

Crowdsourcing and the VO

Matthew J. Graham (Caltech, NVO)
et

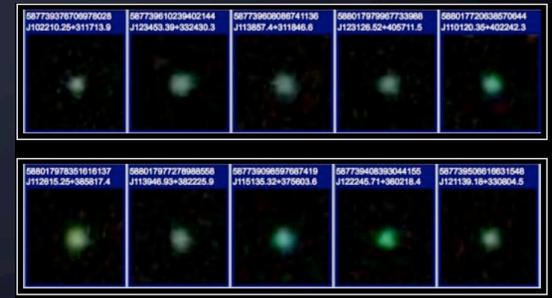
Roy Williams, Andrew Drake, George Djorgovski
Ashish Mahabal, Ciro Donalek

Humans as CPUs

- Unique juncture in history of science:
 - technological capability exists to network large numbers of people
 - data volumes and complexity are still desktop manageable
 - class of problems that are resistant to present machine learning solutions
- Crowdsourcing/human computation/citizen science projects exploit efforts of volunteers to attack particular areas, e.g. image analysis
- Axes:
 - Sweat shop vs. GWAP
 - Idiot vs. savant

Tonight a galaxy, tomorrow the Zoo

- Initial questions:
 - Are galaxies elliptical or spiral?
 - If spiral, rotating clockwise or anticlockwise?
- 34617406 clicks done by 82931 users
- Main result:
 - Spiral galaxies which share a neighbourhood (a region defined as 65 million light years across) are likely to rotate in the same direction – but only if they formed the vast majority of their stars more than 10 billion years ago.
- Other results:
 - Hanny's Voorwerp
 - Green Peas



Galaxy Zoo 2

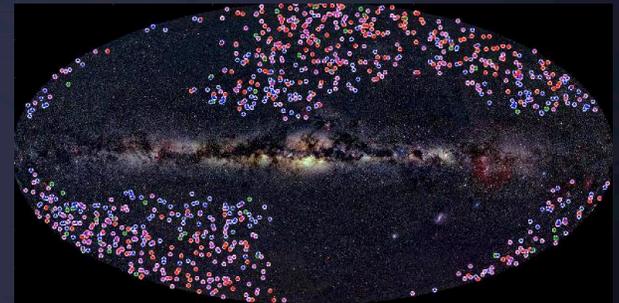
The screenshot displays the Galaxy Zoo 2 website interface. At the top, the title "GALAXY ZOO" is prominently displayed in a stylized font, with a small number "2" to its right. Below the title is a navigation menu with links for "Home", "How To Take Part", "My Galaxies", and "Contact Us". On the right side of the menu, there are buttons for "Profile" and "Logout".

The main content area is divided into two sections. On the left, a large image of a galaxy is shown. Below the image are two buttons: "Invert galaxy image" and "Add to my favourites". On the right, a panel titled "Classify galaxies" contains the following text: "Answer the question below using the buttons provided." Below this is the question "Is there anything odd?".

Two buttons are provided for the question: a checkmark icon labeled "Yes" and an "X" icon labeled "No". Below these buttons, the text reads: "Please click an image below to return to an earlier point in the classification". This is followed by a grid of seven galaxy icons. The second icon from the left in the top row is marked with a red "X". At the bottom right of the panel, there is a "Need help?" link with a question mark icon.

Things that go BANG! in the night

- Catalina Real-Time Transient Survey (<http://crts.caltech.edu>)
 - Repeatedly surveys $\sim 26000 \text{ deg}^2$
 - 3 telescopes: MLS (1.5m), CSS (0.7m), SSS (0.5m)
 - 1067 new discoveries to date
 - Only completely public transient survey
- SkyAlert (<http://www.skyalert.org>)
 - enables users to perform complex queries about discoveries in order to receive personally tailored and filtered event streams.
- The VO is useful for:
 - data discovery
 - semantics
 - data mining



