Managing the Public to Manage Data:

CITIZEN SCIENCE AND ASTRONOMY

PETER DARCH

9TH INTERNATIONAL DIGITAL CURATION CONFERENCE

SAN FRANCISCO, CALIFORNIA

26 FEBRUARY 2014

Overview

Citizen Cyberscience Projects (CCPs)

- Low and high granularity
- Challenges for CCPs

Overview

Citizen Cyberscience Projects (CCPs)

- Low and high granularity
- Challenges for CCPs

Introducing Galaxy Zoo

Overview

Citizen Cyberscience Projects (CCPs)

- Low and high granularity
- Challenges for CCPs

Introducing Galaxy Zoo

Securing credibility for volunteer-generated datasets

Overview

Citizen Cyberscience Projects (CCPs)

- Low and high granularity
- Challenges for CCPs

Introducing Galaxy Zoo

Securing credibility for volunteer-generated datasets

Implications for practice

Overview

Citizen Cyberscience Projects (CCPs)

- Low and high granularity
- Challenges for CCPs

Introducing Galaxy Zoo

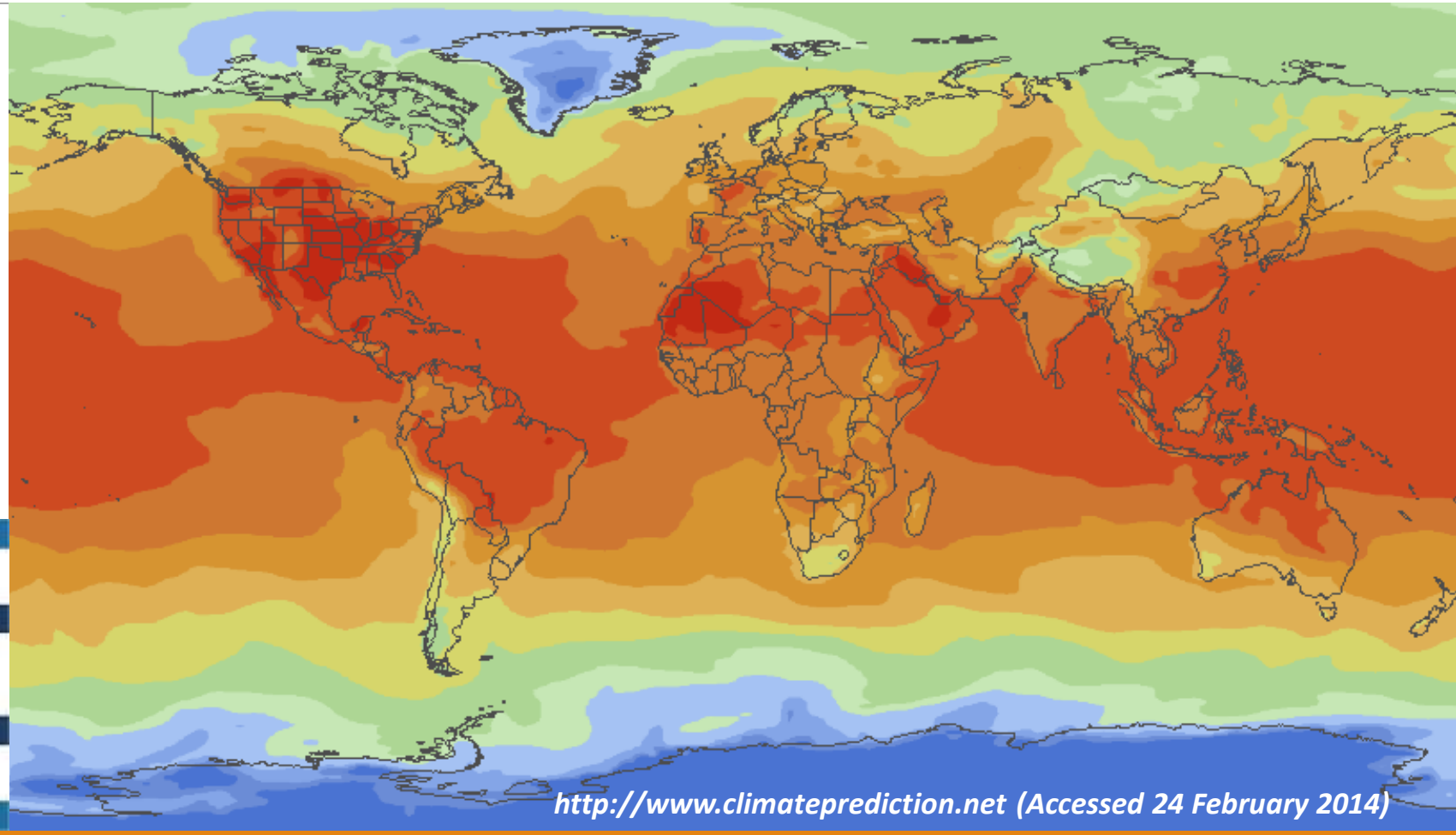
Securing credibility for volunteer-generated datasets

Implications for practice

Other areas of study

Granularity of projects (Nov et al. 2011)

