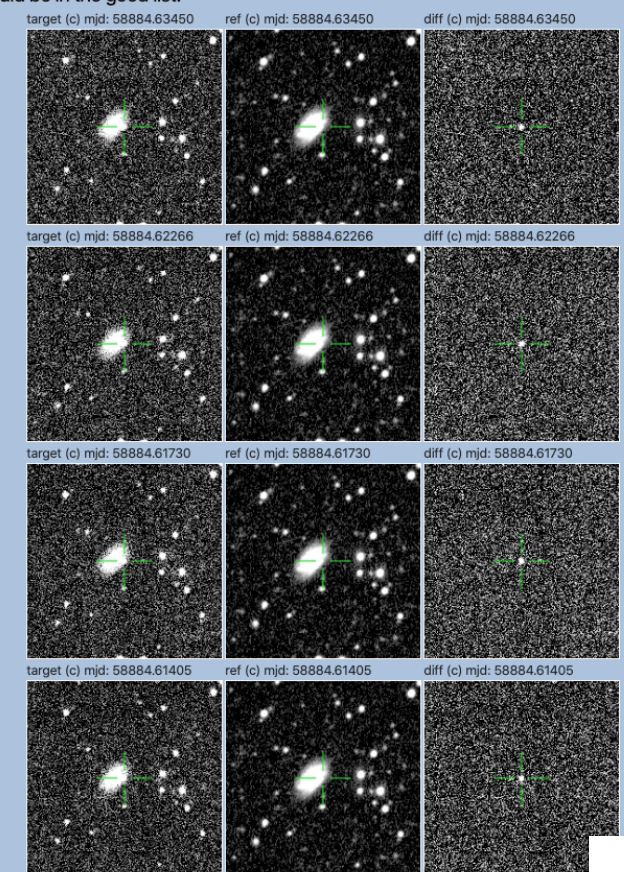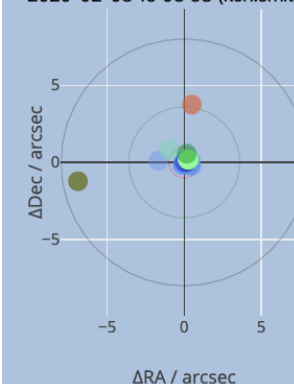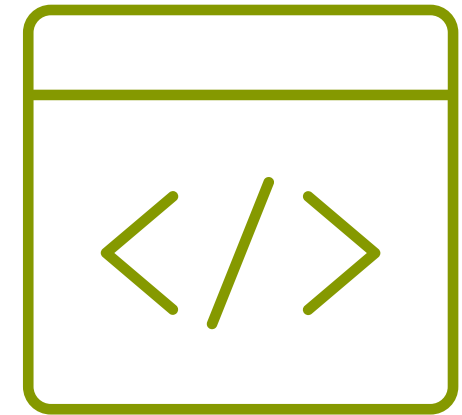
3 Core Components

Catalogue
Database
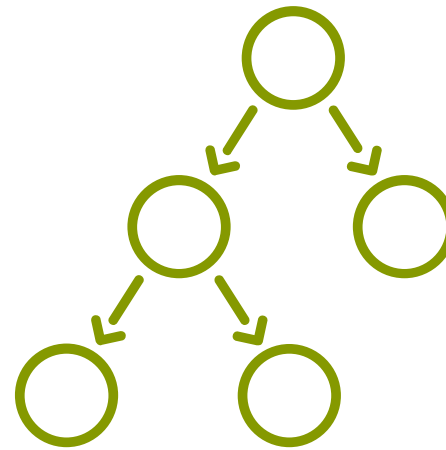
Algorithm
Parameters

Code

Catalogue Database

all-sky survey catalogues
```
PS1 DR1, Gaia DR2, SDSS DR12 PhotoObjAll Table, SDSS DR12
SpecObjAll Table, GSC v2.3, 2MASS XSC & PSC
```

source-specific catalogues
```
Million Quasars Catalog v5.2, Veron-Cett AGN Catalogue v13,
Downes Catalog of CVs, Ritter Cataclysmic Binaries Catalog
v7.21, The GLADE Galaxy Catalogue v2.3, NED-D Galaxy Catalogue
v13.1
```

