

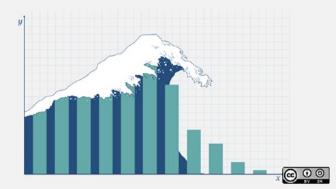
# CITIZEN Crowd sourced image classification

By Sunil Vanmullem

“Space is big. You just won't believe how vastly, hugely, mind- bogglingly big it is.”

[Douglas Adams](http://www.quotationspage.com/quotes/Douglas_Adams/), The Hitchhiker's Guide to the Galaxy

The amount of data that is being acquired and produced by space agencies is truly mind-boggling. NASA for example gathers hundreds of Terabytes of data per day[[1]](#footnote-1). This is a drop in the ocean compared to the volumes of data that future missions and experiments will produce. Big data is amongst the foremost issues that teams at NASA[[2]](#footnote-2) and ESA[[3]](#footnote-3) are working on.



Credit: Opensourceway https://flic.kr/p/9Lz176

The data tsunami that was seen in the mobile telecoms industry is nothing compared to the proportions that has been occurring in the space industry as instrument resolutions have been improving and technology continues to develop at astonishing rates. If 90% of the worlds’ data has been created in the last two years[[4]](#footnote-4), it is unimaginable to think about the Exabytes and beyond, volumes of data yet to come. But imagine we must, as there are opportunities for CGI as we have the experience and resources needed to help the agencies tackle this big space data problem.

There are a number of problems with big space data, each of which could justify papers in their own right.

* How the data is reliably acquired and stored. If the data is lost, it is lost forever. There is no going back to acquire the exact same data again.
* How the data is effectively disseminated to consumers of the data. The sheer volume of data means that downloading product data requires a consistently high bandwidth between the archive and the consumer.
* It also takes considerable resources to process such large amounts of data.
* Besides the volume of the data, how long does it take for the data to be analysed?

## The Data Analysis Backlog

Science takes a long time - to analyse the reams of data can take decades of detailed scientific work. This presents a problem with space missions as the spacecraft are making observations now, are unlikely to last decades and science targets need to be identified quickly. A data analysis backlog is building up.

I liken this to a thought experiment in which a group of scientists are making observations about the breeding of a particular species from the deck of a super tanker travelling at full speed. When a scientist spots a species he is interested in, he tells the captain to stop the ship so that he can make detailed observations. The problem is that by the time the supertanker stops, 5 miles away, the opportunity for the scientific observation is long past. In this thought experiment, the scientific process is the supertanker.



Credit: FactsHumor.com

The constraint is that most space instruments can’t stop, they have a limited life span and need to complete their mission objectives within a limited time. And having multiple missions with the same mission objectives is not an option due to the literally astronomical costs involved. So if it’s not feasible to change the way that the data is acquired, what about changing the way that the analysis is done?

## The Halls of Science

Science is done behind closed doors. Right? A minimum of a PhD and letters after your name is required to even get a chance to look at the data. Wrong. That may have been true before the internet was invented, but things (as they say) are a changing.

Funding for space agencies and their missions comes mainly from governments, but it’s public taxes that fund the government. So the actual source of the funding for space agencies is …. us, the tax payers. And the relationship between the public and governments has changed with members of the public now demanding more accountability for the money being spent in their name. So it is that agencies are being challenged to be more accountable to the people who fund them. In the US there is the Open Government Initiative[[5]](#footnote-5) while in the EU there is the Open government approach[[6]](#footnote-6). And it’s not just about funding, it’s a recognition that making data accessible is necessary to help keep ideas flowing and encourage innovation.

“Build it and they will come” and they did, the citizen crowd is coming with the data tsunami.

## Citizen Crowd

There is a nocturnal breed of people out there who have a casual interest in space imagery and who are actively looking at the images coming back from missions. I count myself amongst these strange people. We are not planetary scientists or even experts, but we are curious about the data coming back and hope to be amongst the first to see something that will influence scientific thinking about our planetary system and wanting to be involved in space exploration, be it from our armchairs.

