Grant agreement no. 607379

SPA.2013.2.1-01 - Analysis of Mars Multi-Resolution Images using
Auto-Coregistration, Data Mining and Crowd Source Techniques

- Collaborative project -

D7.4
Best Practice - Summarising ‘Mars in Motion’ Lessons

WP 7 – Crowd-sourced features for change discovery and validation of data mining

Due date of deliverable:	month 39 – March 2017	
Actual submission date:	31/03/2017	*(*) EC approval pending*
Start date of project:	January 1st 2014	Duration: 39 months

Lead beneficiary for this deliverable: UNOTT

Last editor: James Sprinks

Contributors: Robert J. Houghton, Jessica Wardlaw, James Sprinks (UNOTT)

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

History table

Version	Date	Released by	Comments
1.0	21/03/17	James Sprinks	Version 1.0

Executive Summary

This document summarises lessons learned from activities carried out within Work Package 7 "*Crowd-sourced features for change discovery and validation of data mining"* within the European FP7 project "*iMars: Analysis of Mars Multi-Resolution Images using Auto-Coregistration, Data Mining and Crowd Source Techniques"*. The purpose of this document is to share our reflections and observations regarding the design and implementation of a citizen science project (in our case using the Zooniverse Panoptes platform) and additionally to propose a design process for future projects. Our intent is that this will provide some stimulus to the future designers of citizen science projects and, in particular, promulgate a user-centric approach to the design of citizen science in general.

Table of contents

History table ... 2
Executive Summary ... 3
List of figures ... 4
List of tables ... 4
Key word list ... 5
Definitions and acronyms ... 5
1. Introduction ... 6
  1.1. Scope and objectives ... 6
  1.2. Relationship to activities in WP7 and relationship to other Work Packages ... 6
2. A design process for virtual citizen science projects ... 6
  2.1. A framework for citizen science project design ... 7
  2.2. Citizen science project development process ... 9
3. Data outcomes and volunteer performance ... 14
  3.1. Current data outcomes ... 14
4. Best practice recommendations ... 16
  4.1. The influence of task and interface design on performance ... 17
  4.2. Design guidelines for a planetary citizen science system ... 18
References ... 19

List of figures

Figure 1: An ergonomics systems model for citizen science. ... 8
Figure 2: A citizen science project development process in nine stages. ... 13
Figure 3: New crater impact marked by Mars in Motion volunteers. ... 15
Figure 4: Slope streak marked by Mars in Motion volunteers. ... 15
Figure 5: Dust devil track marked by Mars in Motion volunteers. ... 16
Figure 6: Crater impact and ejecta marked by Mars in Motion volunteers. ... 16

List of tables

Table 1: Features of change detected by volunteers using the Mars in Motion platform. ... 14

Key word list

Citizen Science, Crowdsourcing, Change Detection, Task Design, ICT, Internet Applications, Visual Inspection

Definitions and acronyms

Acronyms	**Definitions**
MSSL	Mullard Space Science Laboratory
TWD	Task Workflow Design

1. Introduction

In this deliverable of the iMars project - *D7.4 Best Practice - Summarising 'Mars in Motion' Lessons* we share our observations and reflections on the development of a citizen science project, "Mars in Motion". The aim of this activity is to provide a set of insights and food for thought to the designers of future citizen science projects based upon our experiences and also to share our process and model for designing citizen science projects.

1.1. Scope and objectives

As discussed in D7.1, while there are some noteworthy exceptions, generally speaking most citizen science projects typically devote more energy to reporting final results than explaining their own basis and implementation - consequently while practice can clearly be seen as developing, it is hard to find guidance that addresses many issues beyond the purely technical. Therefore, this deliverable is intended to address that lacuna within the literature and to be complementary to prior deliverables within Work Package 7 which more formally report the results of a literature review (D7.1); present open source software (D7.2); and present results from an experimental series (D7.3).

It specifically comprises of:

1. A description of our high-level human factor/ergonomics inspired model of citizen science as a designable system of interrelated parts.
2. A description of our design process for "Mars in Motion" that has emerged from WP7 as a whole that we propose for use in future projects.
3. A discussion of data outcomes and volunteer performance.
4. Our final best practice recommendations based on all of the above.

1.2. Relationship to activities in WP7 and relationship to other Work Packages

Within Work Package 7, as mentioned above, this deliverable constitutes a capstone to prior work developed through literature review (*D7.1 Design guidelines for crowd-sourcing software*), software development (*D7.2 Improved toolkits*) and laboratory experimentation and further testing (*D7.3 Report on Data Validation Tests by Citizen Scientific Users)*. The work is also informed by our collaboration across Work Packages, particularly with Work Package 6 (*Change detection from Data mining & Validation)* and Work Package 8 (*Dissemination*) which provided opportunities for both testing and engagement with the needs of different stakeholders around a crowd sourcing activity formative to this deliverable.

2. A design process for virtual citizen science projects

As part of the FP7 iMars project, and together with our experience in implementing a prior citizen science project "Planet Four: Craters", we identified a user-centric design process. This process was informed throughout by reference to a "systems ergonomics" approach that considers the role of people from multiple perspectives (c.f., Houghton, Balfe & Wilson, 2016; Wilson, 2014) developed at the start of the Work Package. We recap on this now.

