



Supported by Nesta, Arts & Humanities Research Council and public funding by the National Lottery through Arts Council England

Literature review

## **Complex game-based crowdsourcing in Cultural Heritage**

A review of the literature informing the design of Ten Most Wanted, a complex, collaborative game that crowdsources the discovery and verification of missing information about collection items. The review addresses challenges in encouraging and sustaining volunteer engagement, discusses game elements and reward structures, explores approaches to improving data quality and examines intellectual property issues in social content production. It concludes with a summary of findings and provides a set of recommendations for the Ten Most Wanted project.

Marcus Winter, University of Brighton

Draft version 25 July 2013

## Contents

1	Introduction .....	3
2	Terminology .....	4
2.1	Crowdsourcing v. Peer Production.....	4
2.2	Crowdsourcing v. Citizen Science.....	6
2.3	Games.....	6
2.4	Intellectual property .....	7
3	Crowdsourcing typologies.....	8
4	Engagement .....	11
4.1	Motivation.....	11
4.2	Encouraging initial engagement.....	13
4.3	Sustaining engagement.....	13
5	Game elements and rewards .....	16
5.1	Casual v. complex games.....	16
5.2	Game elements and reward structures .....	17
6	Data quality and validation .....	19
6.1	Professional v Amateur content creation .....	19
6.2	Approaches to increase rigour .....	20
7	Intellectual property .....	21
7.1	Copyright.....	21
7.2	Alternative frameworks .....	21
7.3	Current practice .....	22
8	Summary and conclusions.....	23
9	Acknowledgements.....	24
	References.....	25
	Recommendations for TMW .....	31

## 1 Introduction

This literature review was carried out in the context of Ten Most Wanted (TMW), a research and development project on complex game-based crowdsourcing in Cultural Heritage funded by the Digital R&D Fund for the Arts<sup>1</sup>. The review aims to situate TMW in a wider research context and inform the development of the TMW crowdsourcing platform which, together with related templates and best practice guidelines, will be made available to support similar efforts in other arts organisations looking to crowdsource aspects of their curatorial activities.

The basic idea of TMW is to use game play and social media as a means to involve the public in discovering and verifying previously undocumented facts about collection items. Unlike previous game-based crowdsourcing projects in the Cultural Heritage sector, which usually involve casual online games where single users complete simple tasks, such as tagging images or correcting Optical Character Recognition (OCR) errors, TMW explores the potential of complex, investigative games for crowdsourcing that involve both online and offline activities, and require sustained engagement and collaboration between players to solve problems posed by curators.

While TMW is being developed in the first place for the Arts University Bournemouth's Museum of Design in Plastics<sup>2</sup> (MoDiP), the project addresses shared challenges across the arts sector. These include finding sustainable models for enhancing digital collections with the help of volunteers, reaching new audiences and keeping them engaged over longer periods, verifying and integrating user-generated content with professionally curated collections and addressing copyright issues in an open and transparent manner that respects contributors' rights while meeting the needs of the arts organisation.

In line with the project's research questions, the review focuses in particular on:

- Ways to encourage and sustain volunteer engagement
- Design of game mechanics and reward structures
- Data quality and validation of contributed content
- Intellectual property aspects

While the review aims specifically to inform the design of a game-based crowdsourcing platform for the Cultural Heritage sector, it draws on literature from a wide range of related fields including non-game-based crowdsourcing, citizen science and peer production in the context of the open source software movement, as well as sociological and legal perspectives on social content production.

Reflecting the emergent and dynamic character of the field, the review includes not only books and peer-reviewed journals, but also conference presentations and reports produced by research groups and other agencies informing and advancing crowdsourcing and peer production in various contexts and domains.

---

<sup>1</sup> Digital R&D Fund for the Arts. <http://www.artsdigitalrnd.org.uk/>

<sup>2</sup> Museum of Design in Plastics. <http://www.modip.ac.uk/>

## 2 Terminology

### 2.1 Crowdsourcing v. Peer Production

There have been various attempts to define the term crowdsourcing and describe its many facets and types. A detailed analytical discussion is offered in Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012), who synthesised a broad, inclusive definition of crowdsourcing from 40 original definitions in the literature:

*"Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken."* (Estelles-Arolas and Gonzalez-Ladron-de-Guevara, 2012)

The careful wording of this definition, which tries to encompass all the possible actors, activities, motivations and outcomes of crowdsourcing, is an indication of the variety of actual implementations and reflects the many different interpretations of crowdsourcing: the field might just be too diverse and too fuzzy at the edges for a meaningful definition that is not unduly limiting or generalising or encroaching on other peer production models.