There are many thousands of such space hobbyists, citizen scientists, out there who are dispersed into various disconnected forums and discussion boards, be it on Facebook, LinkedIn groups, NasaSpaceFlight.com, UnmannedSpaceflight.com, AboveTopSecret or hundreds of other alternatives.

Efforts continue to be made to engage citizen scientists with focussed crowdsourcing efforts such as planetFour.com[[7]](#footnote-7) where citizen scientists have classified millions of MRO images, NASA Image Detective[[8]](#footnote-8) that helps identify locations of photos taken by astronauts aboard the ISS, and NASA Disk Detective[[9]](#footnote-9) locating early planetary systems. And new crowd sourcing initiatives are emerging such as the initiative with planetary resources[[10]](#footnote-10) to detect, track and characterize near-Earth objects (NEOs). ESA is no stranger to this arena either with Invitations To Tender (ITTs) to investigate the use of crowd sourcing[[11]](#footnote-11) and having used the crowd to identify[[12]](#footnote-12) asteroid 2011 SF108 as being close enough to Earth to pose an impact threat.

I present here a prototype crowdsourcing platform focussed on MSL (Mars Science Lab – Curiosity) that I have developed in my own time.

## Why another platform?

I’ve been a fan of images from the Red Planet ever since Mars Global Surveyor (MGS) sent back intriguing other worldly images from Mars taken by the Mars Observation Camera (MOC) instrument. I downloaded every image to pore over them and start to understand the geology and the processes going on down on Mars. The simple truth is that it was an overwhelming task to analyse the 212,000 plus images. I think I managed 35,000 before this realisation dawned on me.

A problem I faced was how did I catalogue what I was seeing so that I could keep track of my analysis – let alone share it with others. It was superb that Malin Space Science Systems (MSSS) made the images available, but no collaborative tools were provided to be able to discuss findings with other users. There was an email address to feedback to MSSS, but responses to that feedback were never received because the agencies don’t have the structures to engage and respond to the thousands of citizen scientists like me who sent in questions and observations.

This lack of ability to respond to the crowd has not changed - the same problem of overwhelming the agency exists on social networks and other channels that the agency is using to try and engage the public. If the agencies truly want to engage the public, as mandated by the various open government initiatives, the relationship needs to change and the public needs to feel they are having a closer relationship with the agency.

Facebook and social media sites are useful places to share information, but as posts slide down the walls in minutes on busy sites the ability to retain information is poor. I needed a solution which was focussed on the space data and was usable by amateurs like myself in which we could easily share information and which would bring together multiple sources of data about the images. The problem was that it didn’t exist.

And then I came across an article[[13]](#footnote-13) by Chris Herwig which talked about an unpublished JSON feed for MSL. Then I started experimenting.

## A Peculiar Crowd and a Scientific Silence.

Before going on further I feel the need to address a concern about the crowd. A colleague recently told me that agencies have a policy about not touching wacky (I put it mildly) citizens even with a 100ft barge pole. The reason behind this is, as I mentioned, the lack of structures to engage the public, and the potential impact it has to distract the agency from its mission objectives. If the agency went to look at every rock about which some crazy person had an unscientific hypothesis that it wasn’t a rock they would not be able to focus on their mission.



Credit: http://cosasdekiko.blogspot.co.uk/2010/05/animanipulacion-12.html

I accept this. Though in my view an approach other than scientific silence is needed. Agencies need to recognise diversity, i.e.; the opportunity to educate the public. I think it’s important to recognise that this is something they need to get to grips with as they simply don’t have the resources to analyse everything. The agency needs the crowd because among the noise there might just be a lone signal that changes science forever.

The solution is simple – crowd sourced triage and moderation. Why not use the power of the crowd to perform a multi-level triage going through successive layers of agency friendly expertise so that agency resources and functions are not impacted?