2.1. A framework for citizen science project design

Key to our conceptualization of citizen science is that it necessitates the creation of a system of work (Houghton et al., 2016). By work, we do not mean necessarily paid or obligatory activity – nearly all citizen science is based around volunteers after all -- rather a structured goal-directed endeavour. That this constitutes a wider system beyond a web interface or an app is not necessarily appreciated but is clear from the fact that citizen science is a form of mass work around a scientific endeavour. A project typically includes a large group of workers (however brief their involvement), the issuing of tasks (be it classifications in a Zooniverse project or data collection through an app) and the recovery and judgement of the product of those tasks. Layered onto this will typically be some form of data aggregation as well as social interactions between workers (e.g., a web forum) and often a group of scientists who provide some form of overarching management and governance. That this activity might be, at least in part if not whole, mediated or created implicitly through webpages, mobile phone apps, servers and databases does not mean these elements are not present or that their manifestation cannot be altered or redesigned.

There are many ways in which a conceptual system model can be defined and represented but in the present case we settled on a variation of the so-called “Onion Model” (Sharples & Wilson, 2015, p. 11) customized to the citizen science case (Figure 1). The goal of adopting this model is to provoke consideration of how design decisions at different level relate and constrain the overall form of the system and can all affect two key variables: motivational impact on users and the quality of the judgements generated.

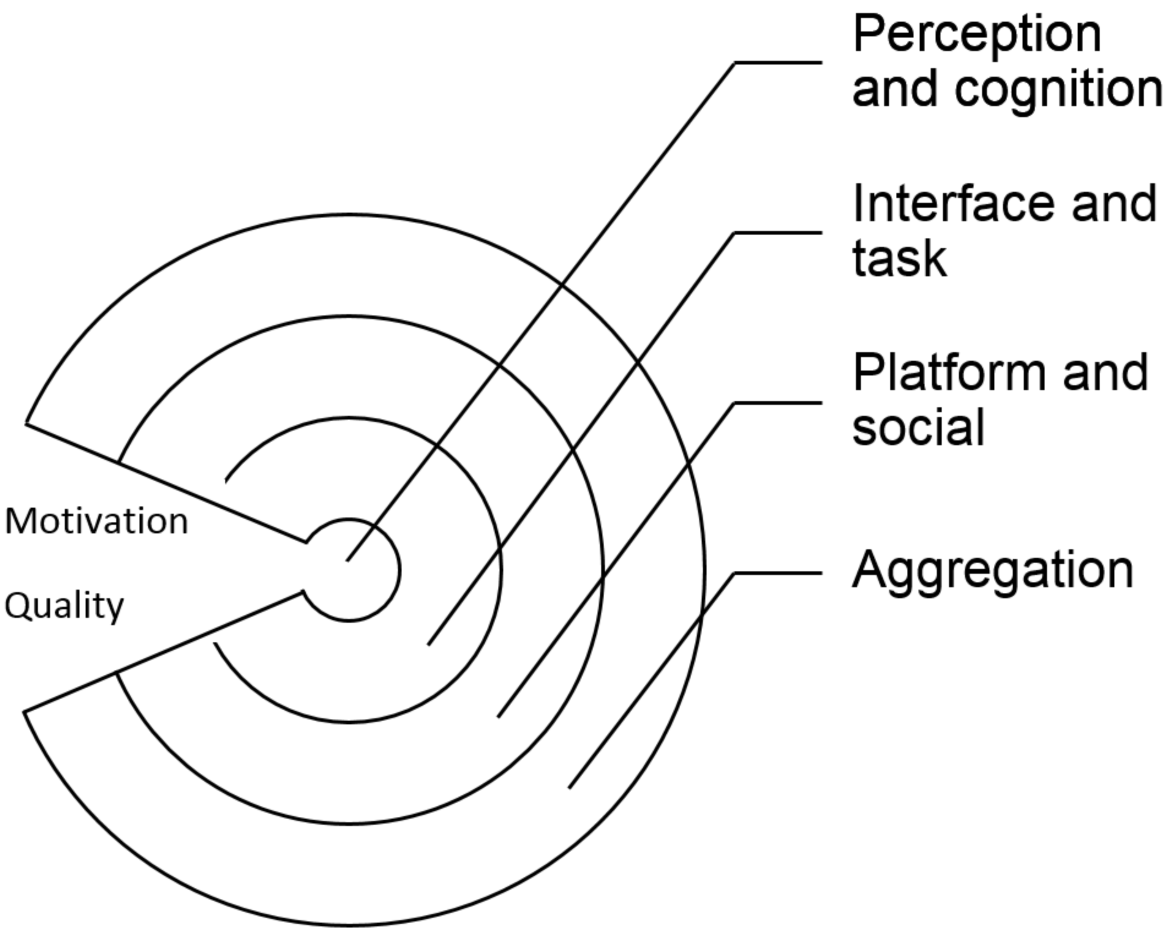


Figure 1: An ergonomics systems model for citizen science.

To illustrate these concepts, consider the engaged citizen scientist sitting in front of her/his computer undertaking a crater counting activity. At the lowest level of analysis, she/he is asked to make some sort of judgment or decision about the imagery presented. The issues here concern human perceptual psychology with regard to the imagery presented and the parameters under which good decisions and judgments can be made. This imagery is of course presented using some sort of user interface; an optimal user interface will allow easy interaction with the imagery and easy reporting of judgments by the citizen scientist. Taking a step further back in terms of abstraction, the sum of activities the citizen scientist is undertaking will add up to some sort of 'work design'; at this point we might begin to wonder what the optimal pattern of workflow would be to keep the citizen scientist engaged, which tasks should be undertaken by people, which tasks should be allocated to algorithms and how can people and algorithms work together? At the highest level we need to consider how the many citizen scientists are organised and managed; what steps are taken to check or balance performance, what opportunities are there for community interaction and how can we aggregate responses into accurate answers to the questions we are most interested in? In truth all these elements are intimately related and affect each other; a difficult task may leave a citizen scientist feeling demotivated, the user interface will be configured to match the job design, the level of explicit organisation will depend upon the kinds of activities being undertaken and so on. What this model does capture is the fact that citizen science projects are eminently *designable* (hence in principle tuneable and optimizable) and contain numerous design decisions that must be deliberately made even if in a particular case they may emerge *en passant* under relatively little reflection.