One approach to avoiding the pitfalls of a single definition is to look at a range of actual crowdsourcing projects and consider the views of practitioners and volunteers involved. Dunn and Hedges (2012) unpack crowdsourcing along various dimensions and give a broad overview of crowdsourcing projects in the Humanities, drawing together the findings of a literature review, a survey of participants, interviews with academics and participants and two academic workshops on the topic.

Another approach is to take a step back, look at patterns of organisational structure and participant behaviour in crowdsourcing projects and relate them to the type and quality of involvement aimed at in TMW. Instead of providing a definition of crowdsourcing per se, this approach identifies characteristics in various flavours of crowdsourcing that can inform the design of a supportive environment helping participants to tackle complex curatorial tasks.

A key difference between various understandings of crowdsourcing relates to the depth of user involvement, with knock-on effects on length of engagement, complexity of tasks, levels of collaboration and participation in project governance. While the term crowdsourcing was originally used by Howe (2006a) to describe the practice of using the Internet to outsource work to a large number of individuals, it was quickly adopted by others who, according to Howe (2006b), often used it "*incorrectly*" (i.e. differently from his intended meaning) in the sense of commons-based peer production (Benkler, 2002).

Whereas the former uses the "*spare processing power of millions of human brains*" (Howe, 2006a) and does not require any deep commitment or social connections between contributors, the latter involves groups of individuals collaborating "*on large-scale projects following a diverse cluster of motivational drives and social signals*" (Benkler, 2002), implying deep engagement and social interaction between participants.

The difference between the two is crystallised in Haythornthwaite's (2009) distinction between "*crowd*" and "*community*" in her discussion of lightweight and heavyweight peer production:

- Lightweight Peer Production (LWPP) involves large numbers of unrelated individuals who contribute to a product but don't have a say in the direction or development of the project itself. Work is broken down into small, discrete, rule-based tasks that require no or little prior domain knowledge. Reflecting the simple nature of tasks and contributions, projects often have quantitative reward structures. LWPP requires no long-term commitment and is not designed to create or maintain relationships between participants.
- Heavyweight Peer Production (HWPP) involves a community of individuals who contribute to the product as well as the direction and development of the project itself. Contributions are more complex and often require participants to make independent decisions and pay attention to the actions and contributions of others. Reflecting the nature of participants' involvement, projects usually have qualitative reward structures based on the level and quality of engagement. HWPP requires a critical mass and a certain level of commitment from participants to sustain the project.

Far from being mutually exclusive, Haythornthwaite (ibid) points out that the distinction between lightweight and heavyweight peer production is useful in the first place to identify "*two overlapping patterns of collaborative, contributory behavior [sic]*" rather than to classify projects, which often mix and match elements of both models to satisfy their specific needs.

Like many other projects inviting members of the public to contribute their time and effort towards a specific goal, TMW mixes LWPP and HWPP models and cannot clearly be defined as crowdsourcing in Howe's (2006a) sense or as commons-based peer production in Benkler's (2002) sense.

On the one hand, the overall direction and purpose of TMW is determined by the project team with participants having little say in strategic decisions. The core idea of TMW - curators selecting collection items they want to know more about and the public contributing missing information about these objects - further emphasises a tiered structure where the lead organisation sets targets for participants' enquiries. These elements in TMW are close to the LWPP model and characterise it as a crowdsourcing project in Howe's (2006a) sense.

On the other hand, TMW seeks to build and sustain a community of contributors, organised in independent and self-directed teams that tackle complex task. The nature of the tasks

requires deep engagement from participants, and in order to optimise communication and collaboration between volunteers, teams and project staff, participants are involved in shaping the platform and practices in the project through a participatory design process. These elements in TMW are close to the HWPP model and share characteristics with Benkler's (2002) idea of commons-based peer production.

Bearing in mind that TMW integrates both LWPP and HWPP patterns, this review uses the terms crowdsourcing and peer production interchangeably, depending on the specific aspect under discussion.

## 2.2 Crowdsourcing v. Citizen Science

Some scholars make an explicit distinction between crowdsourcing and citizen science, the latter being defined by Cohn (2008) as "*a form of research collaboration involving members of the public in scientific research projects to address real-world problems*".