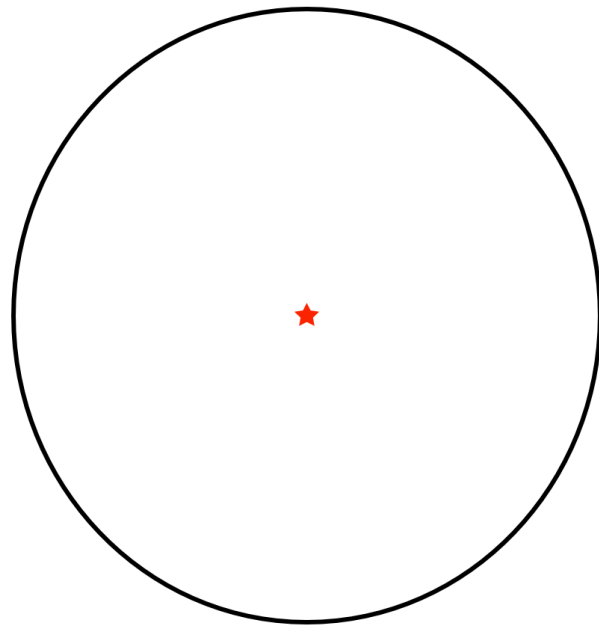NASA/IPAC Extragalactic Database - remote query with local caching

Catalogue Database

- one table per catalogue
- tables (and views) follow a strict naming scheme
- 4.6TB storage footprint - this includes many legacy tables
- larger catalogues host a minimal column set to reduce footprint and index sizes
- indexed with 3 HTM levels (10, 13, 16)

```python
# CONSTANTS
pi = (4 * math.atan(1.0))
DEG_TO_RAD_FACTOR = pi / 180.0
RAD_TO_DEG_FACTOR = 180.0 / pi


aa = (90.0 - dec1) * DEG_TO_RAD_FACTOR
bb = (90.0 - dec2) * DEG_TO_RAD_FACTOR
cc = (ra1 - ra2) * DEG_TO_RAD_FACTOR
one = math.cos(aa) * math.cos(bb)
two = math.sin(aa) * math.sin(bb) * math.cos(cc)

# Because acos() returns NaN outside the ranges of -1 to +1
# we need to check this.  Double precision decimal places
# can give values like 1.0000000000002 which will throw an
# exception.
three = one + two
if (three > 1.0):
    three = 1.0
if (three < -1.0):
    three = -1.0

# BE CAREFUL WITH PRECISION PROPAGATION
thisVal = math.acos(three)
angularSeparation = float(thisVal) * RAD_TO_DEG_FACTOR * 3600.0

angularSeparation = "%0.*f" % (precision, angularSeparation)
```
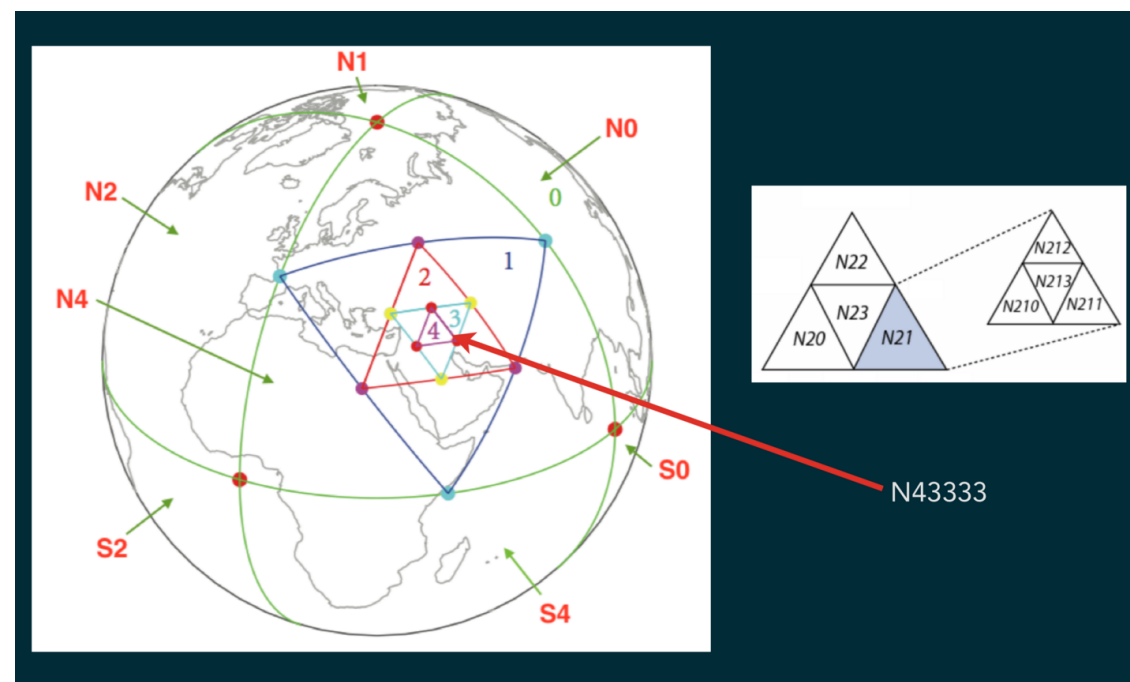
$\times 2.9e^{9}$