2.2. Citizen science project development process

The phased development process used to develop “Mars in Motion”, a citizen science project designed to engage participants with change detection and feature identification on the surface of Mars is shown in Figure 2. The project was specifically designed to be expressive of the unique iMars FP7 imagery developed in WP2, 3, 4 and 6, which by bringing together and co-registering imagery from 40 years of Mars exploration offers unique opportunities. Whilst on the one hand we could have used this imagery for relatively straightforward tasks perhaps less evocative of that presented to specialists by the iMars webGIS (WP5), we were particularly concerned, first, that the engagement of citizen scientists should reflect the core science case in the project (discovery of changes on the surface of Mars) and second that change detection was manifest in the project as we felt ‘discovery’ was a potentially exciting and satisfying form of activity.

The process outlined has an obvious familial similarity to many used in work system and computer system development (see Houghton, Balfe & Wilson, 2015 for a review), with amendments made to reflect the particular needs and constraints manifest in citizen science work.

The first stage is to acquire and understand the underlying science case. For our purposes here, the science case can be understand as broadly analogous to role a business case plays in the development in a new enterprise as it defines the reason for the existence of the system (a system designed purely for educational or engagement purposes without any scientific aims could possibly substitute this stage with normal user requirements or pedagogical requirements). Our goal throughout the next four stages is essentially to reconcile this science case with the capabilities and needs of human participants. Ensuring this ‘fit’ is essential to producing a productive and sustainable project that creates data of an adequate quality (i.e., with reference to the systems model above - that manages both performance and motivation). This is a concrete expression of our opposition to views of crowdsourcing and human computation that consider the participants as if they were “human processing units” or “clickworkers” and similar popular but arguably dehumanising notions (see Reeves & Sherwood, 2010, and Kittur et al., 2013 for a wider discussion).

The second stage considers engagement with the imagery or data that participants need to be interactive with. At this stage, relatively early in the process, it may still be possible to interact with the producers of that imagery or to consider pre-processing or re-representing if it appears particularly problematic for human viewing. Either way, understanding of this is necessary to the following steps.

The third stage represents the most complex element of this reconciliation between science case and the citizen user's capabilities, tendencies and needs. We cannot offer generic advice that would suit every image set and data requirement but we can outline the kinds of things that might be worth considering and the kinds of conclusions we drew in the present project. First, we used a systems ergonomics model to consider the work from four different but closely interlinked perspectives (see above section). In terms of 'perceptual', a key issue for us was in terms of the expected form and detectability of different types of features (which we assessed with reference to the literature and formative outcomes from WP6) - these outcomes are reported D7.3. The difference in approach represents a certain amount of hedging in so far as we are designing for imagery as yet unseen, and indeed, by one person, unseeable hence the need for crowd sourcing! Nonetheless, we were able to use the literature to come to an assessment of the parameters involved and to operationalise this down to the likely size and extent of these features in the imagery available to us. This then informed the extent to which it was realistic to think about a task in which multiple forms of change detection and feature recognition were present. An issue of further consideration under the heading of 'task and interface' was the best way to help participants carry out a change detection task (feature recognition and reporting were less controversial and lent themselves to radio buttons and a bounding box respectively). Under the heading of "platform & social", while it was not obvious that our work would require any particularly novel social interaction, a question to be asked regards whether one envisages the activity as being a 'one shot' situation where the site is only visited once or whether it is a good idea to build in social features (such as discussion boards, imagery collections and commenting on individual images) to encourage a so-called participant career (Luczak-Roesch et al., 2014; Cox et al., 2015). Our determination was that the task was of sufficient scope and variety that a participant career model was ideally to be encouraged. Finally, under the heading 'aggregation', it is worth considering what type of aggregation scheme is required. In our case this, and all other stages, were further informed by our position on automation - which is to say in this case image-pairs were to be for the most part pre-selected by algorithmic means (WP6 of the iMars project).