Dunn and Hedges (2012) point out that "*the use of the word 'science' (at least in the usual Anglophone sense) confines the activities reviewed (in terms of both the methods and the content) to a particular epistemic bracket which inevitably excludes some aspects of humanities research*", while Ridge (2012) suggests that citizen science projects are "*focused on the production of data as accurately and efficiently as possible*" whereas crowdsourcing in GLAM<sup>3</sup> sectors is "*as much about engaging people with the content as it is about content production*".

Given that TMW aims to crowdsource the systematic discovery and verification of information about collection artefacts, which in many ways is a characteristic of citizen science, while also aiming to broaden public engagement with the collection, which is a familiar feature of crowdsourcing in Cultural Heritage, a strong distinction between these concepts seems problematic. One of the key aspects in TMW is to develop a rigorous methodology that produces valid, reliable and credible results, which, independent of ontological and epistemological traditions, is the overarching goal of research in both, humanities and natural sciences. Moreover, many projects classed as citizen science have a strong secondary agenda of engaging the public with STEM<sup>4</sup> subjects, mirroring similar engagement efforts in the humanities.

The review will therefore draw on both crowdsourcing and citizen science literature, to inform the design of TMW while using the terms crowdsourcing and peer production throughout for simplicity and consistency.

## 2.3 Games

Games are being researched in a wide range of contexts, all of which have different perspectives and offer their own, sometimes conflicting, definitions and terminologies.

For instance, taking an anthropological perspective, Orwant (2000) defines games as all leisure activities that are not play or sport, with play having no explicit goal and sport

---

<sup>3</sup> GLAM is an acronym for Galleries, Libraries, Archives and Museums

<sup>4</sup> STEM is an acronym for Science, Technology, Engineering and Mathematics

involving a test of physical ability, while Prensky (2001), taking an educational perspective, understands game to be a subset of play that is organised and helps us to learn.

Play, a primary formative element in human culture (Huizinga, 1950), and game are clearly related, with some languages even using the same word for both, e.g. the phrase "*to play a game*" would be "*jouer un jeu*" in French and "*ein Spiel spielen*" in German (Parlett, 1999).

However, while the meaning of play is by and large the same across cultures, games are understood differently throughout different cultures, and what might be considered a game in one country might not be considered a game or even appropriate in another country (Hinske et al., 2007).

Further complicating the picture is the cleavage between narratologist and ludologist conceptions of games (Frasca, 2003), with the former emphasising representation and narrative structure and the latter focusing on the underlying abstract rule systems.

Reflecting these different perspectives and accounting for TMW's pragmatic approach to games as a means to motivate and engage volunteers, we use a loose, inclusive definition of game synthesised from Dempsey et al. (1996), Fabricatore (2000), Prensky (2001) and Hinske et al. (2007), the latter itself being an amalgam of definitions by Costikyan (2002), Lindley (2002), Crawford (2003), Klabbers (2003), Juul (2003) and Salen and Zimmermann (2003).

A game consists of actors, resources, and a set of structural elements:

- Framework of agreed constraints and/or rules
- Challenge/Opposition/Competition, either with oneself or others
- Management of resources and/or time
- Pursuit of objectives and goals
- Outcomes and feedback
- Representation or story
- Emotional attachment

Given the explorative character of TMW and the fact that it integrates both LWPP and HWPP patterns into what has been termed "*serious games*" (Abt, 1970), we will use this working definition as a superset of potential game elements rather than a strict checklist of mandatory elements to qualify as a game.

## 2.4 Intellectual property

Intellectual Property (IP) refers to the exclusive rights associated with original creation and broadly encompasses patents, trademarks, designs and copyright (IPO, 2013a). This review focuses in particular on copyright which applies to written or otherwise recorded original works.

As researching, obtaining and asserting IP is an integral part of curatorial work, museums and other keepers of digital collections usually have internal guidelines on IP related issues. However, such internal guidelines tend to focus primarily on preventing the infringement of

others' IP and protecting the organisation's own IP when presenting collection items to the public, e.g. on the museum's website.

Crowdsourcing introduces new complexities into this already problematic aspect, as social, collaborative production is difficult to reconcile with the focus in current copyright on the relationships between a single, sovereign author/owner of content and non-owners wishing to make use of it (Elkin-Koren, 2011).

As a consequence, the integration of content produced in diverse, collaborative environments into professionally curated collections considerably complicates curators' work and requires a comprehensive IP framework which on the one hand respects the rights of both, individual participants and the community, while ensuring on the other hand the rights of the organisation in the collection it keeps and maintains.