# Curiosity Browser

<http://www.mars-browser.co.uk/curiosityb/>

The curiosity browser is a prototype web application that demonstrates that it is possible to cheaply construct a platform that allows the public to be engaged with space imagery by providing collaborative features to tag, highlight and comment on images and share these discoveries with others.

The prototype application took approximately 300 man hours to write, is written in PHP and JQuery and can be run on almost any computer or web hosting that can support Apache and PHP. The application is simple to deploy and run and does not use a database server and has a low CPU overhead.

The application is a continuing work in progress. At this stage the prototype shows the functionality of the browser and the ability to seamlessly combine multiple sources of data. Further functions continue to be added, and the application is being re-designed.

## Sols and instruments

The curiosity browser provides an easy way to locate images for specific sols (Martian days) and instruments. The Sols are shown in separated by divisions of 50 sols



Instruments relevant to the sol are retrieved asynchronously using JQuery and shown in the list



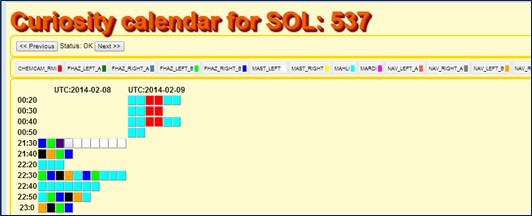
Selecting an instrument shows the images available for the sol/instrument combination



Any tags and highlights made by all users are shown in this view also



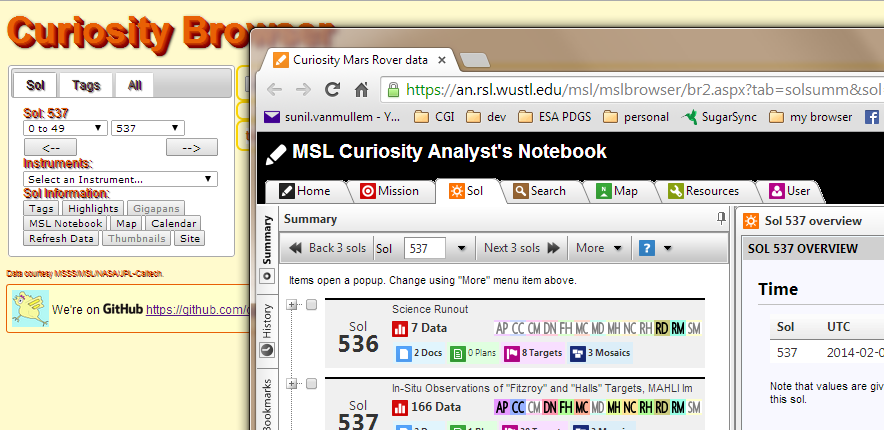
A click thru calendar is provided to see what other images are available in the sol and what instruments provided the image:



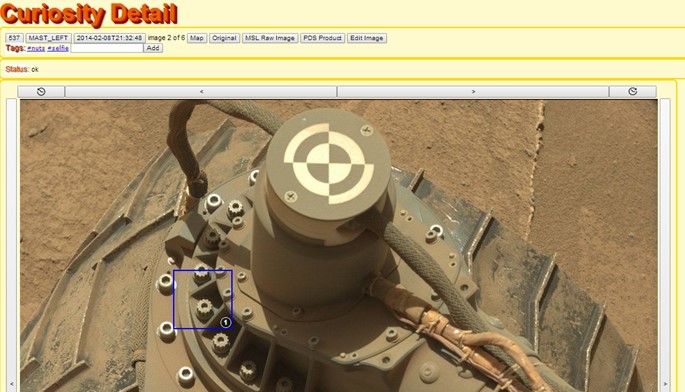
A visual album of tags and highlights is shown