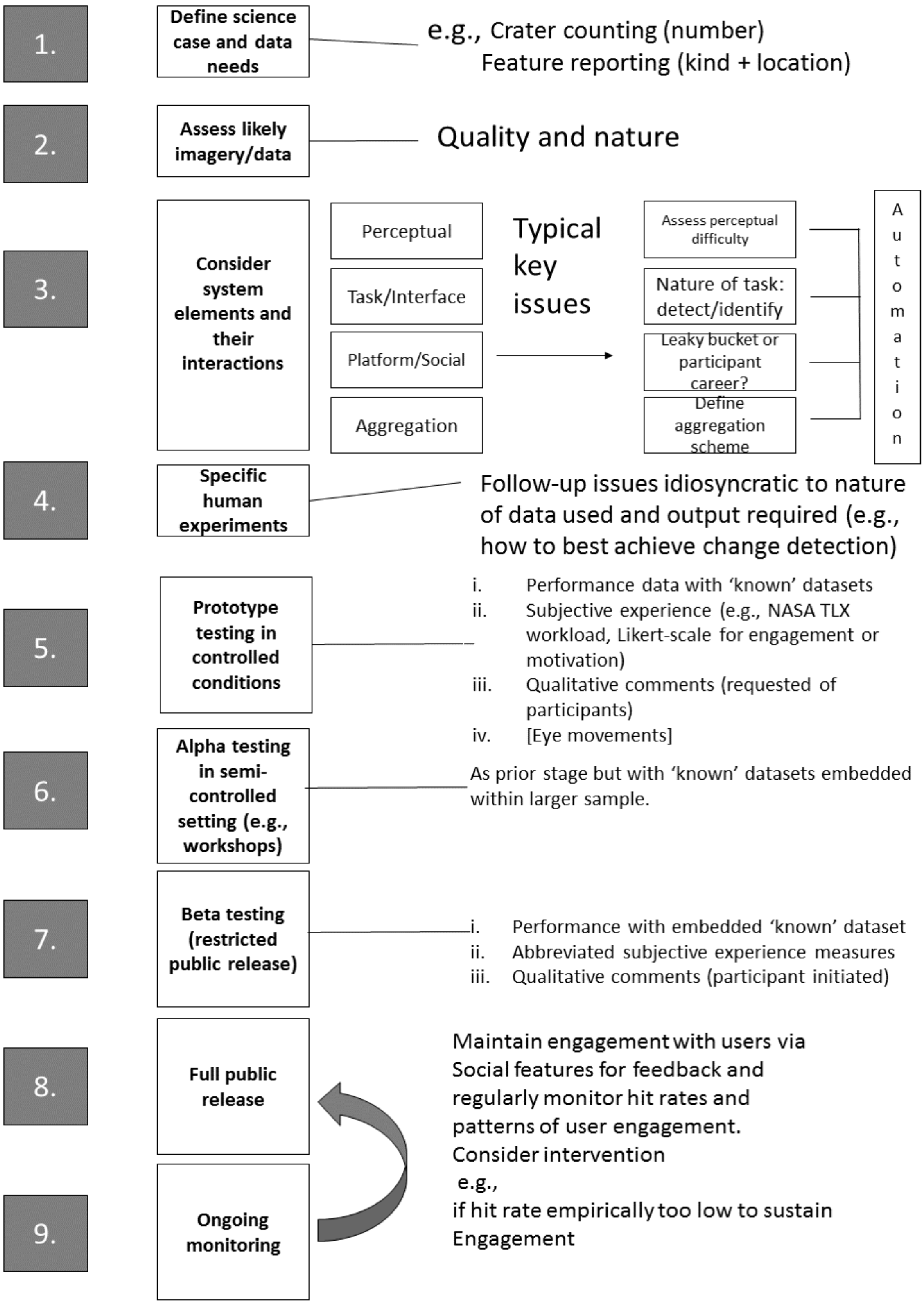
What should be clear even from this short discussion is that issues under all these headings are mutually interacting and constraining. Once identified, and following consultation with the relevant literature, we arrived at a set of key questions that were then addressed in the fourth stage with specific human experiments (i.e., form of task interface - flicker or side-by-side; how many features should we ask humans to recognise and how does this affect motivation and performance; how do humans relate to the statistical properties of image-set shaped by algorithmic pre-selection). Armed with the outcomes of these experiments we were in a position to build a prototype.

In the fifth stage, we tested a prototype under controlled laboratory conditions. We used imagery where ground truth had already been established so it was possible to measure performance. The aim here is to collect as much data as possible from humans. We explored perceptions of workload using different prototypes using standard tools (including the NASA Task Load Index or TLX) together with subjective ratings of motivation and detailed interviews to elicit qualitative feedback, concerns and so on. This phase of pre-testing led to particularly useful feedback that could be fed back into further prototyping activity. Because we could control and observe how users interacted with the system, as well as interview them in person, while the absolute numerosity of the participants was relatively low (¬30 people), the quality of the data was high and its content rich and provocative at a point where the design was not yet fixed in any way. Although it was not possible in our laboratory owing to resources, it also struck us that eye-tracking measures might be a particularly useful additional measure to gain insight not only into what participants did, but as an objective source of clues about *how* they might be doing it. Other indirect measures such as heart rate, skin conductivity, keystroke analysis and pointer tracking might also be mooted depending on the needs of the situation.

In the sixth stage, having produced a prototype that we and our participants were reasonably happy with, we undertook an ‘Alpha test’ with a workshop audience. The represents a middle ground between the laboratory experiment and a public release. The environment is slightly less controlled (and may elicit behaviours would not produced when closely observed and perhaps embolden participants towards freer discourse around limitations of the system) yet it is still possible to exert influence over participants, particularly to justify the collection of subjective and qualitative measures. This also offers a small initial test of social features (which themselves must work) and, of course, a staged increase in the amount of load and simultaneous users to provide a test of technical elements of the system. It is usually also possible at this point to increase the size of imagery set but to embed the smaller ‘known’ imagery subset within it to maintain a measure of performance within this.

In the seventh stage, a limited public release for testing purpose is carried out. In our case this was facilitated by the Zooniverse project on our behalf but could also be created through light seeding of publicity materials in a particular context (e.g., within a university). At this stage we are interested in looking at behaviour and performance ‘in the wild’ and it becomes necessary, as there has been an increase in scale and scope, to restrict direct participant data capture (this is now harder to carry out anyway as the relationship between experimenter and citizen participant is largely mediated by a web interface and it is harder to persuade compliance). We found it was possible to embed some subjective measurement tools but the experimenter should be aware compliance rates are likely to be lower and from self-selecting individuals. In order to encourage maximum compliance, the use of abbreviated short-form instruments is advisable. It is also possible at this stage to still receive qualitative comments but these will be in written (and often ‘complaint’) form and will be user initiated rather than as a result of requests.