In order to address these challenges, this review explores the conflicts between current copyright and social production, looks at alternative frameworks aiming to address these conflicts, attends to practical issues of using these alternatives and discusses the intricate relationships between copyright, attribution and motivation in social production.

### 3 Crowdsourcing typologies

In order to situate the research in the wider field and better understand its contribution, this section discusses TMW in the context of well-known typologies of crowdsourcing and peer production, including Shirky's (2009) levels of group action, Bonney et al.'s (2009) categories of citizen science, Wiggins and Crowston' (2011) application areas for citizen science, Oomen and Arroyo's (2011) key stages in GLAM's digital content lifecycle which crowdsourcing can support, Ridge's (2012) list of activity types in game-based crowdsourcing in the Cultural Heritage sector and Dunn and Hedges' (2012) overarching typology framework for crowdsourcing in the Humanities.

Shirky (2009) identifies three levels of group action:

- *Sharing*: individuals share their works with the group, creating value through aggregation
- *Cooperation*: mutual synchronisation of group members' actions, involving a higher level of reciprocity
- *Collective Action*: decisions adopted by the group become binding upon its individual members

Adapted from Elkin-Koren (2011)

TMW investigates in particular the two higher levels of group action, cooperation and collective action. Participants loosely cooperate with others or form small groups in which they carry out enquiries through collective action. In both cases they depend on social communication channels that provide awareness towards other group members' activities. The project also provides opportunities to explore a combination of these levels as groups of individuals pursuing collective action might compete with each other in what can be seen as a form of cooperation from an overall project perspective.

Bonney et al. (2009) define three categories of crowdsourcing:

- *Contributory*: participants respond to an open call and contributions are defined and directed by scientists
- *Collaborative*: participants contribute data in the context of a central research design, help to analyse data, disseminate findings and improve the research design
- *Co-creative*: participants work together with scientists to design, promote and coordinate a research project

Adapted from Dunn and Hedges (2012)

TMW falls primarily into the category of contributory projects: it involves an open call for participants to discover, document and verify facts about collection items selected by curators, and participants receive feedback, advice and support from the project team during their investigation. However, it also incorporates collaborative aspects, as TMW employs a user-centred design approach where participants have a say in the design and direction of the platform and practices, and co-creative aspects as participants are encouraged to actively promote the project and recruit new participants.

Wiggins and Crowston (2011) identify five distinct types of crowdsourcing:

- *Action-oriented*: participatory action research encouraging participant intervention in local concerns, using scientific research as a tool to support civic agendas
- *Conservation*: engaging citizens in stewardship and natural resource management, primarily in the area of ecology, often with educational goals
- *Investigation*: data collection from the physical environment for scientific research, often with secondary educational agenda and involving large scales of participation
- *Virtual*: project where all activities are ICT-mediated and have no physical elements; goals often similar to *Investigation* projects
- *Education*: projects with education and outreach as primary goals, providing formal and informal learning resources, often focusing on a specific area or place

Adapted from Wiggins and Crowson (2011)

While TMW does not fit any of these categories particularly well, it is probably closest related to *Investigation* projects as participants try to uncover facts about collection items in a distributed fashion and in the process acquire both domain and methodological knowledge pertinent to curatorial research in this field. Some aspects in the project, including the social-constructivist, community-based approach to knowledge generation (Vygotsky, 1978; Wenger, 1998), the situated learning approach providing practice-based guidelines and background information to participants (Rogoff, 1982; Lave, 1988), and the declared goal of outreach and engagement, also resonate with educational crowdsourcing projects.

Oomen and Arroyo (2011) describe how specific aspects of the digital content lifecycle in the GLAM<sup>5</sup> sector can be supported by crowdsourcing initiatives:

- *Correction and Transcription*: correct or transcribe outputs of digitisation processes
- *Contextualisation*: add contextual knowledge to objects
- *Complementing*: suggest additional objects to be included in a collection

---

<sup>5</sup> GLAM is an acronym for Galleries, Libraries, Archives and Museums

- *Classification*: gather descriptive metadata related to objects in a collection.
- *Co-curation*: provide inspiration/expertise to create exhibits
- *Crowdfunding*: pool money and other resources to support efforts initiated by others

Adapted from Oomen and Arroyo (2011)

TMW primarily supports aspects of contextualising and classification. It creates new knowledge as participants investigate specific aspects of collection items flagged up by curators. As these aspects relate to specific metadata fields in the collection, the activities help to classify objects and make them searchable. To some degree, TMW also supports complementing the collection as participants might contribute variations of selected items (e.g. previously undocumented colours or sizes) either in physical or digital form.