Catalogue Database

- one table per catalogue
- tables (and views) follow a strict naming scheme
- 4.6TB storage footprint - this includes many legacy tables
- larger catalogues host a minimal column set to reduce footprint and index sizes
- indexed with 3 HTM levels (10, 13, 16)

```
select * from big_catalogue where htmID
in (123,1234,5673,7823 …)
```
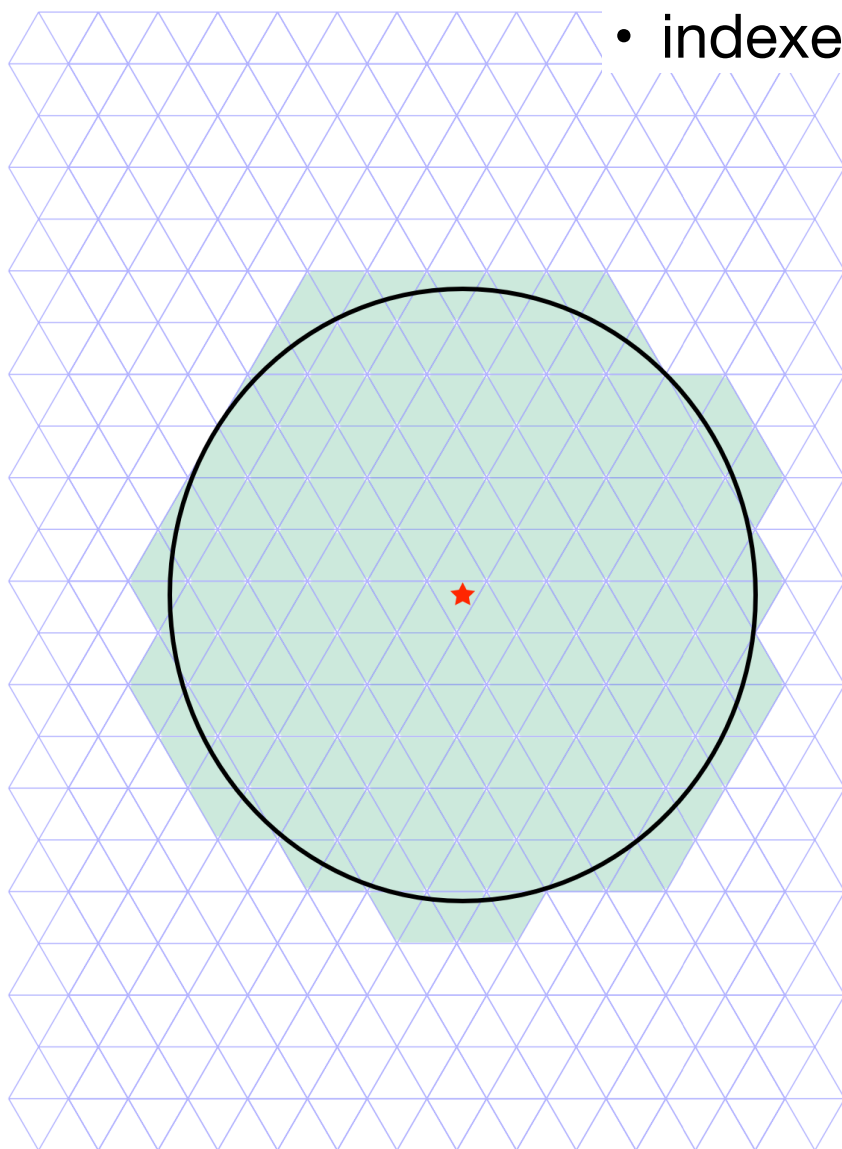
Catalogue Database