Once running, the eight stage, it is still necessary to monitor the running of the system. At this point this is likely to be relatively indirect. For example, the scientists involved in the project should maintain a presence within chat and discussion forms to both gain a feeling for how the project is perceived, as well as to answer citizen participant questions. This monitoring will also take a quantitative form (rates of completion, Google analytics data et.). An example of a quantitative measure of particular interest to iMars and Mars in Motion is that of empirical hit rate; a variable identified through earlier stages and experimentation as particularly important to motivation and the production of persistent engagement within a change detection task. In the event this fell too low (despite our projections based on algorithm performance that this should be found at a healthy 1:5 change/no change ratio) we should be prepared to consider what interventions could be made either in terms of amended design or imagery pre-filtering to avoid a task that feels like ‘hunting for a needle in a haystack’ that may repel users. In such a case it may be necessary to reiterate back to an earlier stage of development.



Follow-up issues idiosyncratic to nature of data used and output required (e.g., how to best achieve change detection)

i. Performance data with ‘known’ datasets
ii. Subjective experience (e.g., NASA TLX workload, Likert-scale for engagement or motivation)
iii. Qualitative comments (requested of participants)
iv. [Eye movements]

As prior stage but with ‘known’ datasets embedded within larger sample.

i. Performance with embedded ‘known’ dataset
ii. Abbreviated subjective experience measures
iii. Qualitative comments (participant initiated)

Maintain engagement with users via Social features for feedback and regularly monitor hit rates and patterns of user engagement. Consider intervention e.g., if hit rate empirically too low to sustain Engagement

Figure 2: A citizen science project development process in nine stages.

3. Data outcomes and volunteer performance

The current iteration of 'Mars in Motion' went live on the Zooniverse platform in June 2016 - in conjunction with the Europlanet RPIF 3D Training Workshop that took place at the Mullard Space Science Laboratory (MSSL). It was launched with an image set of 868 image pairs (2 images of the same region of Mars taken at different times) all of which demonstrated change derived from an initial change-detection algorithm developed at MSSL as part of Work Package 6 (*Change detection from data mining & validation*).

3.1. Current data outcomes

As of March 2017, over 9,000 classifications have been made using the Mars in Motion platform, looking for geomorphological changes on the surface of Mars. Table 1 below summarises these classifications in terms of the types of change identified by volunteers, and their level of agreement.

Table 1: Features of change detected by volunteers using the Mars in Motion platform.

Feature Type	**Number Detected**	**Mean Volunteer Agreement**	**Median Volunteer Agreement**	**Average Number of Volunteers**
Dune	105	0.77	0.77	2.07
Dust Devil Track	170	0.84	0.84	2.88
Gully Slide	65	0.53	0.52	2.77
Impact Crater	83	0.63	0.63	2.27
Seasonal Fan	32	0.50	0.50	2.00
Slope Streak / RSL	154	0.75	0.75	2.59

Through these volunteer classifications a number of these changes have been verified by expert comparison, providing some good examples of surface activity on Mars. Figures 3 - 6 show some of these examples, along with the average marking position of the volunteers who detected the change (the yellow box).

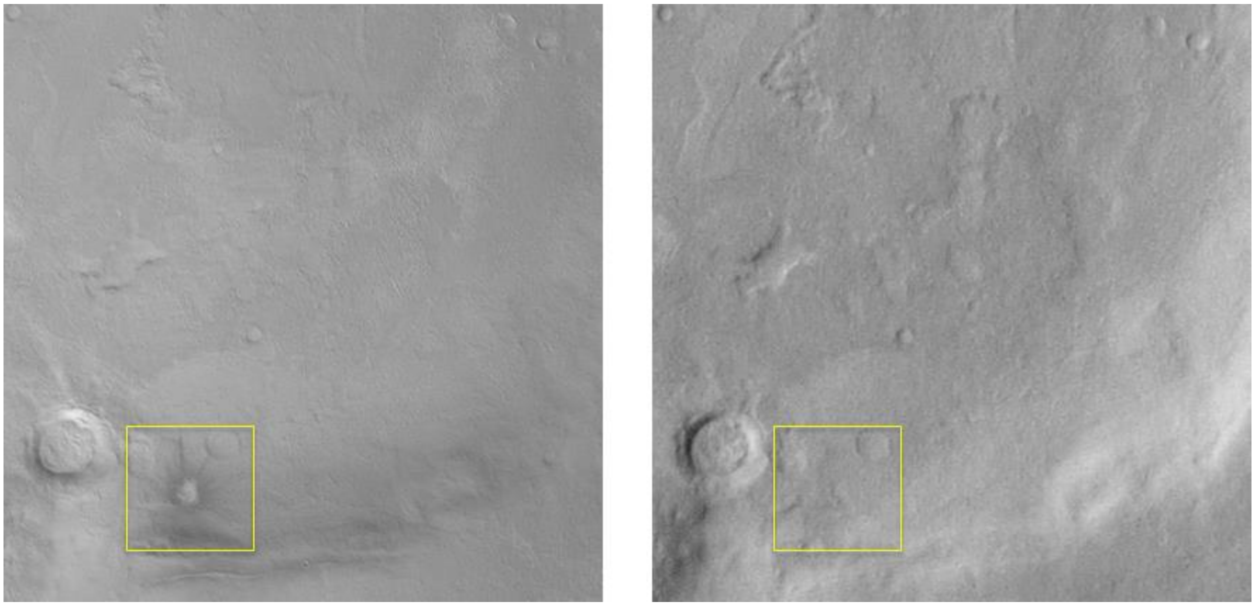


Figure 3: New crater impact marked by Mars in Motion volunteers.