Low granularity

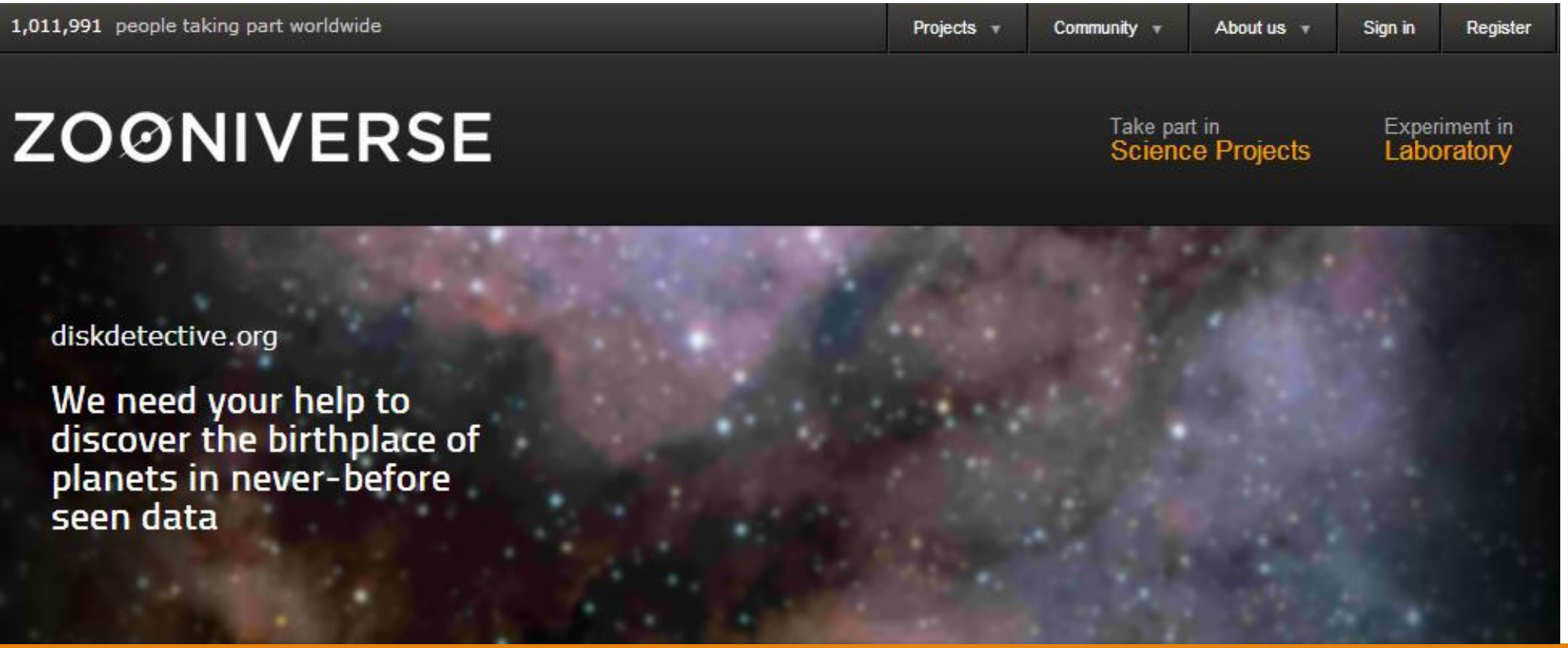


http://www.climateprediction.net (Accessed 24 February 2014)

Granularity of projects (Nov et al. 2011)

High granularity

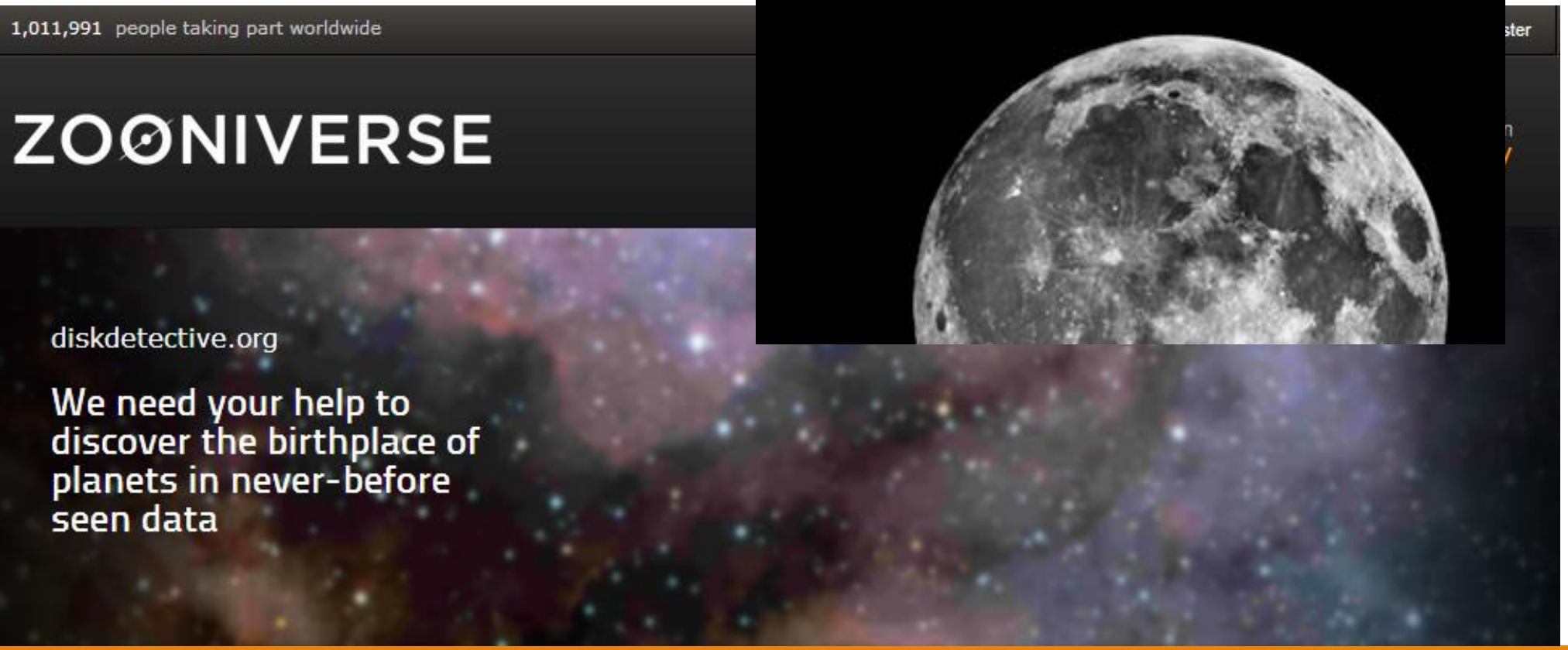
All images from: https://www.zooniverse.org/ (Accessed 24 February 2014)



Granularity of projects (Nov et al. 2011)

High granularity

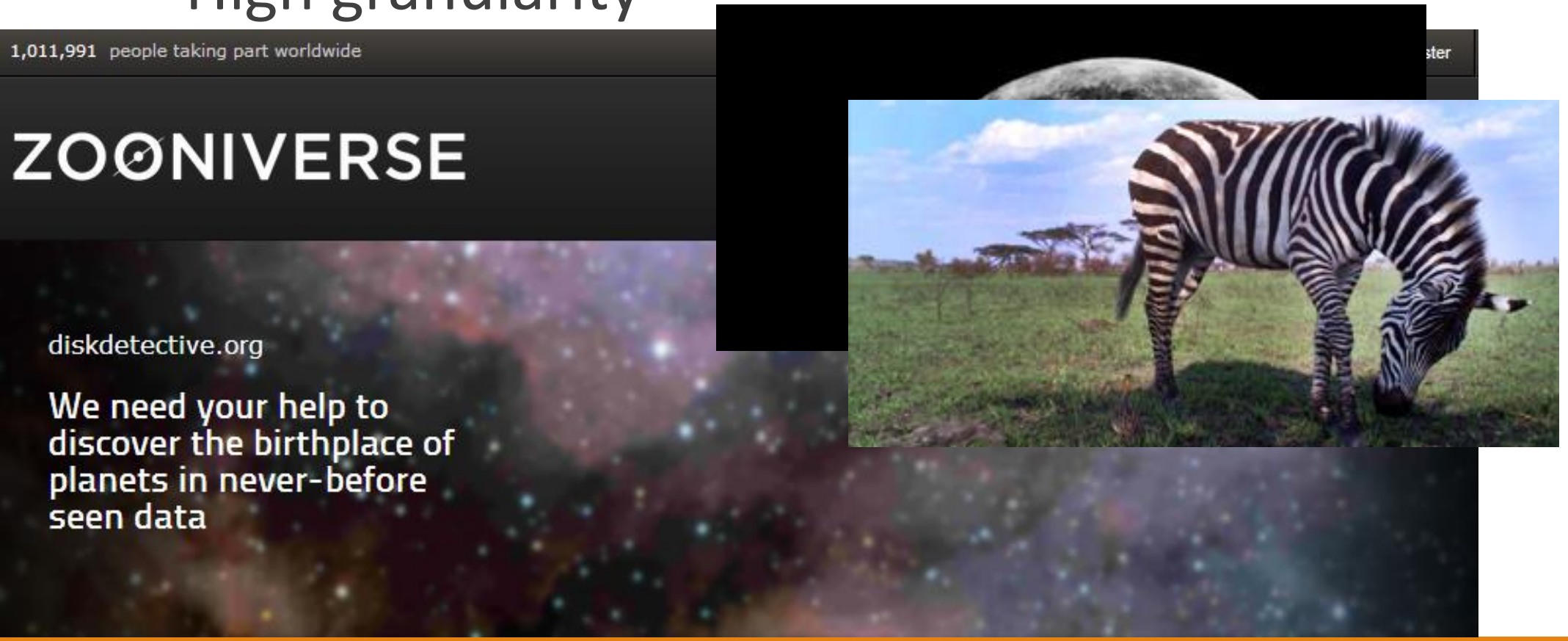
All images from: https://www.zooniverse.org/ (Accessed 24 February 2014)