- one table per catalogue
- tables (and views) follow a strict naming scheme
- 4.6TB storage footprint - this includes many legacy tables
- larger catalogues host a minimal column set to reduce footprint and index sizes
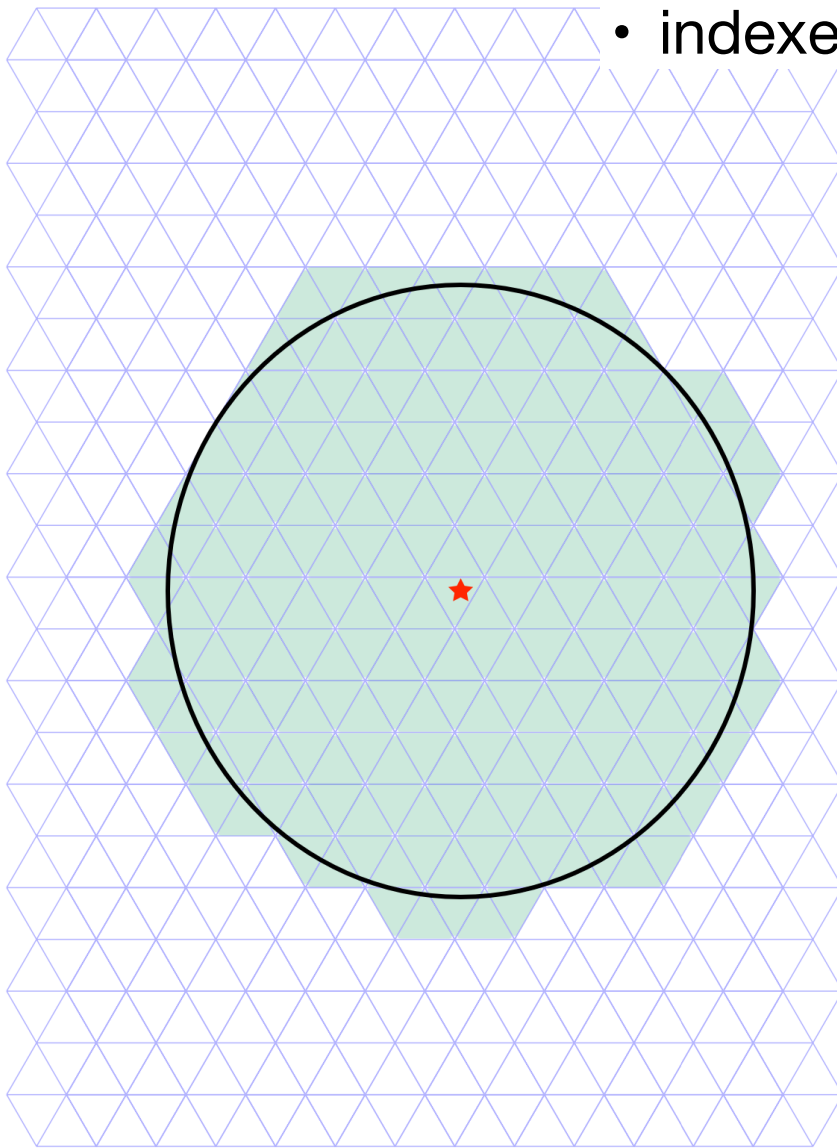- indexed with 3 HTM levels (10, 13, 16)