Links are provided direct to NASA catalogues and other sources of information

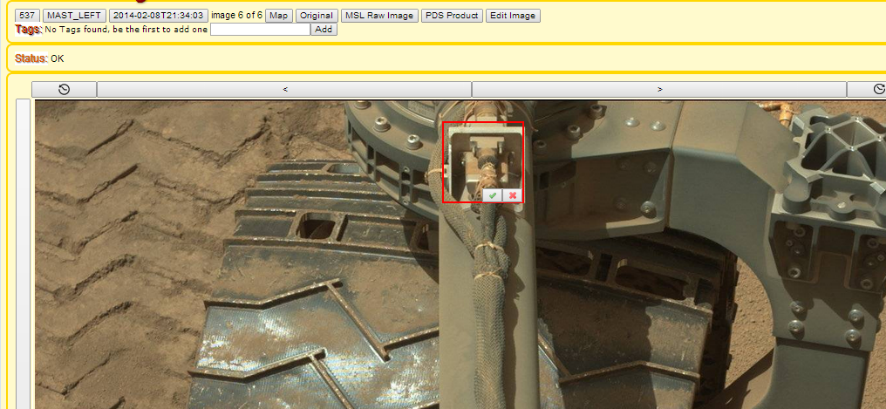


Click thru any image or calendar item to be taken through to the detail



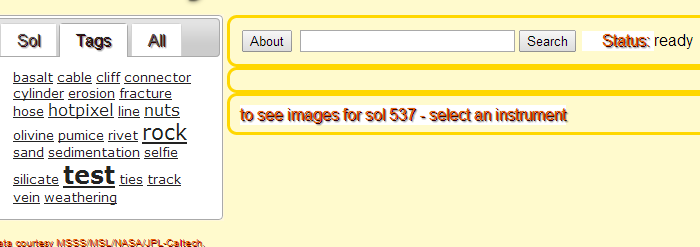
## Collaboration

The detail view allows highlights to be easily added by clicking on the area of the image of interest. A box is shown which is accepted by pressing the tick

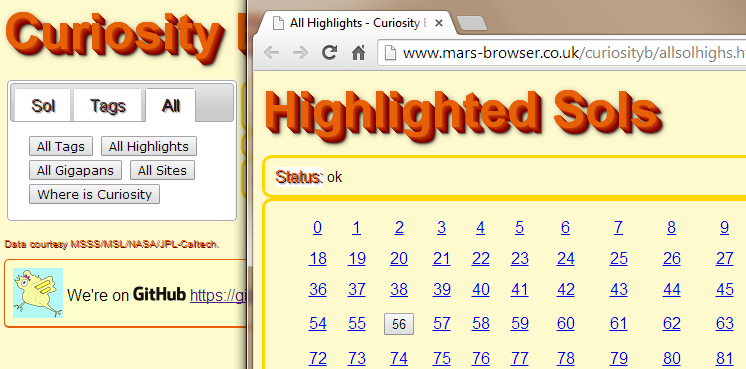


Tags and comments are added using the boxes provided.

A catalogue of tags and highlights is provided from the home page, the tags are shown as a tag cloud, with the more popular tags shown in a larger font.



A catalogue of all highlights is provided (sols shown as buttons have highlights associated)

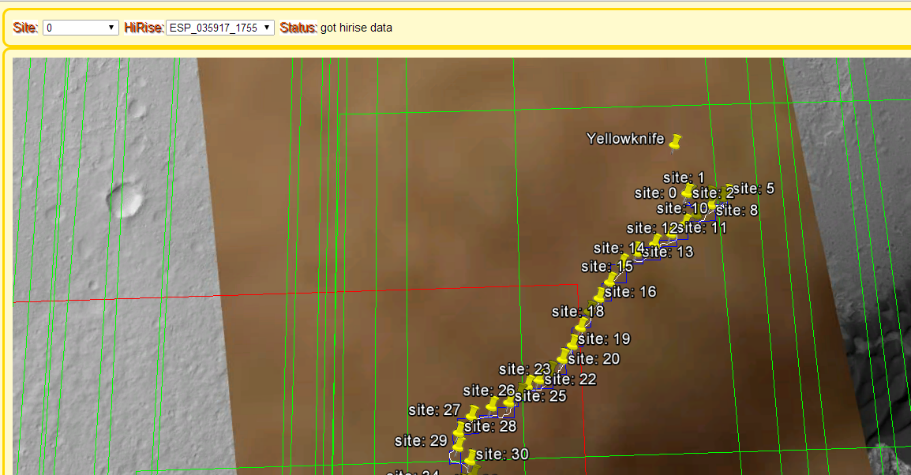


A catalogue of sols that have Gigapans (high resolution interactive panoramas created by stitching multiple images) associated to them is provided; these Gigapans were created by a fellow citizen scientist. Shown below is a progressive click thru.