Ridge (2012) proposes activity categories from the particular perspective of game-based crowdsourcing, pointing out that they "*should be read in the light of discussion about the difference between crowdsourcing and user-generated content and in the context of things people can do with museums and with games*" (ibid):

- Tagging (e.g. adding key words and thereby improving searchability)
- Debunking (e.g. flagging content for review, providing corrections)
- Recording a personal story (e.g. contextualising items, enriching collections)
- Linking (e.g. linking to other objects, subject authorities, related media or websites)
- Stating preferences (e.g. choosing between two objects)
- Categorizing (e.g. applying structured labels to a group of objects)
- Creative responses (e.g. writing a fake history for an object)

Adapted from Ridge (2012)

Surprisingly, TMW fits none of these categories as it involves participants in the role of investigators who uncover facts about collection items. One reason for this lack of congruence might be that Ridge's main focus is on casual games (Ridge, 2011a) involving LWPP patterns of participation, while TMW aims to integrate game elements to support deep, long-term engagement with the project using HWPP patterns. In this respect TMW marks new territory and has the potential to complement existing activity categories based on casual games with new categories pertaining to longer-term investigative games.

Dunn and Hedges (2012) synthesise a range of crowdsourcing categorisations discussed in the literature into an overarching typology framework composed of four key facets and the relationships between them:

- *Process*: composed of tasks operating on an asset; producing outputs
- *Task*: an activity undertaken in order to create, process or modify an asset
- *Asset*: the content transformed through a crowdsourcing activity
- *Output*: what is produced as the result of applying a process to an asset

Adapted from Dunn and Hedges (2012)

According to Dunn and Hedges (2012), the *Process* facet is the key determinant in this typology as it corresponds to research methods in the Humanities. While the authors give a wide range of examples for each of these facets drawing on their review of crowdsourcing in

the humanities, they stress that the list is not exhaustive and that their typology "*will develop and evolve as the field of humanities crowdsourcing itself evolves*" (ibid).

TMW has the potential to help extend this framework by developing and evaluating a complex game-based approach to crowdsourcing involving HWPP patterns, as opposed to a casual games based approach involving primarily LWPP patterns (e.g. Trant, 2009; Brooklyn Museum, 2009; Oomen et al., 2010; Ridge, 2011a). In the *Process* facet, for instance, TMW explores new categories for discovering, documenting and verifying new knowledge about collection items. With regard to other facets, TMW will provide new perspectives on investigative *Tasks*, potentially operate on new compound *Asset types* mixing digital and physical content, and involve complex *Output types* that not only complement and enrich collections but also include a reference- and review-trail relating to provenance and content verification.

## 4 Engagement

Regarding the sustainability of crowdsourcing projects as a means to enrich collections and engage the public, Dunn (2013) points out "*rampant short-termism*" across the spectrum of research funding where particularly vulnerable projects stall when the funding period ends because they cannot sustain themselves. An important question in crowdsourcing projects is therefore how to encourage participation over longer periods while minimising the need for costly human resources to facilitate and sustain engagement. A key aspect in this context is motivation: Why do people take part in crowdsourcing projects in first place? What do they expect to get out of the experience?

### 4.1 Motivation

Motivations to participate in crowdsourcing projects tend to differ widely (Roy et al, 2012), reflecting the diversity of topics and approaches. In the context of LWPP patterns of crowdsourcing, Ridge (2011a), for instance, cites altruism, validating procrastination, stress relief, cognitive exercise and fun as motivating factors for taking part in casual crowdsourcing games, while in the context of HWPP patterns of crowdsourcing Benkler (2004) cites self-expression, creative satisfaction, a desire to establish one's online reputation and a wish to strengthen one's self-esteem as motivating factors.

A high-level discussion of motivational factors in peer production is provided by Rafaeli and Ariel (2008), who explore the concept of motivation from psychological, sociological, media studies and economic perspectives. One recurring theme in their discussion is the distinction between "*intrinsic motivation*" and "*external rewards*" (Deci, 1975), a concept which resonates with Hars and Ou's (2001) distinction between motivations innate to an individual's psychological makeup ("*internal factors*") and motivations originating from the environment ("*external factors*"), and Peddibhotla and Subramani's (2007) distinction between "*self-oriented*" (e.g. fun, self-expression, personal development ) and "*other-oriented*" (e.g. social affiliation, altruism, and reciprocity ) motives for contribution.