MySQL™


```
# CONSTANTS
pi = (4 * math.atan(1.0))
DEG_TO_RAD_FACTOR = pi / 180.0
RAD_TO_DEG_FACTOR = 180.0 / pi


aa = (90.0 - dec1) * DEG_TO_RAD_FACTOR
bb = (90.0 - dec2) * DEG_TO_RAD_FACTOR
cc = (ra1 - ra2) * DEG_TO_RAD_FACTOR
one = math.cos(aa) * math.cos(bb)
two = math.sin(aa) * math.sin(bb) * math.cos(cc)

# Because acos() returns NaN outside the ranges of -1 to +1
# we need to check this.  Double precision decimal places
# can give values like 1.0000000000002 which will throw an
# exception.
three = one + two
if (three > 1.0):
    three = 1.0
if (three < -1.0):
    three = -1.0

# BE CAREFUL WITH PRECISION PROPAGATION
thisVal = math.acos(three)
angularSeparation = float(thisVal) * RAD_TO_DEG_FACTOR * 3600.0

angularSeparation = "%0.*f" % (precision, angularSeparation)
```
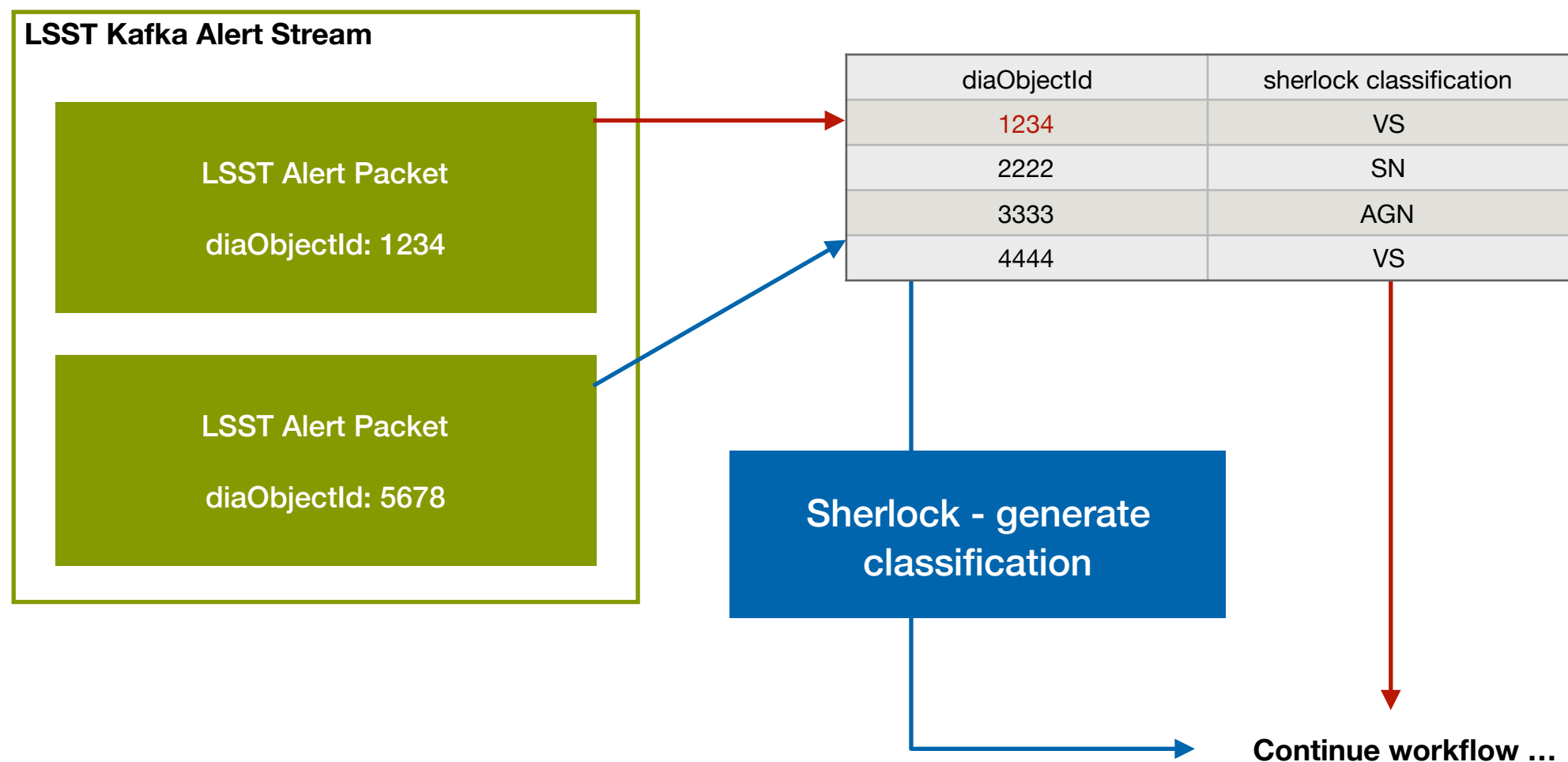
x ~100

Side Note on Classification Caching

LSST Kafka Alert Stream

LSST Alert Packet

diaObjectId: 1234

LSST Alert Packet

diaObjectId: 5678

diaObjectId	sherlock classification
1234	VS
2222	SN
3333	AGN
4444	VS

Sherlock - generate classification

Continue workflow …

Integer matching is cheap!