Granularity of projects (Nov et al. 2011)

High granularity

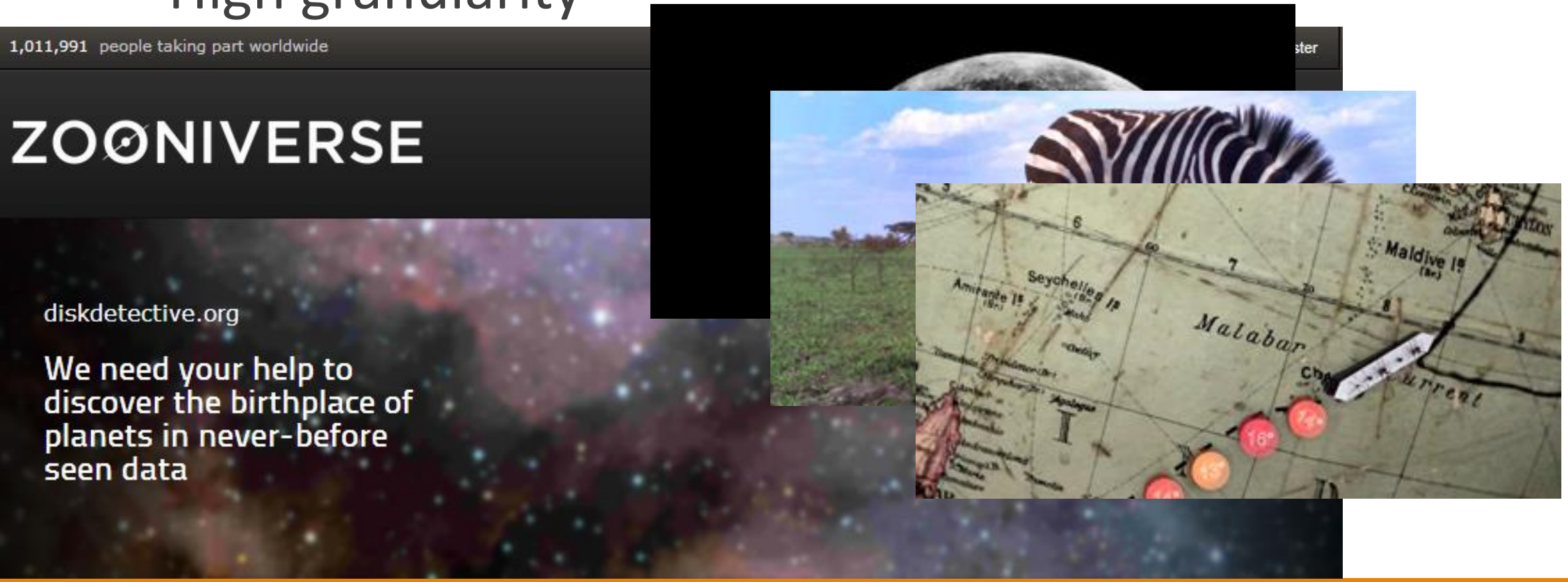
All images from: https://www.zooniverse.org/ (Accessed 24 February 2014)



Granularity of projects (Nov et al. 2011)

High granularity

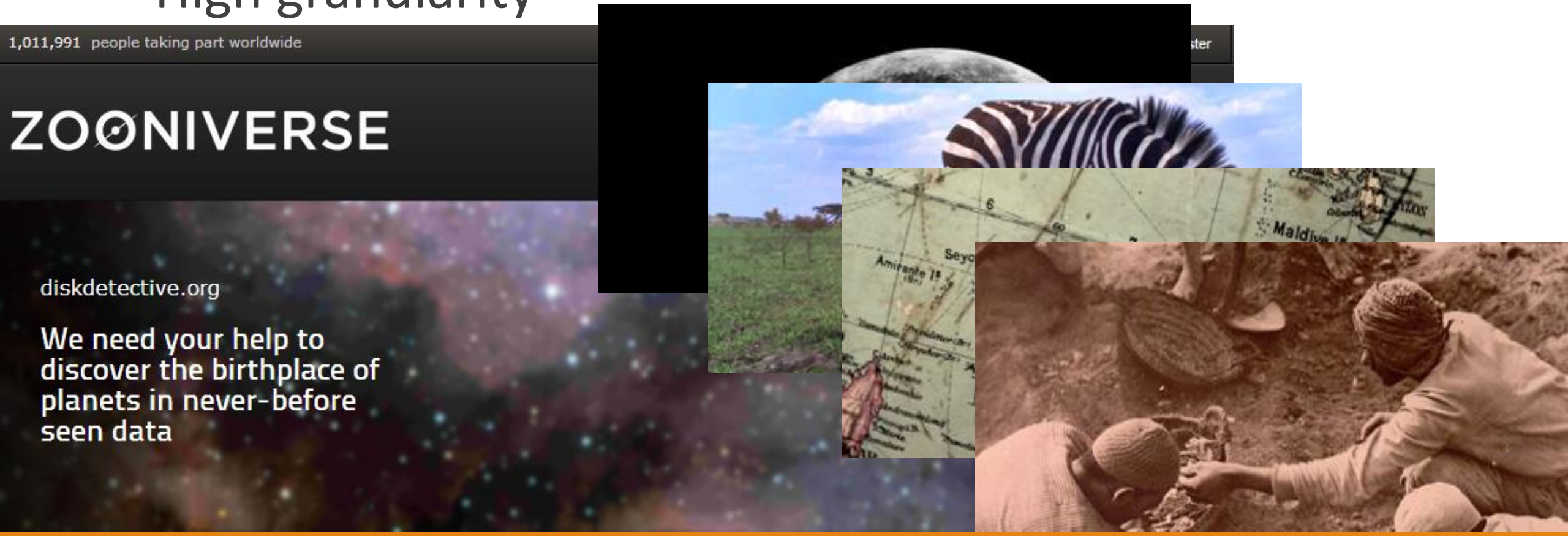
All images from: https://www.zooniverse.org/ (Accessed 24 February 2014)



Granularity of projects (Nov et al. 2011)

High granularity

All images from: https://www.zooniverse.org/ (Accessed 24 February 2014)



Challenges for Citizen Cyberscience

Increase quantity of scientific work

- Technical solutions
- Volunteer motivations

Challenges for Citizen Cyberscience

Increase quantity of scientific work

- Technical solutions
- Volunteer motivations

Ethical issues of volunteer credit

Challenges for Citizen Cyberscience

Increase quantity of scientific work

- Technical solutions
- Volunteer motivations

Ethical issues of volunteer credit

Securing credibility for CCP-generated datasets

Challenges for Citizen Cyberscience

Increase quantity of scientific work

- Technical solutions
- Volunteer motivations

Ethical issues of volunteer credit

Securing credibility for CCP-generated datasets

Credibility of datasets

What methods do scientists in a Citizen Cyberscience Project employ to manage volunteers' behaviour in order to secure credibility for volunteer-generated datasets?

http://www.galaxyzoo.org (Accessed 24 February 2014)

Galaxy Zoo is a **Zooniverse** project. | Our Projects

username | password | Login | Sign Up

Forgot Password?