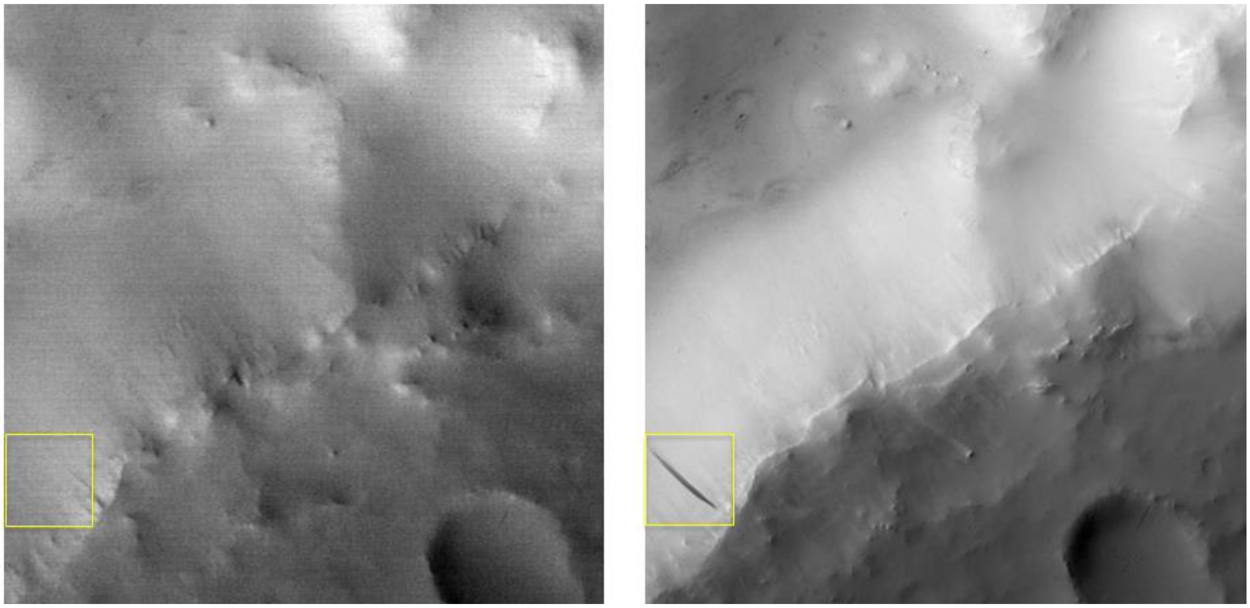


Figure 4: Slope streak marked by Mars in Motion volunteers.

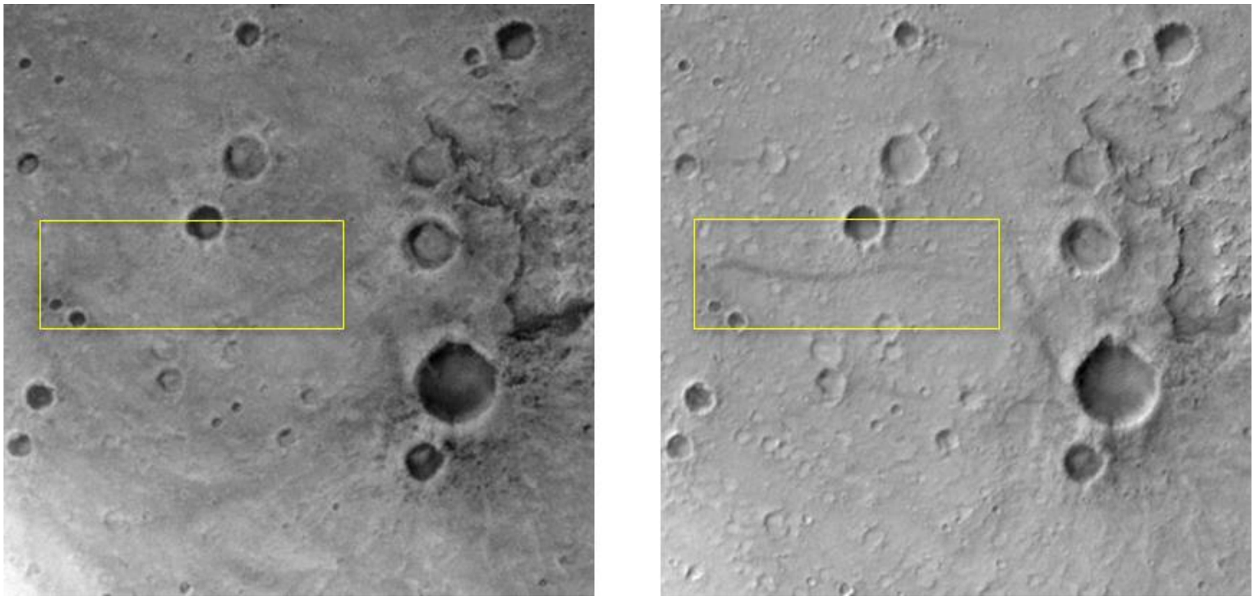


Figure 5: Dust devil track marked by Mars in Motion volunteers.