Catalogue Database

- helper-table that maps heterogeneous columns in catalogues to a custom defined set of column-names (RA, DEC, redshift, distance, magnitudes ….)
- Database views (attempt to) define specific source types
- Quality of the data between catalogues and intra-catalogue vary vastly. Important to understand strengths, weaknesses and foibles of each catalogue.
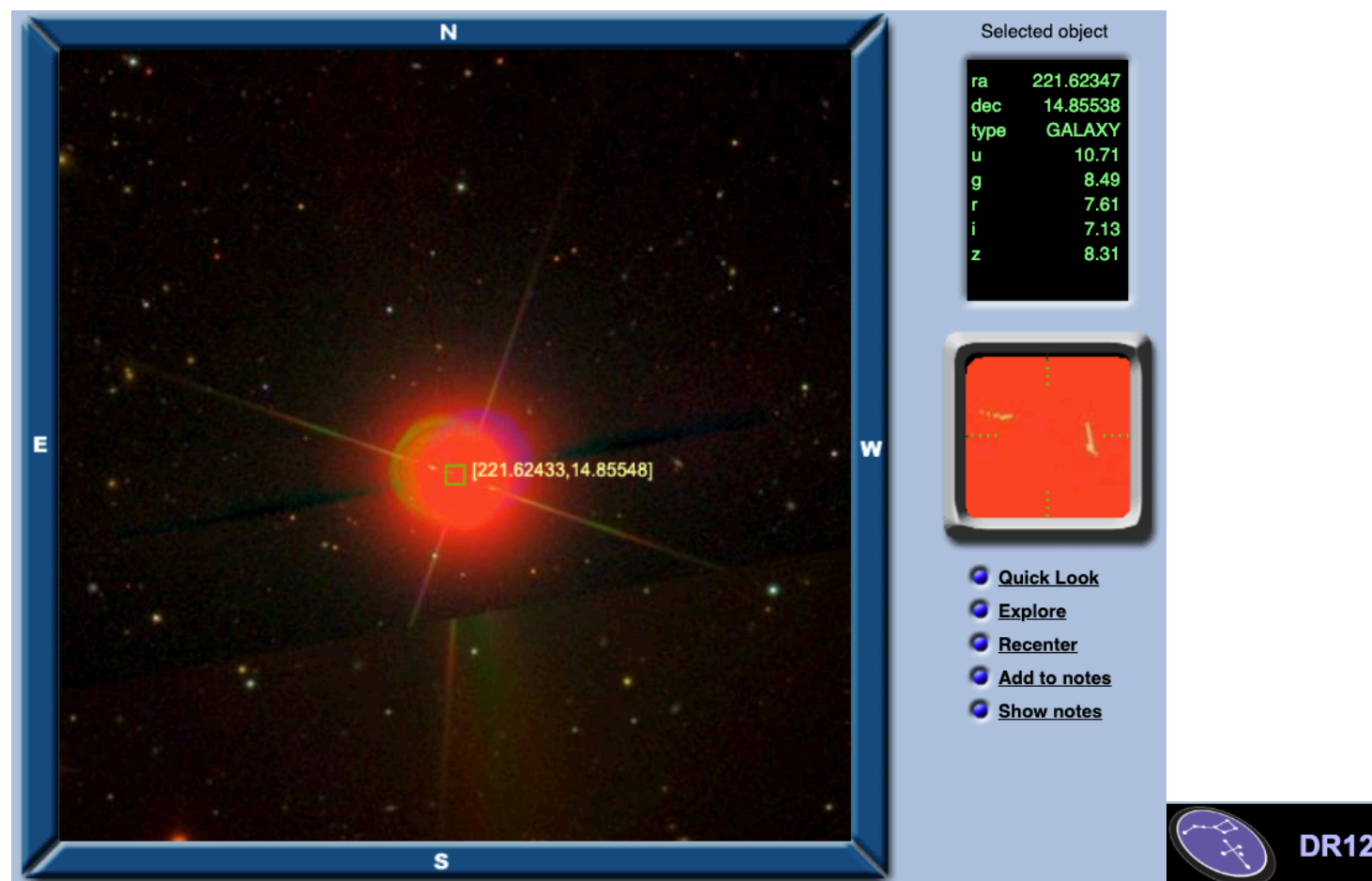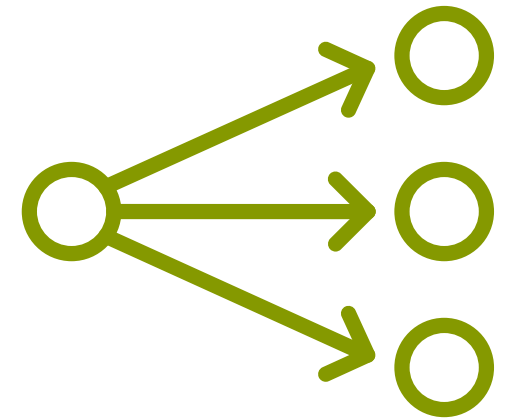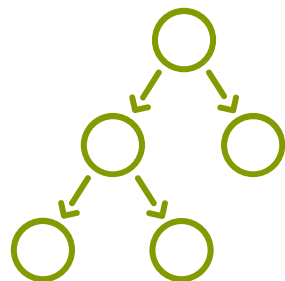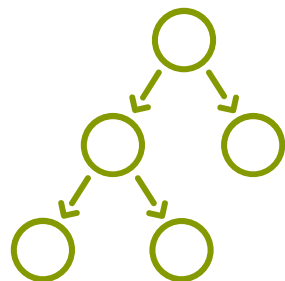
Algorithm Parameters