CLASSIFY | STORY | SCIENCE | GALAXY ZOO | DISCUSS | PROFILE | LANGUAGE

Few have witnessed what you're about to see

Experience a privileged glimpse of the distant universe as observed by the SDSS, the Hubble Space Telescope, and UKIRT

GALAXY ZOO QUENCH

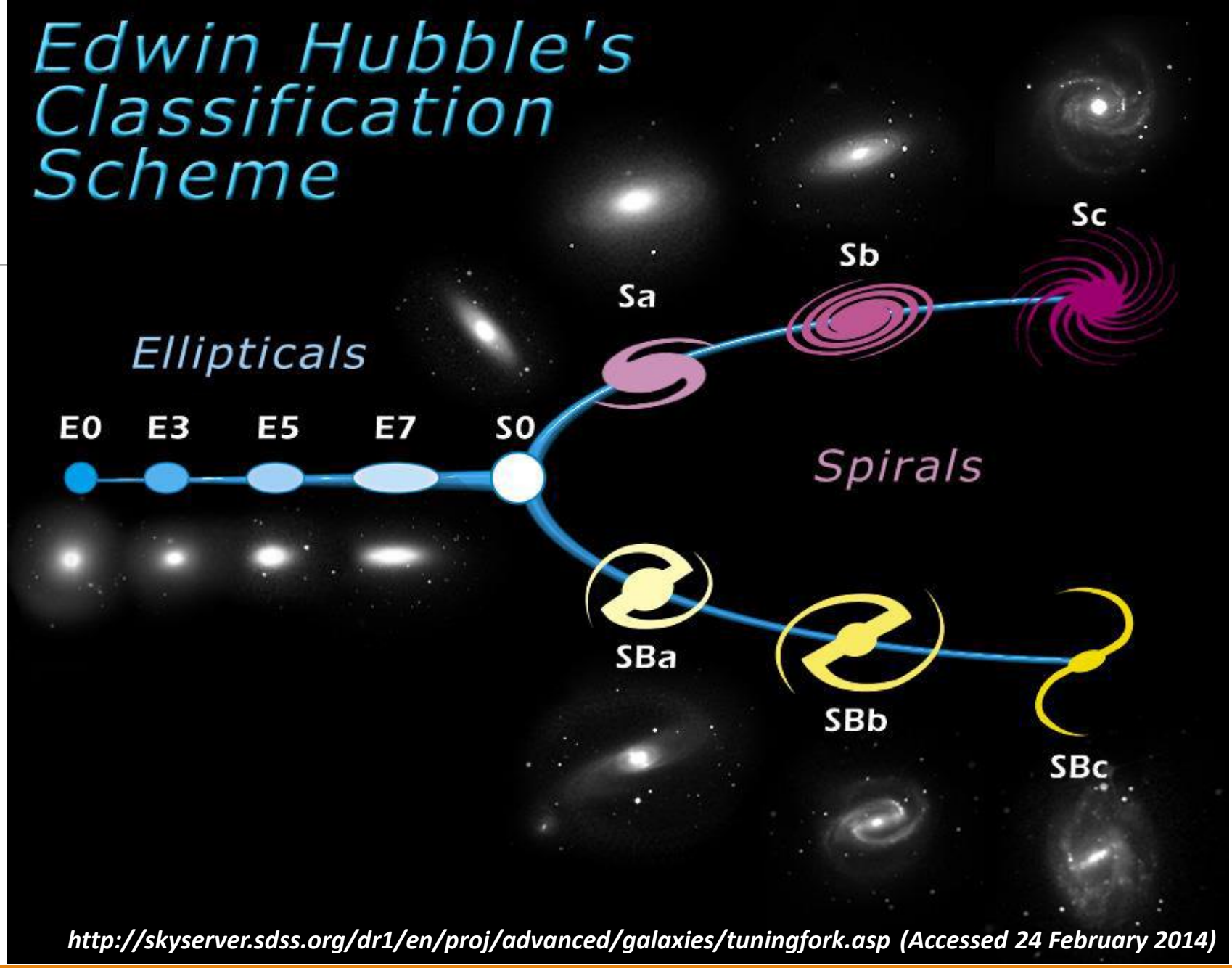
We are trying something new! Come help us understand a very specific type of galaxy and experience science from start to end. Take part

Classify Galaxies

To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify.