Dunn and Hedges (2012) point out that most individuals taking part in crowdsourcing projects have multiple intrinsic, extrinsic and altruistic motivations, but that there is usually one dominant motivating factor which is often based on genuine interest in the subject area.

- As an example for intrinsic motivation, Dunn and Hedges (2012) cite research by Raddik et al. (2010) which found the top motivation for participants in the popular Galaxy Zoo<sup>6</sup> project to be a personal interest in astronomy. Similar findings were reported by Nov et al. (2011) for the Stardust@home<sup>7</sup> project where volunteers cited enjoyment and enthusiasm for the goals of the project as their main motivations to take part. Intrinsic motivations can also relate to the activities carried out by volunteers, e.g. Grove-White et al. (2007) found enjoyment when working outdoors in a natural environment to be an intrinsic motivation of volunteers in a conservation project.
- An example for altruistic motivations is provided by Bradford and Israel (2004) who found that volunteers in projects with strong advocacy goals often cite a wish to help protected species as their main motivation. This is supported by Grove-White et al. (2007) who found that participants often cite the protection of local species and habitats as a major motivation to volunteer. Altruistic motivations also play a part in projects without advocacy goals. For instance, a large group (13%) of volunteers in Galaxy Zoo say their primary reason for participation is to help (Raddik et al., 2010).
- Extrinsic motivations often refer to rewards volunteers can expect from their participation. Apart from symbolic rewards, such as points or badges in game-based crowdsourcing projects, these can include psychological rewards such as the feeling of being part of a community, which is cited by many reports on crowdsourcing as an important motivation to participate (cf. Dunn and Hedges, 2012). Extrinsic motivations can overlap with altruistic motivations, for instance Holley (2009) notes that many participants in the Australian Newspaper Digitisation Programme<sup>8</sup> are family history researchers who feel a responsibility to help fellow genealogists.

One particular motivation mentioned by participants in a range of projects is discussed by Prestopnik and Crowston (2011) in the context of game-based crowdsourcing. They point out that "*fun*" is a motivation cited across many different projects and postulate that the more fun a project promises, the more people will join it, and the more fun a project actually involves, the more motivated volunteers will be to continue or expand their participation.

Based on these findings, it seems important for TMW to address a range of different motivations that participants may have. This includes catering for intrinsic motivations, such as satisfying a genuine interest in design or plastic, altruistic motivations, such as appealing for support for MoDiP as a small organisation and the only museum in the UK focusing on design in plastics, and extrinsic motivations such as fostering a supportive community and designing a suitable reward system. Pervading all these measures should be an emphasis on

---

<sup>6</sup> Galaxy Zoo, <http://www.galaxyzoo.org/>

<sup>7</sup> Stardust@home, <http://stardustathome.ssl.berkeley.edu/>

<sup>8</sup> Australian Newspaper Digitisation Programme. <http://trove.nla.gov.au/newspaper>

making participation fun and emphasising this aspect in the project's communication and branding.

## 4.2 Encouraging initial engagement

An initial hurdle for many crowdsourcing projects is to recruit a critical mass of participants. While projects adopting a LWPP model have linear progression with rising user numbers and therefore can start small and scale up operations as participation picks up, projects with a HWPP model need a critical mass of participants to function at all (Haythornthwaite, 2009).

Causser and Wallace (2012) describe a wide range of measures and activities to publicise and recruit participants in the context of Transcribe Bentham<sup>9</sup>. Segmenting their target audience, they cumulatively targeted campaigns at the general public, at the academic and professional community, and at schools with a view to educational aspects of the project:

- The general public was targeted through the project website, an active presence in a various social media outlets, leaflets and a press release distributed to major British newspapers and magazines.
- The academic and professional community was targeted through mailing lists, online forums and presentations at seminars, conferences and workshops. Further efforts included contacting individual academics, libraries, archives, and educational bodies, as well as developing educational resources and sending out a press release to publications targeting relevant communities.
- Schools were targeted with specific information pages, reading lists and links to manuscripts of particular relevance to the curriculum. Other measures included a press release sent to publications targeting teachers and schools and a highly publicised visit of school children to test the newly designed project website.