synonyms vs associations

Examples of synonym matches are CVs, AGN and variable stars (VS)

Examples of associations matches are SNe and bright-stars (BS)

Transient Classifications

1. ***Variable Star (VS)*** if the transient lies within the synonym radius of a catalogued point-source,
2. ***Cataclysmic Variable (CV)*** if the transient lies within the synonym radius of a catalogued CV,
3. ***AGN*** if the transient lies within the synonym radius of a catalogued AGN,
4. ***Bright Star (BS)*** if the transient is not matched against the synonym radius of a star but is associated within the magnitude-dependent association radius,
5. ***Nuclear Transients (NT)*** if the transient falls within the synonym radius of the core of a resolved galaxy,
6. ***Supernova (SN)*** if the transient is not classified as an NT but is found within the magnitude-, morphology- or distance-dependant association radius of a galaxy, or
7. ***Orphan*** if the transient fails to be matched against any catalogued source.

Algorithm Parameters

crafting meaning from crossmatches

```yaml
search algorithm:
    GSC star 1:
        database table: tcs_view_star_guide_star_catalogue_v2_3
        mag column: V
        bright:
            mag limit: 16.
            angular radius arcsec: 100.0
            synonym: VS
            association: BS
        faint:
            mag limit: 19.5
            angular radius arcsec: 2.0
            annotation: SN
        general:
            angular radius arcsec: 0.5
            synonym: VS
    GSC star 2:
        database table: tcs_view_star_guide_star_catalogue_v2_3
        mag column: B
        bright:
            mag limit: 16.
```

Search Algorithms are written in YAML - plain text

Search Module Parameters

- *angular separation crossmatch radius*
- *physical separation crossmatch radius*
- *source magnitude filtering*
- *magnitude dependent search radii for bright stars and galaxies*

Classifications & Reliabilities

- *transient given a predicted classification based on the parameters of the catalogued source it matches against*
- *transients can be given multiple classifications which are later ranked*
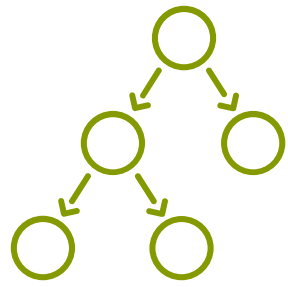
Algorithm Parameters