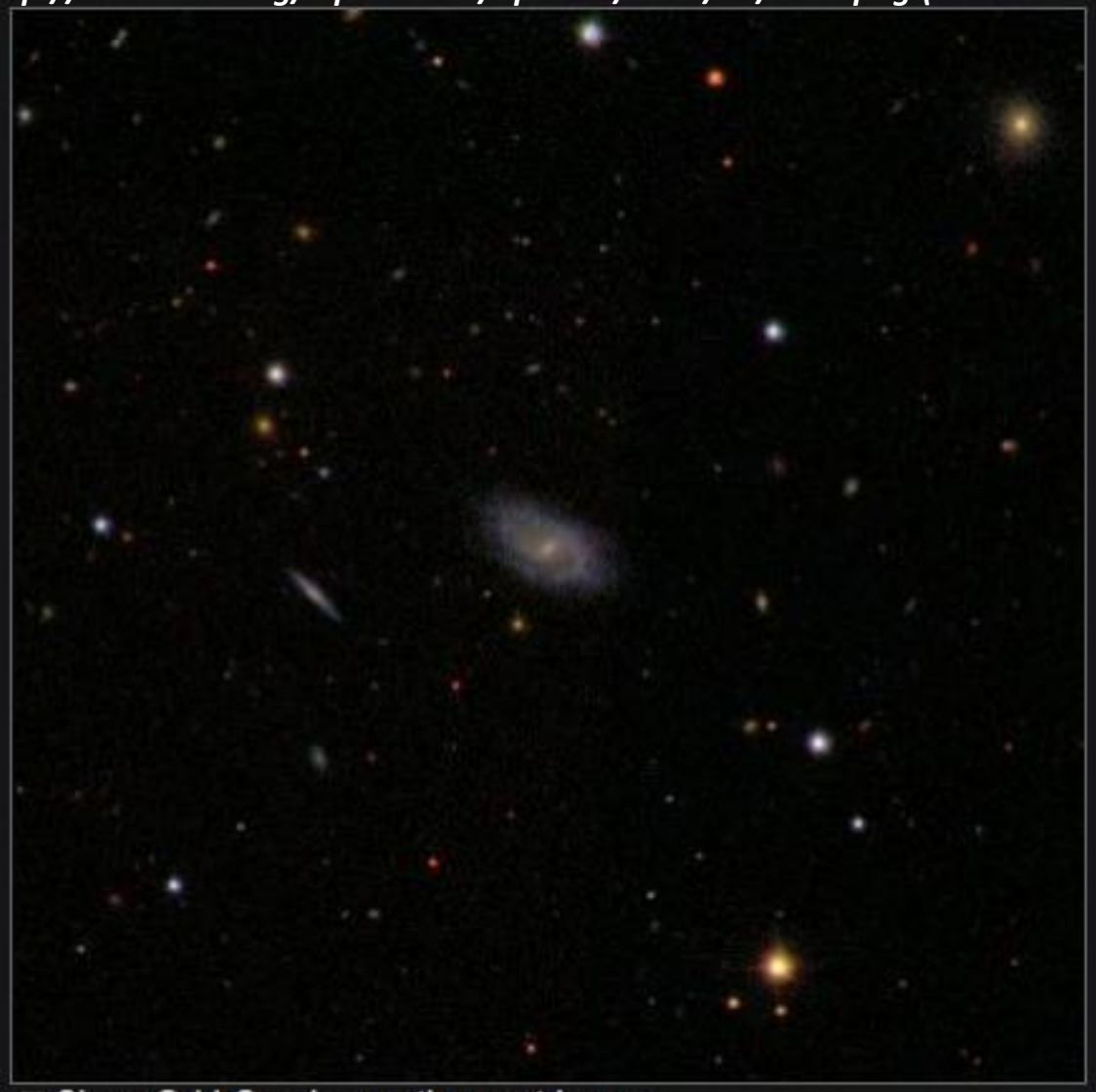
Begin Classifying

Core work



http://skyserver.sdss.org/dr1/en/proj/advanced/galaxies/tuningfork.asp (Accessed 24 February 2014)

http://astrobites.org/wp-content/uploads/2012/01/zoo1.png (Accessed 24 February 2014)



Galaxy Ref:
588017991230685339

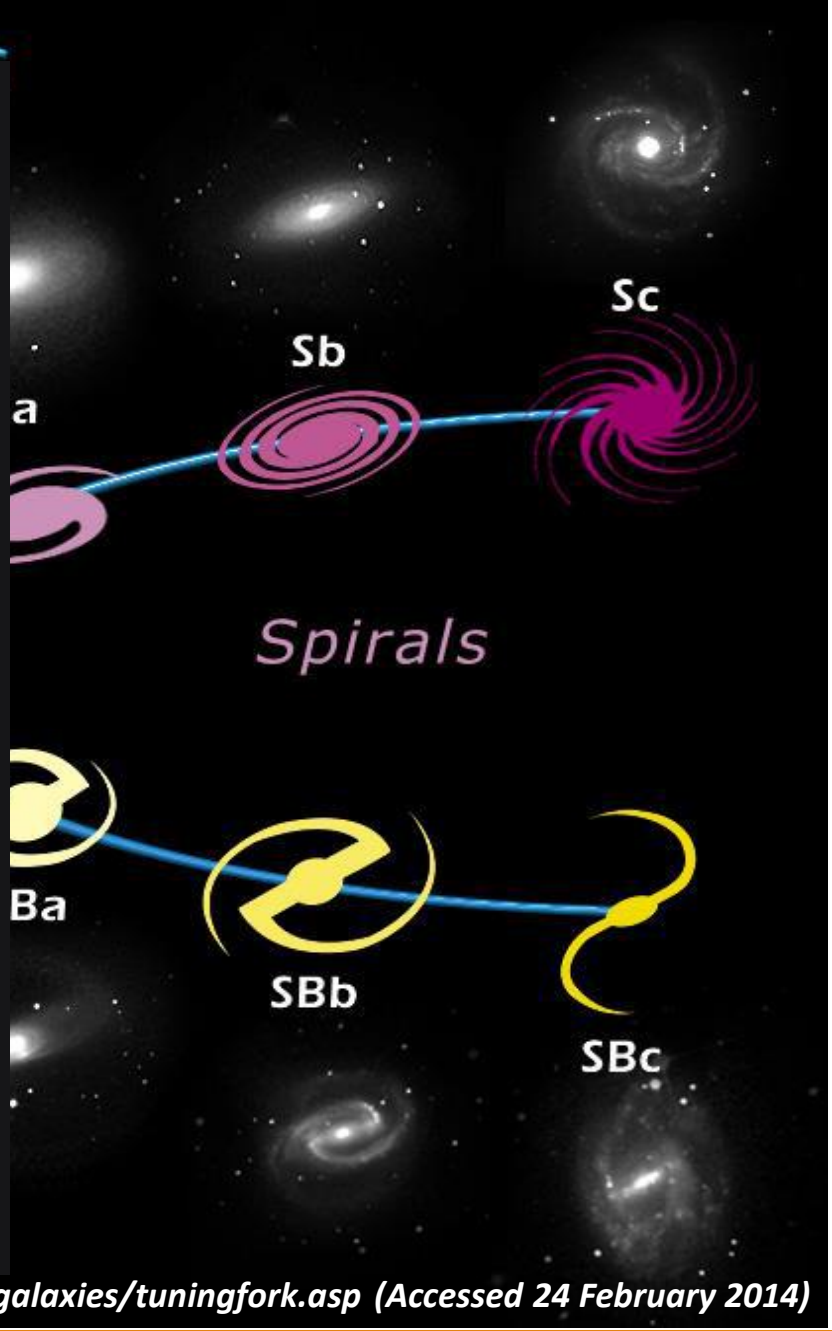
Choose the Galaxy Profile by clicking the buttons below

CLOCK | ANTI | EDGE ON/ UNCLEAR
SPIRAL GALAXY

ELLIPTICAL GALAXY

STAR / DON'T KNOW | MERGERS

Show Grid Overlay on the next Image



http://skyserver.sdss.org/dr1/en/proj/advanced/galaxies/tuningfork.asp (Accessed 24 February 2014)

Serendipitous discoveries

Hanny's Voorwerp

http://cas.sdss.org/astro/en/tools/chart/chart.asp?ra=145.2671505&dec=34.73290502 (Accessed 24 February 2014)

Serendipitous discoveries

Hanny's Voorwerp

http://cas.sdss.org/astro/en/tools/chart/chart.asp?ra=145.2671505&dec=34.73290502 (Accessed 24 February 2014)

Green Peas

Mon. Not. R. Astron. Soc. **399**, 1191–1205 (2009) doi:10.1111/j.1365-2966.2009.15383.x

Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies★

Carolin Cardamone,[1,2]† Kevin Schawinski,[2,3] Marc Sarzi,[4] Steven P. Bamford,[5]

http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2966.2009.15383.x/full (Accessed 24 February 2014)

Securing credibility

"There were people who have spent years of their lives learning how to classify galaxies…so they're sceptical that some supposedly random person could just click a few buttons"

GALAXY ZOO TEAM MEMBER

Volunteer motivations

Interest in astronomy

Volunteer motivations

Interest in astronomy

Awe at scale of universe

Volunteer motivations

Interest in astronomy

Awe at scale of universe

Desire to contribute to science

- Core work
- Serendipitous discoveries

Methods of crediting volunteers

Early league table

Top participants

http://climateapps2.oerc.ox.ac.uk/cpdnboinc/top_users.php (Accessed 24 February 2014)