With respect to the effectiveness of these efforts, the authors note that the publicity campaign as a whole was a success and that the press release in particular raised public awareness of the project as it led to media coverage in multiple countries (Causser and Wallace, 2012). Other projects, too, emphasise the importance of broadcast media for an initial push in publicising the project and the role of social media in terms of amplifying and further spreading the message among specific communities, potentially leading to drastic increases in participation (Raddick et al., 2010).

These accounts provide valuable ideas for publicising TMW. They suggest that the project should develop a broad range of dissemination activities and materials and in particular make efforts to get broadcast media interested in the project.

## 4.3 Sustaining engagement

Sustaining engagement in crowdsourcing projects, especially over longer periods and beyond the funding period, is a crucial aspect with respect to delivering long-term value for public money and avoiding the "*rampant short-termism*" in research funding criticised by Dunn (2013). The literature describes a wide range of design decisions, mechanisms and practices

---

<sup>9</sup> Transcribe Bentham, <http://blogs.ucl.ac.uk/transcribe-bentham/>

aimed at keeping volunteers interested in crowdsourcing projects and ensuring their continued participation.

### **Communication**

One recurring theme across many research reports is the importance of communication, not only between project team members and volunteers, but also between participants themselves as they shape their community. Roy et al. (2012) found that several projects use social media platforms for their communication and note that this approach has the added benefit that messages and conversations are visible even to outsiders, giving the impression of activity within the project and thereby potentially attracting new participants. As pointed out by Haythornthwaite (2009), communication is particularly important in HWPP communities that require participants to follow each other and be aware of other's activities in order to coordinate collaboration.

While most projects rely mainly on online communication, some also provide opportunities for face-to-face communication. According to Oomen and Arroyo (2011), Oxford University's Great War Archive<sup>10</sup> involved meet-ups with volunteers and project staff supporting the collection of items contributed by the public. In a similar fashion, Haklay and Weber (2008) describe local "*mapping parties*" as part of the engagement efforts in the federated Open Street Map<sup>11</sup> project, which are designed to recruit new participants and offer first-hand experience of collecting and uploading content to the system. According to the authors, "*mapping parties play an essential part in creating and fostering local OSM user groups and creating a vibrant social community around the project*" (Haklay and Weber, 2008).

### **Incentives and rewards**

While a majority of participants in crowdsourcing projects cite in the first place intrinsic and altruistic motivations (Dunn and Hedges, 2012), a large number of projects provides rewards to encourage sustained engagement.

Some projects offer rewards as part of an overall attempt to make participation more game-like, such as for instance participants being "*promoted to 'captain' of a vessel if they complete the most transcriptions for that ship*" in the Old Weather project<sup>12</sup> (Romeo and Blaser, 2011). Other projects, such as the Atlas of Living Australia<sup>13</sup>, use digital badges as straightforward rewards for specific achievements. Hill et al. (2012) point out, however, that there are no rigorous studies to date demonstrating the effectiveness of digital badges in enhancing motivation.

Financial rewards, while commonplace in other contexts, are problematic in a crowdsourcing context. Rafaeli, Raban and Ravid (2007) found in empirical studies that while financial rewards can lead to higher levels of participation, participants' behaviour was still strongly moderated by social variables. A more serious problem with financial rewards is pointed out

---

<sup>10</sup> Great War Archive. <http://www.oucs.ox.ac.uk/ww1lit/gwa>

<sup>11</sup> Open Street Map. <http://www.openstreetmap.org/>

<sup>12</sup> Old Weather Project. <http://www.oldweather.org/>

<sup>13</sup> Atlas of Living Australia (ALA) Biodiversity Volunteer Portal. <http://volunteer.ala.org.au/>

by Mason and Watts (2009), who found in experiments using Amazon's Mechanical Turk<sup>14</sup> that direct monetary incentives increased the quantity but not the quality of contributions.

While direct financial rewards might therefore not be suitable in projects employing HWPP patterns, which depend on high quality contributions, indirect financial rewards might be more appropriate. Elkin-Koren (2011) points out the importance of indirect financial rewards in the form of improved employment prospects or potential contract work based on one's standing in particular online communities. In this context it is important for participants to be publicly credited for their contributions (see below and also *6.1 Copyright* in this context).

### ***Recognition and attribution***

Recognition and attribution have been linked to participants' motivation and sustained engagement in several reports. Roy et al. (2012) emphasise the importance of maintaining "*strong links between the data and data providers*" as a key insight across research studies in this context. As an example they cite maps used in many citizen science projects to link data to locations and participants and visualise how "*contributions are closing gaps in knowledge*".