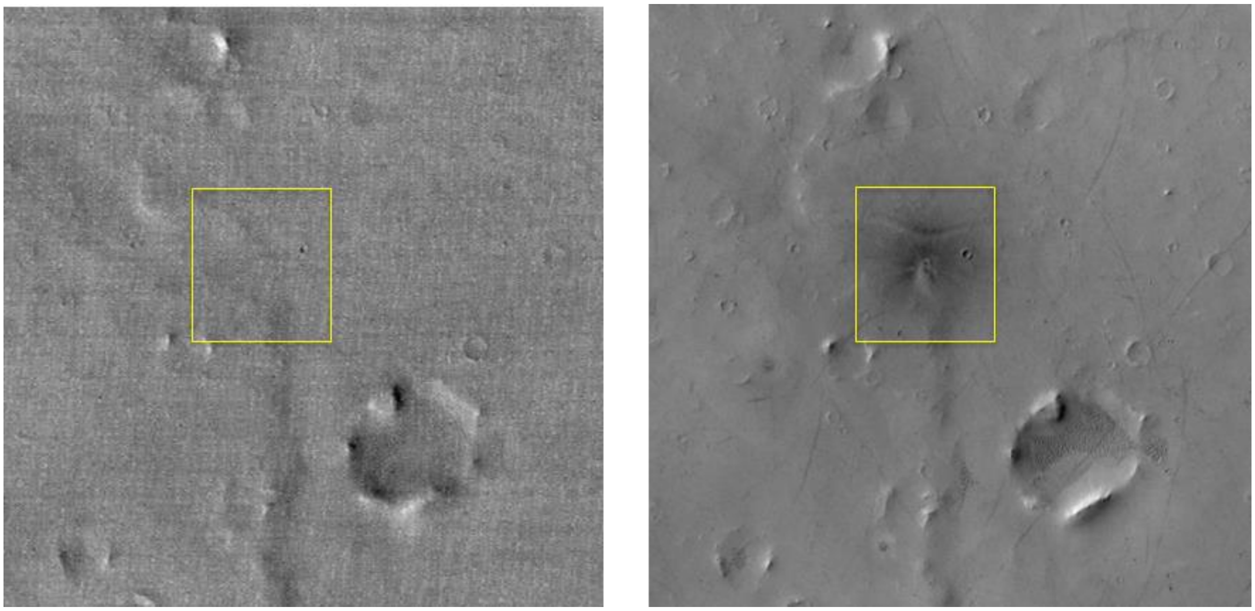


Figure 6: Crater impact and ejecta marked by Mars in Motion volunteers.

4. Best practice recommendations

This section presents both the lessons learnt regarding the interface design of a Mars citizen science project, and also a number of recommendations based on these outcomes. They have been reached both through the implementation of the iMars' 'Mars in Motion' project (Sprinks et al., 2015; 2016) and through previous work on "Planet Four: Craters project" (Sprinks et al., 2017). Deliverable 7.3 describes this process in more detail.

4.1. The influence of task and interface design on performance

The experimental work of deliverable 7.3 has shown that interface and task design factors can influence:

1. The intrinsic motivation and therefore engagement of citizen science volunteers
2. The data collected and analysis performed by citizen science volunteers

In reference to point 1, laboratory studies carried out investigating the manipulation of task design and the presentation of the imagery (side by side vs. flicker etc.) revealed that a number of different volunteer concerns that could be directly related to the process of analysing the data. Hence, the conclusion is that interface and task design factors can affect the intrinsic motivation of volunteers. For instance, participants reported that they felt much more successful at carrying out the task of identifying change when the imagery was 'flickered' compared to inspecting them 'side by side' ($p < 0.05$). Additionally, participants reported a preference for interface designs that afforded the user greater autonomy to choose the task to complete, and included a greater variety of tasks. This reported preference translates into improved engagement with the platform, with volunteers more likely to return and more often when the interface involved greater autonomy and task variety.

Concerning point 2, analysis revealed the interplay of different interface design factors and its influence on participants change detection and identification analysis. The subsequent design-related phenomena have surfaced as important volunteer data mediators:

- ***Task Type & Judgement*** – Simpler tasks with fewer volunteer judgements result in a significantly greater data volume being collected, and at a faster rate; however accuracy is reduced in terms of an increase in false-positive markers.
- ***Task Type*** – Volunteer accuracy (in terms of inter-participant and expert agreement) is significantly improved when the task can be related to its aim – i.e. the volunteer is aware of the measurement that the task is attempting to capture.
- ***Autonomy*** – Task workflow designs that constrain autonomy, forcing volunteers to complete every task and in a set order (and therefore spend more time on each classification), result in significantly greater volunteer accuracy - both in terms of inter-participant and expert agreement.
- ***Task Variety*** – VCS Interface designs that constrain task variety, forcing volunteers to do only one task on several classifications before moving on to a different one, result in significantly greater volunteer accuracy – both in terms of inter-participant and expert agreement.
- ***Autonomy & Variety*** – TWD's that involve both greater autonomy and variety, allowing the volunteer to do a number of different task types in their chosen order, result in significantly

reduced volunteer accuracy, opposing the expected outcome assimilated from previous human factors research.

- ***Image Pair Presentation*** – Images that are presented one after the other in a 'flicker' process will be analysed more quickly by volunteers compared to those presented side by side.
- ***Image Pair Presentation*** – Volunteers will be more accurate in terms of change detection when images are presented 'flickered' rather than side by side.
- ***Feature type*** – Volunteers are more accurate at detecting 'new feature' changes (i.e. features that are in one image, but not the other) compared to 'feature movement' (i.e. changes in the same features position between two images).

Therefore it is clear that different design factors influence online citizen science outcomes (such as engagement, data volume, analysis accuracy) in different ways - and it is the interplay between these effects that determines the answer to the best design approach. This being the case, it is clear that designing a 'one size fits all' interface to ensure the best results for a planetary science project is a highly ambiguous undertaking that is unlikely to be suitable for 100% of the scientific aims involved. It would be more rational for the developers of citizen science systems to balance the effects of task and interface design factors against the aims of the project, and the needs of their volunteer community on an individual basis.