Rank	Name	Recent average credit	Total credit	Country	Participant since
1	ibmcommerce	23,730	9,650,117	United States	16 Aug 2013 21:42:28 UTC
2	robert.mouris	21,671	7,715,682	Hungary	16 Jul 2005 3:53:16 UTC
3	happywetter.at	10,556	1,946,723	Austria	9 Sep 2012 13:42:52 UTC
4	[AF>WildWildWest] RLDF	8,240	4,333,251	France	3 May 2007 6:41:25 UTC
5	boinc_qc	7,382	95,157,395	United States	31 Aug 2004 19:33:12 UTC
6	oppen	7,115	12,459,487	Germany	25 Nov 2005 7:32:43 UTC
7	Eric in Seattle	7,112	502,178	United States	27 Sep 2010 23:44:43 UTC
8	Marius Schamschula	7,038	20,526,489	United States	1 Sep 2011 21:10:11 UTC

http://zoo1.galaxyzoo.org/poster/GalaxyZoo_lowres.jpg (Accessed 24 February 2014)

http://zoo1.galaxyzoo.org/poster/GalaxyZoo_lowres.jpg (Accessed 24 February 2014)

Motivations for egalitarian methods

Quality of individual classifications

"What we consciously don't want is a competition because we'd rather people classify well than classify as fast as they can"

GALAXY ZOO TEAM MEMBER

Motivations for egalitarian methods

Quality of individual classifications

"I was surprised by how many people wanted to cheat the system, there were just kind of trolls sending bogus clicks, writing robots"

GALAXY ZOO TEAM MEMBER

Motivations for egalitarian methods

Robustness of datasets

"There are people who have classified one million galaxies, but that's not where our data comes from, data comes from people who have done 100…The reason why these classifications are so good is because we have 70 independent classifications of every galaxy"

GALAXY ZOO TEAM MEMBER

Serendipitous discoveries

http://hannysvoorwerp.zooniverse.org/sandbox/Downloads/VoorwerpComic.pdf (Accessed 24 February 2014)

Serendipitous discoveries

http://hannysvoorwerp.zooniverse.org/sandbox/Downloads/VoorwerpComic.pdf (Accessed 24 February 2014)

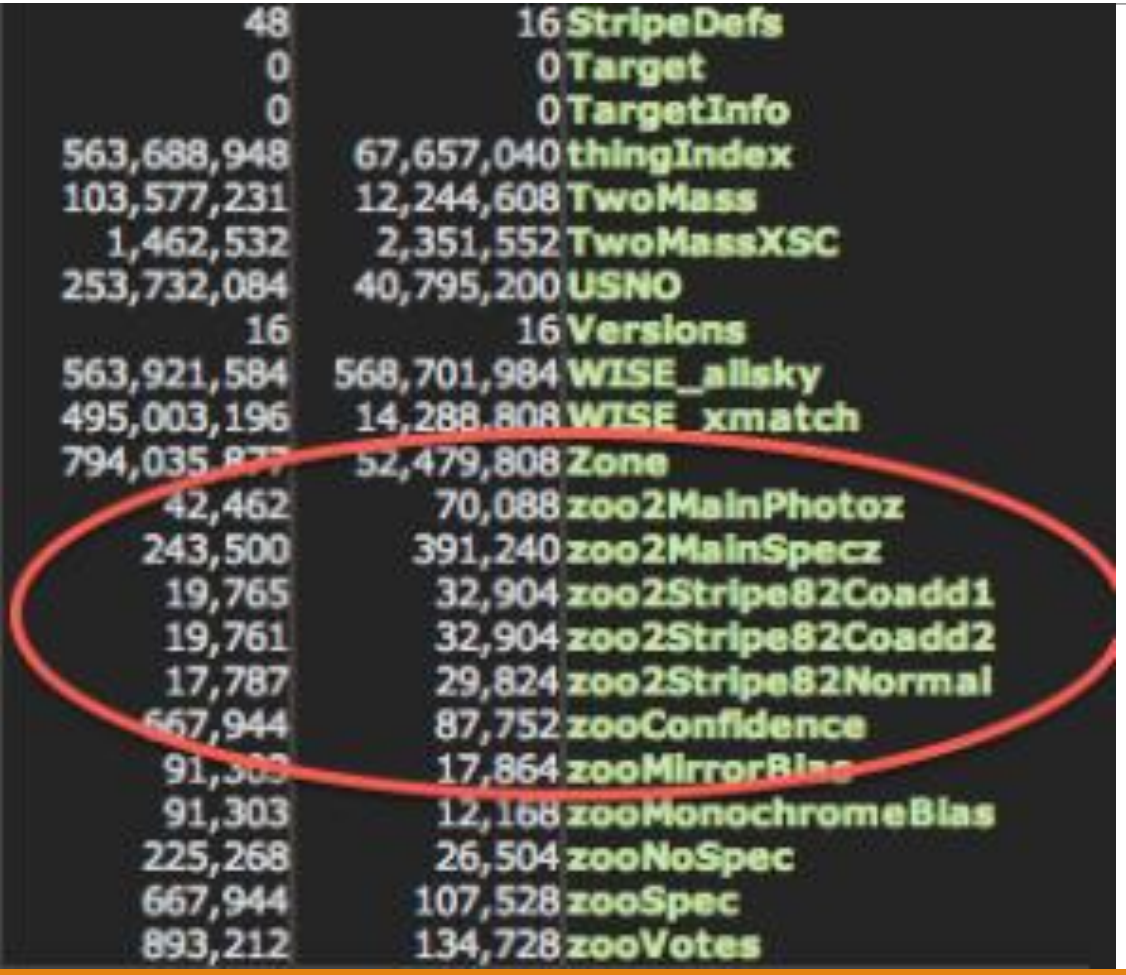
http://blog.galaxyzoo.org/2011/01/10/hannys-voorwerp-and-hubble-what-did-we-learn/ (Accessed 24 February 2014)

Success in securing credibility

"[Scepticism] has decreased…and some people have got excited about it so people are coming and saying let's do a Galaxy Zoo with the cosmos or other surveys"

