Astronomy and Computing 16 (2016) 99–108

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Crowdsourcing quality control for Dark Energy Survey images☆

P. Melchior[1,2,*], E. Sheldon[3], A. Drlica-Wagner[4], E.S. Rykoff[5,6], T.M.C. Abbott[7], F.B. Abdalla[8,9], S. Allam[4], A. Benoit-Lévy[10,8,11], D. Brooks[8], E. Buckley-Geer[4], A. Carnero Rosell[12,13], M. Carrasco Kind[14,15], J. Carretero[16,17], M. Crocce[16], C.B. D'Andrea[18,19], L.N. da Costa[12,13], S. Desai[20,21], P. Doel[8], A.E. Evrard[22,23], D.A. Finley[4], B. Flaugher[4], J. Frieman[4,24], E. Gaztanaga[16], D.W. Gerdes[23], D. Gruen[25,26], R.A. Gruendl[14,15], K. Honscheid[1,2], D.J. James[7], M. Jarvis[27], K. Kuehn[28], T.S. Li[29], M.A.G. Maia[12,13], M. March[27], J.L. Marshall[29], B. Nord[4], R. Ogando[12,13], A.A. Plazas[30], A.K. Romer[31], E. Sanchez[32], V. Scarpine[4], I. Sevilla-Noarbe[32,14], R.C. Smith[7], M. Soares-Santos[4], E. Suchyta[27], M.E.C. Swanson[15], G. Tarle[23], V. Vikram[33], A.R. Walker[7], W. Wester[4], Y. Zhang[23]

[1] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[2] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[3] Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA
[4] Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA
[5] Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA
[6] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[7] Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile
[8] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[9] Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa
[10] CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[11] Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[12] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[13] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[14] Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA
[15] National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
[16] Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Bellaterra, Barcelona, Spain
[17] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
[18] Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[19] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[20] Excellence Cluster Universe, Boltzmannstr. 2, 85748 Garching, Germany
[21] Faculty of Physics, Ludwig-Maximilians University, Scheinerstr. 1, 81679 Munich, Germany
[22] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[23] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[24] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[25] Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany
[26] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany
[27] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[28] Australian Astronomical Observatory, North Ryde, NSW 2113, Australia
[29] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[30] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
[31] Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK

☆ This code is registered at the ASCL with the code entry ascl:1511.017.

* Corresponding author at: Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA.
E-mail address: peter@pmelchior.net (P. Melchior).

http://dx.doi.org/10.1016/j.ascom.2016.04.003

[32] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[33] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

ARTICLE INFO

Article history:
Received 23 November 2015
Accepted 22 April 2016
Available online 11 May 2016

Keywords:
Surveys
Information systems: Crowdsourcing
Human-centered computing: Collaborative filtering

ABSTRACT

We have developed a crowdsourcing web application for image quality control employed by the Dark Energy Survey. Dubbed the "DES exposure checker", it renders science-grade images directly to a web browser and allows users to mark problematic features from a set of predefined classes. Users can also generate custom labels and thus help identify previously unknown problem classes. User reports are fed back to hardware and software experts to help mitigate and eliminate recognized issues. We report on the implementation of the application and our experience with its over 100 users, the majority of which are professional or prospective astronomers but not data management experts. We discuss aspects of user training and engagement, and demonstrate how problem reports have been pivotal to rapidly correct artifacts which would likely have been too subtle or infrequent to be recognized otherwise. We conclude with a number of important lessons learned, suggest possible improvements, and recommend this collective exploratory approach for future astronomical surveys or other extensive data sets with a sufficiently large user base. We also release open-source code of the web application and host an online demo version at http://des-exp-checker.pmelchior.net.

1. Introduction

Large astronomical surveys produce vast amounts of data for increasingly demanding science applications. At the same time, the complexity of the instruments, operations, and the subsequent data analyses renders glitches and flaws inevitable, particularly during the early phases of experiments. Thus, mechanisms that facilitate the discovery and reporting of problems in the data, whether they originate from unexpected instrumental behavior or insufficient treatment in software, are important components of a data quality program. Due to the often unexpected nature of these features, algorithmic approaches to identify artifacts are generally infeasible and human inspection remains necessary. Human-generated reports can then be fed back to algorithm developers and hardware experts to mitigate and eliminate problems whenever possible. For current and upcoming surveys, the demands of carefully inspecting sizable portions of the data volume exceed the capabilities of individual, or a small team of, data management experts.

Crowdsourcing has seen tremendous success in the past few years in many applications where a critical task cannot be performed by computers but where the amount of data to be gathered, processed, or analyzed exceeds the capabilities of even the most dedicated human. Examples can be found in non-profit, academic, commercial, or activist settings. In astronomy, one of the first implementations of crowdsourcing was to gather information about the 1833 Leonid meteor storm (Olmsted, 1834a,b; Littmann and Suomela, 2014). In recent years, widespread access to the internet has made such efforts easier to realize, allowing for larger crowds and quicker turnaround of results. The preeminent early adopter of this web-based mode of operations is the Galaxy Zoo project (Lintott et al., 2008), designed to visually classify the morphology of galaxies from the Sloan Digital Sky Survey (York et al., 2000). Galaxy Zoo has led to Zooniverse,[1] currently the largest online portal for citizen science projects. At the time of writing, Zooniverse findings have been published in over 80 articles across several disciplines of science.

We have built a web-based crowdsourcing application for image quality control for the Dark Energy Survey (DES; Dark Energy Survey Collaboration, 2005).[2] DES is a 5000 deg^2 survey in five photometric bands (*grizY*) operating from the Blanco 4 m telescope at the Cerro Tololo Inter-American Observatory (CTIO). Its 570 megapixel imager DECam (Flaugher et al., 2015) comprises 62 science CCDs (2048 × 4096 pixels) and 12 focus and guiding CCDs (2048 × 2048 pixels), covering a roughly hexagonal footprint of 3 deg^2. Each region of the survey footprint will be observed 10 times in each band over the course of five years. The total number of exposures is thus expected to be approximately 10^5, with a data volume in science images of order 100 TB. An overview of the data processing and management pipeline (DESDM) is given by Mohr et al. (2012) and Desai et al. (2012).

The application, dubbed the "DES exposure checker", is geared for a professional user base of several hundred scientists and seeks to identify flaws in the DES images that will be used for most science analyses. Problems discovered with this application can then be fixed in hardware or in subsequent data processing runs. Our approach ties in with other quality control efforts, which automatically analyze the latest exposures and flag cases of e.g. bad observing conditions (Honscheid et al., 2012; Diehl et al., 2014), and allow the inspection of final coadded data products, both images and object catalogs (Balbinot et al., 2012). Our concept profits from the experience of a prior ad hoc crowdsourcing effort in DES. During the so-called Science Verification phase in 2012, flaws in DES imaging have been identified by a small "eyeball squad", whose reports were relayed to DES operations and data management experts on a daily basis. While essential to improving the performance of the instrument during this early phase, the effort did not scale well with the increasing number of incoming images. In the remainder of this paper, we will show how to create a scalable solution to data quality control by providing an engaging user experience, while simultaneously maximizing the utility of the report collection. The reports resulting from the DES exposure checker have already been used to inform the modification of existing algorithms, the development of new algorithms, and a general improvement the quality of the DES data.

Concept

For many crowdsourcing applications, classification is the critical task performed by humans. Specifically, which class, of a predefined set, does a given test object belong to? To render this question accessible to as wide an audience as possible, the task

[1] https://www.zooniverse.org/.
[2] http://www.darkenergysurvey.org.