4.2. Design guidelines for a planetary citizen science system

Based on the design influences outlined in the previous section, we now present guidelines derived through our experience of devising and studying the manipulation of interface and task design factors in a real-world citizen science application. They are framed to take into account the perspective of different stakeholders (developers, scientists and volunteers), and have been formulated through a subjective review of the findings of deliverables 7.1-7.3 and complimentary work on the Planet Four: Craters project. As such, these guidelines epitomise subjective interpretations, and should not be mistaken as aspiring to a maximum objectivity suitable for all Planetary Science examples.

The design guidelines may be considered as exercises conducted in a loosely sequential order:

1. ***Consider the public reaction*** to the science case you are presenting, what you are asking them to contribute to and the level of media attention this might receive, in order to gauge realistic expectations.
2. ***Consider the data needs*** in order to achieve the project's scientific aims – in terms of its volume, accuracy and granularity. This follows the findings of the experiments in deliverable 7.3, showing that task workflow and interface design factors can influence the data collected in differing ways.
3. ***Consider the role of the human volunteer*** and the machine in a human-computer system. Many citizen science projects now use volunteer analysis in conjunction with algorithmic solutions (as

is the case with Mars in Motion) and it is important to ensure both are being used as per their abilities, and that the volunteer understands their role and contribution to the project.

4. ***Consider if the project is realistic*** in terms of the compatibility of public reaction and data needs. For instance, does an interface design configuration that will produce the accuracy required also produce enough classifications over time based on your expected visitor size? If not, can a compromise be reached?
5. ***Conduct a laboratory trial*** of the platform, using non-expert participants. This follows the findings of deliverable 7.3 that revealed limitations of the system in terms of both its usability and the data collected that might not be apparent to an expert view. By doing this, many issues can be addressed before the platform is released en-masse – and before interface design changes can be an interruption.
6. ***Monitor volunteer behaviour*** and the data produced regularly. As revealed by their own opinions in the empirical part of this thesis, their preferences can change regarding task and interface design depending on their time on the platform. Whilst new users might prefer more guidance, experienced users might want a greater challenge and more freedom. Identifying when this change occurs and reacting to it in terms of the presented interface could prolong their engagement with the platform, and improve the data they produce.
7. ***Acknowledge failure is inevitable***, at least to some degree, and that crowd-sourced information will never be 100% accurate. Online citizen science is not conducted in a controlled environment, and therefore a degree of 'noise' will always be present. Consider how the platform will deal with erroneous, null or biased data.

These guidelines were derived from the experience gained through the process of conducting the work and derivation of the Mars in Motion and Planet Four: Craters projects. As such, they are unlikely to be complete or comprehensive as a suitable manual for 100% of citizen science endeavours. As discussed earlier in this section, the consideration of task and interface design is a complex undertaking, and its design should be as context-specific as reasonably possible in order to achieve a project's scientific goals.

References

Cox, J., Oh, E.Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., Graham, G., Holmes, K. (2015). Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering,* 17(4), 28-41.

Houghton, R.J., Balfe, N., & Wilson, J.R. (2015). Systems analysis and design. In J.R. Wilson & S.Sharples (Eds.) *The Evaluation of Human Work* (*4th Edition)*, 221-249, CRC/Taylor and Francis

Houghton, R.J., Wardlaw, J., Sprinks, J., Giordano, M., Bamford, S. & Marsh, S. (2016). Martian Factors: A systems ergonomics approach to citizen science. *Human Factors in Complex Systems,* Nottingham UK, June.

Kittur, A., Nickerson, J.V., Bernstein, M.S., Gerber, E.M., Shaw, A., Zimmerman, J., Lease, M., & Horton, J.J. (2013). The future of crowd work. *Annual Meeting of the ACM Special Interest Group on Computer-Human Interaction (CHI'13)*, San Antonio, TX, February.

Luczak-Roesch, M., Tinati, R., Simperl, E., van Kleek, M., Shadbolt, N., Simpson, R. (2014). Why won't aliens talk to us? Content and community dynamics in online citizen science. *Eighth International AAAI Conference on Weblogs and Social Media,* Michigan, USA, June.

Reeves, S. & Sherwood, S. (2010). Five design challenges for human computation. *NordiCHI'10: Proceedings of the 6th Nordic conference on Human-computer interaction*, 383-392.

Sharples, S. & Wilson, J.R. (2015). Methods in understanding human factors. In J.R. Wilson & S.Sharples (Eds.) *The Evaluation of Human Work (4th Edition)*, 1-36, CRC/Taylor and Francis.

Sprinks, J., Wardlaw, J., Houghton, R.J., Bamford, S., Morley, J. (2017). Task workflow design and its impact on performance and volunteers' subjective preference in Virtual Citizen Science. *International Journal of Human-Computer Studies,* 104, 50-63.

Sprinks, J., Wardlaw, J., Houghton, R.J., Bamford, S., Mardh, S. (2016). Mars in Motion: An online citizen science platform looking for changes on the surface of Mars. *AAS/Division for Planetary Sciences Meeting Abstracts,* 48, 426.1, Pasadena, CA, October.

Sprinks J., Houghton R.J., Bamford, S., Morley J., Wardlaw, J. (2015). Is that a crater? Designing citizen science platforms for the volunteer and to improve results. *European Planetary Science Congress 2015,* EPSC2015-695, 10.

Wilson, J.R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, *45*, 1, 5-13.