GALAXY ZOO TEAM MEMBER

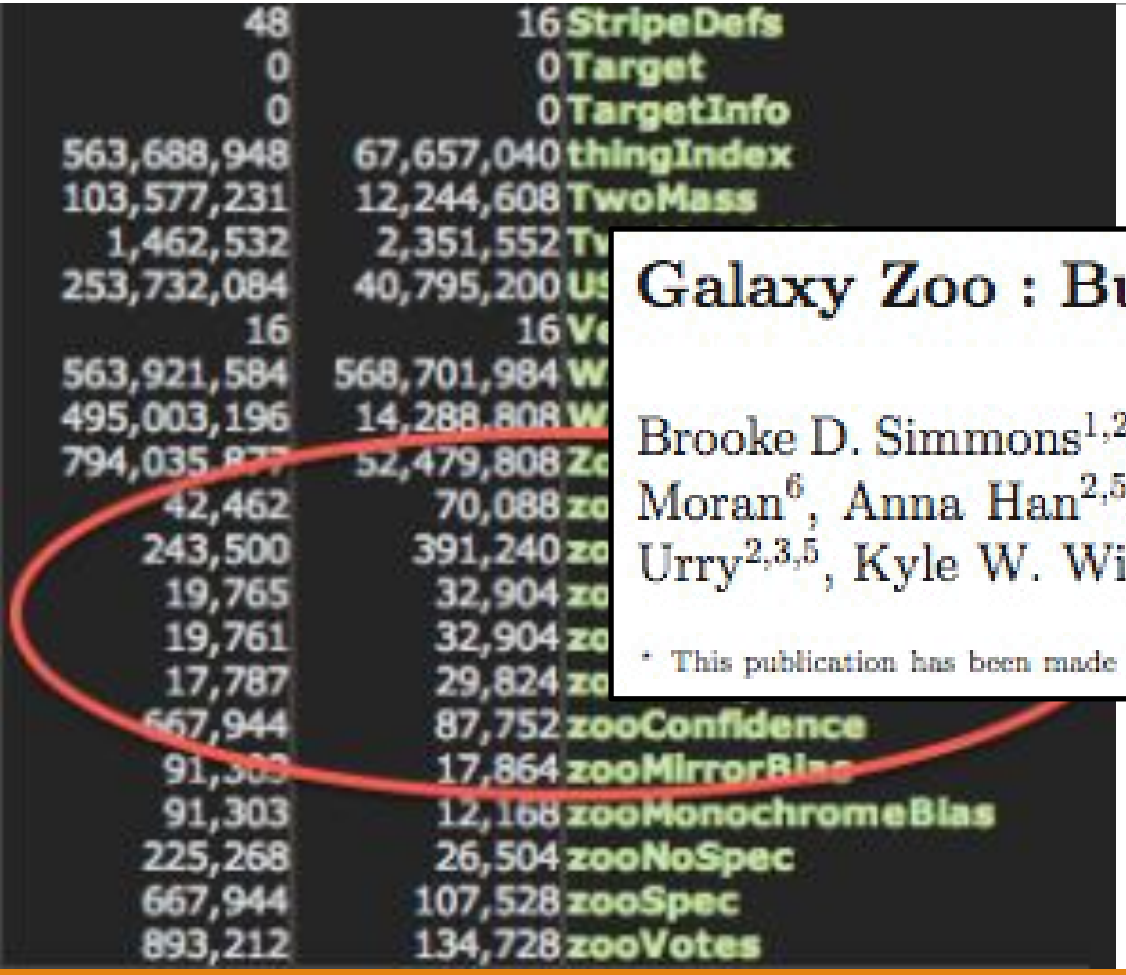
Success in securing credibility



48	16	StripeDefs
0	0	Target
0	0	TargetInfo
563,688,948	67,657,040	thingIndex
103,577,231	12,244,608	TwoMass
1,462,532	2,351,552	TwoMassXSC
253,732,084	40,795,200	USNO
16	16	Versions
563,921,584	568,701,984	WISE_allsky
495,003,196	14,288,808	WISE_xmatch
794,035,877	52,479,808	Zone
42,462	70,088	zoo2MainPhotoz
243,500	391,240	zoo2MainSpecz
19,765	32,904	zoo2Stripe82Coadd1
19,761	32,904	zoo2Stripe82Coadd2
17,787	29,824	zoo2Stripe82Normal
667,944	87,752	zooConfidence
91,303	17,864	zooMirrorBias
91,303	12,168	zooMonochromeBias
225,268	26,504	zooNoSpec
667,944	107,528	zooSpec
893,212	134,728	zooVotes

http://galaxyzooblog.files.wordpress.com/2013/07/screen-shot-2013-07-31-at-9-36-05-am.png?w=295 (Accessed 24 February 2014)

Success in securing credibility



48	16	StripeDefs
0	0	Target
0	0	TargetInfo
563,688,948	67,657,040	thingIndex
103,577,231	12,244,608	TwoMass
1,462,532	2,351,552	Tv
253,732,084	40,795,200	Us
16	16	Ve
563,921,584	568,701,984	W
495,003,196	14,288,808	W
794,035,8[illegible]	52,479,808	Zo
42,462	70,088	zo
243,500	391,240	zo
19,765	32,904	zo
19,761	32,904	zo
17,787	29,824	zo
667,944	87,752	zooConfidence
91,303	17,864	zooMirrorBias
91,303	12,168	zooMonochromeBias
225,268	26,504	zooNoSpec
667,944	107,528	zooSpec
893,212	134,728	zooVotes

http://galaxyzooblog.files.wordpress.com/2012/11/bulgeless_title_authors.gif?w=532&h=129 (Accessed 24 February 2014)

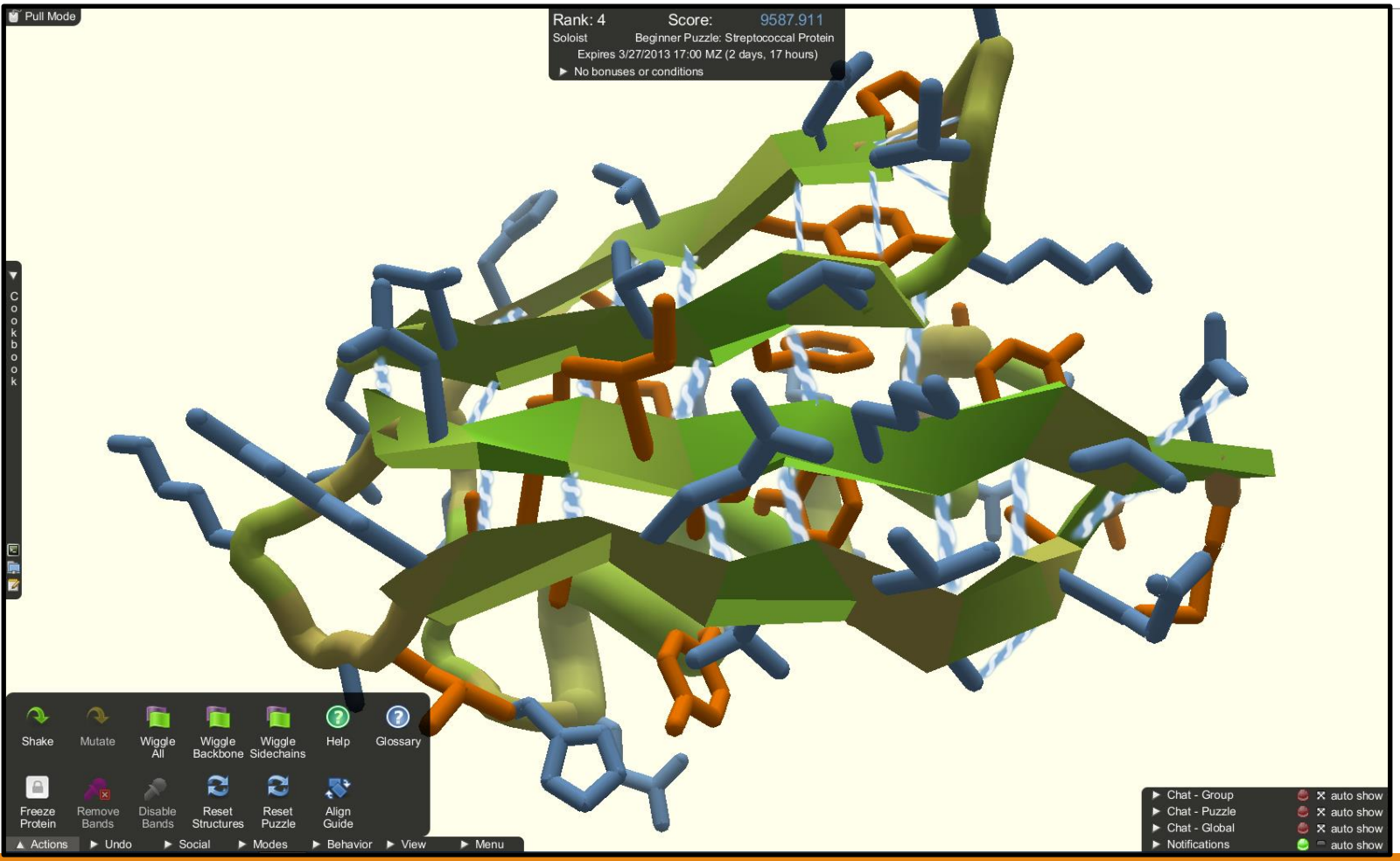
Galaxy Zoo : Bulgeless Galaxies with Growing Black Holes*

Brooke D. Simmons[1,2,3]†, Chris Lintott[1,4], Kevin Schawinski[2,5]‡, Edward C. Moran[6], Anna Han[2,5], Sugata Kaviraj[1,7], Karen L. Masters[8,9], C. Megan Urry[2,3,5], Kyle W. Willett[10], Steven P. Bamford[11], Robert C. Nichol[8,9]

* This publication has been made possible by the participation of more than 200,000 volunteers in the Galaxy Zoo project.