## Google earth

I’m currently experimenting with Google Earth as I would love to be able to show the images captured in the context of a 3d environment. The results are encouraging; I can show the MSL traverse in the context of HiRISE data.



# The Future

What I’ve presented is a taster of crowd collaboration. The prototyping approach has allowed for rapid experimentation and to fail fast so that new ideas can be tried and built upon lessons learned from previous experiments.

The aim is to get buy-in from the agencies to provide tools, inspired by this experiment, that encourage engagement and collaboration with the crowd. I’ve been working with Geoff Buswell to engage with the NASA open government innovation initiative and am pleased to say that we’ve had a very positive initial response.

The browser continues to be developed with the following aims

* Aim to become a one stop solution for citizen scientists to collaborate on Pathfinder, Spirit, Opportunity, Curiosity and future rovers.
* Plan to add more interactivity – e.g. whiteboards, simple image processing, panorama generation, and voting, social networks.
* Plan to add citizen science missions to theme collaboration according to mission. e.g. identify Aeolian features
* Plan to link up to HiRISE, Mars Express and other Mars data sources
* Ask a Geologist – crowd triage - foster 2-way communication with experts, who can then escalate to the mission scientists.
* Industrialise - today is a fully working prototype used to develop and demonstrate the concept.

We’re also redesigning the site – here’s a glimpse:

## 

Copyright Tristam Sparks

## Get involved

If you have been inspired by this article and have an idea to develop, please just make a start and do something. Someone somewhere has to innovate and that someone should be you. When I started this project I didn’t know how to do JQuery, parse PDS data or integrate Google Earth. My advice is don’t let the unknown phase you – knowledge is like the Michael Buble song that goes “I just haven’t met you yet”.

If you are interested in helping out with the curiosity browser I would welcome your input. Please check out my subjects on Cynergi Conversations or drop [me](mailto:sunil.vanmullem@cgi.com) a note.

1. http://www.jpl.nasa.gov/news/news.php?release=2013-299 [↑](#footnote-ref-1)
2. http://www.informationweek.com/big-data/hardware-architectures/how-nasa-manages-big-data-/d/d-id/899791 [↑](#footnote-ref-2)
3. http://www.esa.int/Our\_Activities/Observing\_the\_Earth/Handling\_big\_data\_is\_no\_small\_feat [↑](#footnote-ref-3)
4. http://www.sciencedaily.com/releases/2013/05/130522085217.htm [↑](#footnote-ref-4)
5. http://www.whitehouse.gov/open [↑](#footnote-ref-5)
6. http://ec.europa.eu/digital-agenda/en/open-government [↑](#footnote-ref-6)
7. http://www.planetfour.org/ [↑](#footnote-ref-7)
8. http://eol.jsc.nasa.gov/sseop/ImageDetective/ [↑](#footnote-ref-8)
9. http://www.diskdetective.org/ [↑](#footnote-ref-9)
10. http://goo.gl/JLSiox [↑](#footnote-ref-10)
11. http://artes-apps.esa.int/open-data-and-crowdsourcing-itt-expression-interest [↑](#footnote-ref-11)
12. http://www.esa.int/Our\_Activities/Operations/Space\_Situational\_Awareness/Amateur\_skywatchers\_help\_space\_hazards\_team [↑](#footnote-ref-12)
13. https://www.mapbox.com/blog/tracking-mars-curiosity-rover/ [↑](#footnote-ref-13)