contextual classification is harder than you think!
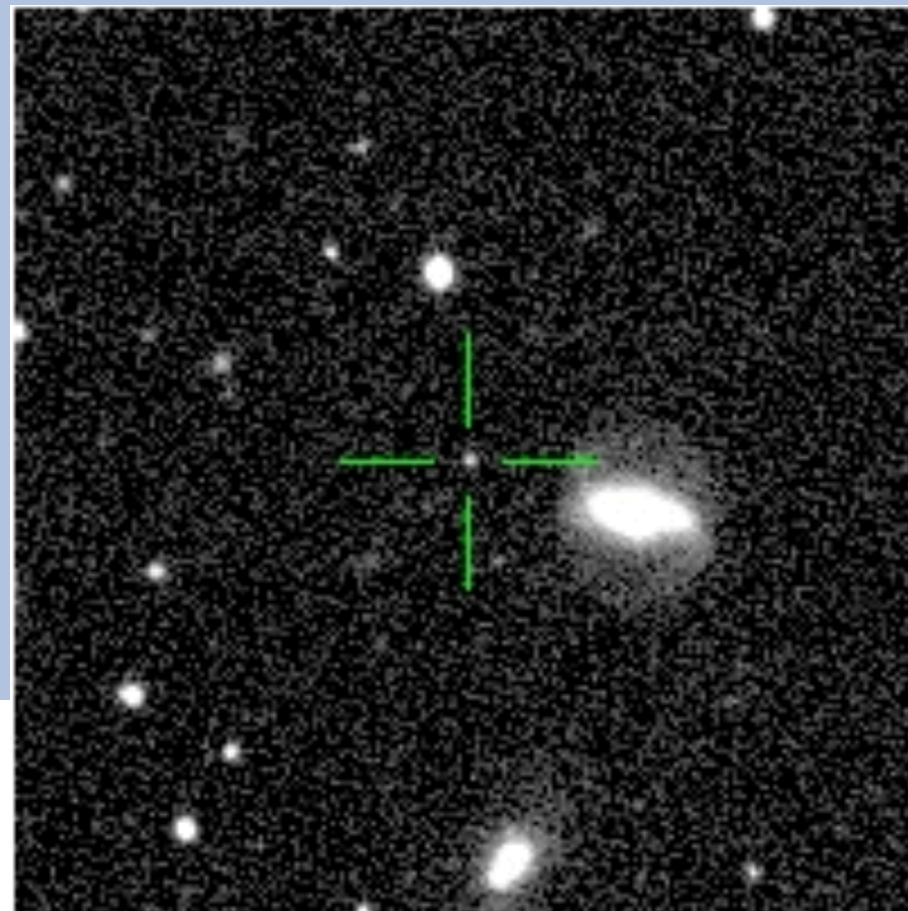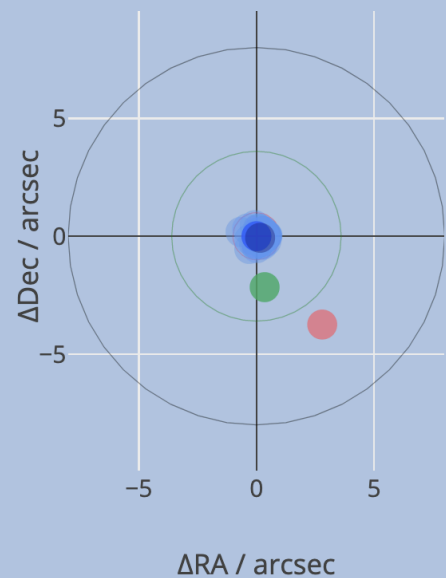

Spectral Type:

SN Ia

Context:

ORPHAN

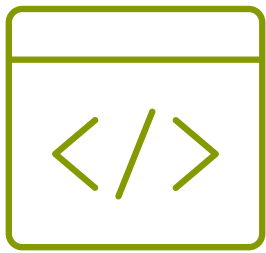The transient location is not matched against any known catalogued source

Comments:

2020-03-03 10:40:47 (michael.fulton): Looks like a very good object to me. Why classed as an orphan though?!
2020-03-03 10:50:06 (ken.smith): This is a fast track object - possibly associated with NGC 5804 at 59Mpc.

- angular separation
- galaxy morphology
- physical-distance separations

Code

1. ***Parallel crossmatching*** of the transient location against each catalogue in the library.

2. ***Internal crossmatching*** of the results of stage 1. to find and merge the information from the same unique astrophysical sources recorded across multiple catalogues.

3. ***Ranking of the post-merge results*** based on the secureness of the classifications and the quality if the underlying catalogue data.

4. ***Reporting*** of the top ranked result as the most likely nature of the transient.

Lᴀꜱᴀɪʀ - Enhancement Suggestions

1. **Classification Caching**
2. **Visualisation**

Questions/Discussion