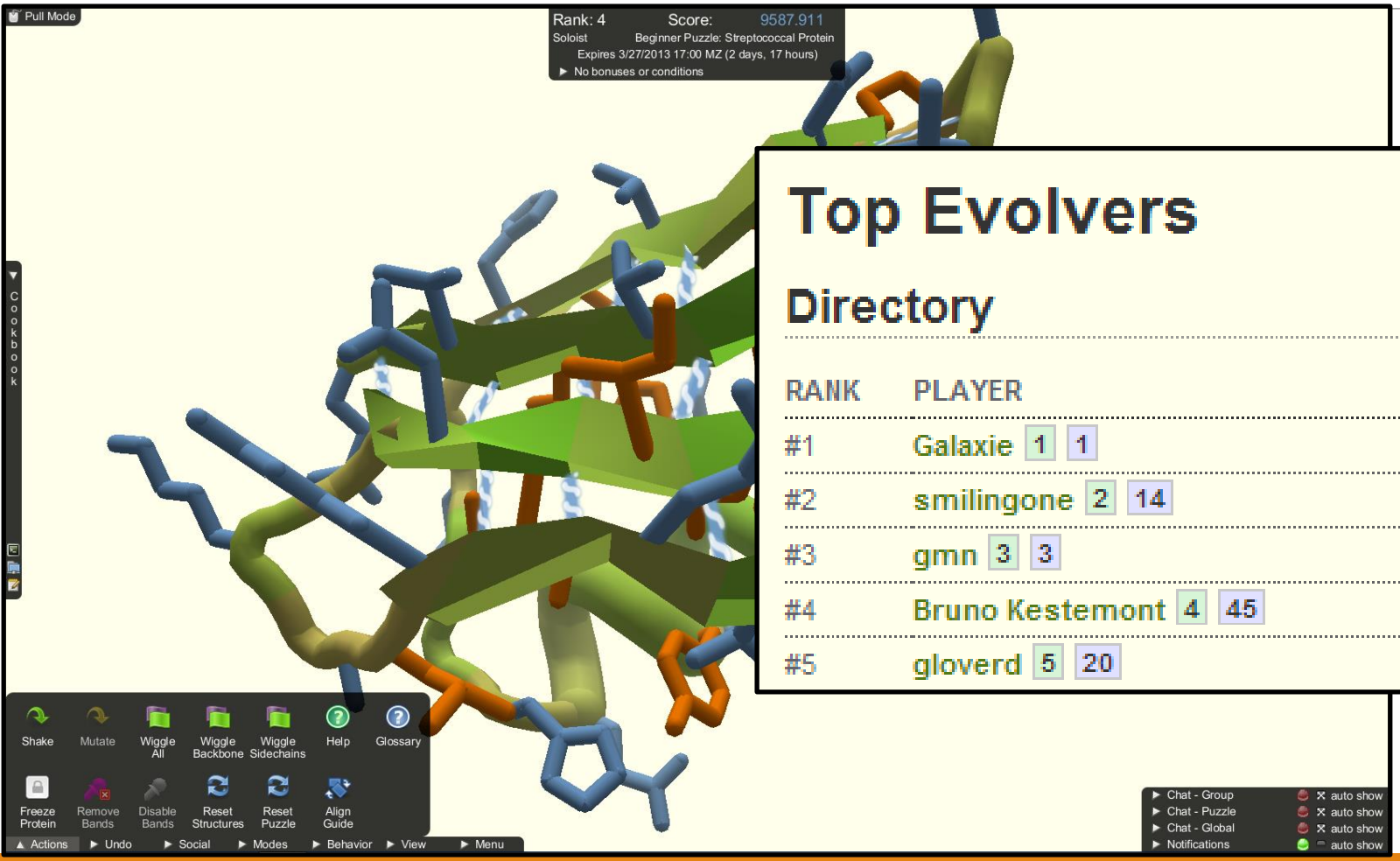
http://galaxyzooblog.files.wordpress.com/2013/07/screen-shot-2013-07-31-at-9-36-05-am.png?w=295 (Accessed 24 February 2014)

Implications for practice



http://fold.it/portal/files/newcompetition.png (Accessed 24 February 2014)

Implications for practice



http://fold.it/portal/players (Accessed 24 February 2014)

Top Evolvers

Find

Directory

RANK	PLAYER	GLOBAL EVOLVER SCORE
#1	Galaxie 1 1	4008
#2	smilingone 2 14	3601
#3	gmn 3 3	3575
#4	Bruno Kestemont 4 45	3537
#5	gloverd 5 20	3466

http://fold.it/portal/files/newcompetition.png (Accessed 24 February 2014)

Implications for practice

What is the nature of the task that volunteers perform?

Implications for practice

What is the nature of the task that volunteers perform?

Can the quality of volunteer contributions be scored according to well-established criteria?

Implications for practice

What is the nature of the task that volunteers perform?

Can the quality of volunteer contributions be scored according to well-established criteria?

What assumptions are made when volunteer contributions are combined into a dataset?

Implications for practice

What is the nature of the task that volunteers perform?

Can the quality of volunteer contributions be scored according to well-established criteria?

What assumptions are made when volunteer contributions are combined into a dataset?

What other scientific outputs might there be?

Other areas of study

Other ways to mediate relationships with volunteers

March 28, 2008

by The Zooniverse

in Papers, Site News

4 Comments

What happens next… Peer Review

http://blog.galaxyzoo.org/2008/03/28/what-happens-next-peer-review/ (Accessed 24 February 2014)

With the first Galaxy Zoo paper submitted (kudos to Kate and Anze!), we'd like to describe to you what happens next. What's scientific publishing all about? How does it work? If you've followed the blog and the forum, you have a pretty good idea of the first part of the scientific process: discovery!

We set out on the Galaxy Zoo project in part to test whether spiral galaxies

Other areas of study

Future developments in sky surveys

http://www.lsst.org/lsst/ (Accessed 24 February 2014)

Acknowledgements

Galaxy Zoo scientists and software engineers

Galaxy Zoo volunteers

Doctoral Supervisors: Dr Annamaria Carusi (University of Copenhagen) and Dr Marina Jirotka (University of Oxford)

Doctoral Examiners: Professor David de Roure (University of Oxford) and Professor Robin Williams (University of Edinburgh)

UCLA Knowledge Infrastructures Team: Professor Christine Borgman, Professor Sharon Traweek, Dr Jillian Wallis, Ashley Sands

Sloan Foundation: Award #20113194

And thank you for listening