Citizen science with CRTS

CRTS (Catalina) Event 6006

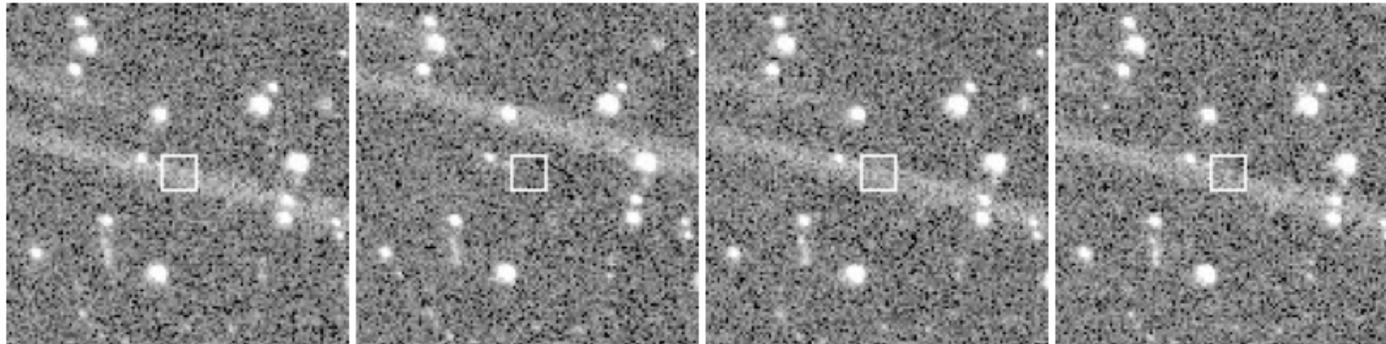
CRTS (Catalina) Event 6774.

Logg

Logged in as mjpg

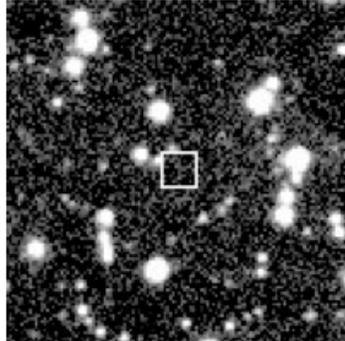
New

New images



Refe

Reference Image



First we would like to understand if the set of images that you see represents a real astronomical object, such as a star getting much brighter, or if it is an artifact or other man-made cause that is nothing to do with what is really happening in the deep sky.

Does this look like a real star field? Is it [Real](#) or is it [Artifact?](#)

Further questions about the nature of the artifact:

- Is there a bright star in the field, or just outside, that may be confusing things? [...more](#)
- Is it an edge? [...more](#)
- Is there a line across the central box? [...more](#)
- Is there a satellite trail across the image? [...more](#)

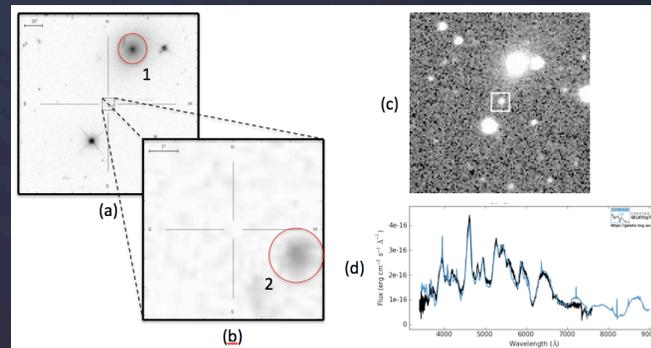
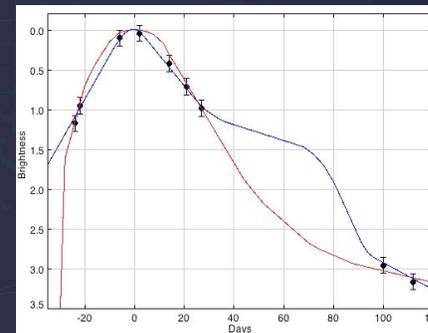
Click to submit results:

[Skip this one, try another ...](#)

[Show all explanation](#)

AstroCollation – I

- Next generation collaborative science venture
- Data mining algorithms applied to transient event data to produce conceptual models describing them
- Models presented to citizen scientists for value judgements, deciding which of a set of models provides the best description
- Citizen scientists can also provide contextual information to aid the classification process



AstroCollation – II

- Decisions and information factored back into the system and consolidated to produce a *consensus* description of an event that can always be retrieved (and reused)
- Produce better (ideal) training sets
- Built upon semantic technologies, CRTS and SkyAlert
- Issues to address:
 - How to formally represent uncertainty in data and *description* in a machine-processible fashion
 - Optimal method to achieve consensus opinion