An even deeper and more persistent way of crediting participants for their contributions is reported by Mendez (2008) in the context of Stardust@Home<sup>15</sup> where participants are invited to search for dust particles captured in aerogel during a space mission to comet Wild 2 in January of 2004. Participants who discover a dust grain are listed as a co-authors in scientific publications announcing the discovery and in addition are given the privilege of naming the dust grain, promising lasting fame as a potential reward for their work.

### ***Ownership***

Ownership, in terms of co-determination and governance of a project, has been identified as an important factor regarding sustained engagement in particular for HWPP models of crowdsourcing. Citing research by O'Brien, Townsend and Ebdon (2008) into the motivations, barriers and benefits of environmental volunteering, Roy et al (2012) emphasise that community involvement in decision-making helps to develop partnerships and increase participation.

While co-determination might not be viable in all projects and sometimes run against the basic design of a project, it can be limited to specific aspects, such as the tools and platforms used by the project. Far from being neutral as claimed by Shirky (2008), tools and platforms profoundly shape the nature of peer production through their design (Nissenbaum, 2005; Elkin-Koren, 2011), for instance determining whether users can stay anonymous or not, or whether content can be transferred for uses in other contexts. By employing a User Centred Design approach (Norman, 1988) in the development of tools and platforms, crowdsourcing projects can therefore cover some aspects of co-determination without the need to relinquish overall strategic control.

---

<sup>14</sup> Amazon Mechanical Turk. <https://www.mturk.com/mturk/>

<sup>15</sup> Stardust@Home. <http://stardustathome.ssl.berkeley.edu/>

### **Training**

Closely related to intrinsic motivation and volunteers' genuine interest in project goals (e.g. Raddik et al., 2010), volunteers appreciate training opportunities in crowdsourcing projects that enable them to increase their skills and expertise (Roy et al., 2012). One particular example in this context is provided by Holley (2009), who noticed that many participants in the Australian Newspaper Digitisation Programme<sup>16</sup> were family history researchers who were "highly motivated to learn new skills in order to get the information they need" (ibid).

Depending on the size, structure and approach of projects, training can be provided either in person, for instance in the form of volunteer workshops (e.g. Bradford and Israel, 2004; Cohn, 2008; Haklay and Weber, 2008) or, as is more often the case in crowdsourcing projects, in the form of online support materials (Roy et al., 2012).

Overall, the literature describes a wide range of approaches to tap into volunteers' motivations and sustain long-term engagement in crowdsourcing projects. These include

- facilitating communication between all actors in a project through online channels and offline events
- designing a suitable reward structure that reflects engagement patterns (LWPP/HWPP) in the project
- providing incentives that help participants to enhance their professional profile
- recognising participants' contributions through public attribution and establishing lasting connections between users and user-generated data
- respecting participants' views on project governance, tools and practices and providing ways for co-determination in these aspects
- providing training opportunities that enable participants to develop their skills through workshops and online training materials

All of these approaches are relevant for TMW and can be related to specific aspects of the project. Some link directly to secondary research questions about reward structures and intellectual property issues. The literature therefore provides not only solutions, but also questions that require further exploration.

## **5 Game elements and rewards**

### **5.1 Casual v. complex games**

Looking at Oomen and Aroyo's (2011) classification of how crowdsourcing can support core activities in the digital content life cycle of heritage organisations, it becomes obvious that most game-based crowdsourcing projects fall into *Classification* or *Correction and Transcription* of content.

Examples for classification games include ESP game (Von Ahn, 2006), *steve.museum* (Trant, 2009), Tag! You're It (Brooklyn Museum, 2009), *Waisda?* (Oomen et al, 2010) and *Museum*

---

<sup>16</sup> Australian Newspaper Digitisation Programme. <http://trove.nla.gov.au/newspaper>

Metadata Games (Ridge, 2011a; 2011b), all of which involve tagging content in order to add semantic metadata and increase its discoverability.

Examples for correction and transcription games include Freeze Tag! (Bernstein, 2009), Old Weather (Romeo and Blaser, 2011) and Transcribe Bentham (Causar and Wallace, 2011). While the former involves the correction of user-generated tags to improve data quality, the latter two involve the transcription of materials and qualify as games in the context of this review because they incorporate game elements as set out in 2.3 *Games*.

Both of these types of activities, classification and correction & transcription, can be broken down into simple tasks suitable for casual, single-user games that don't require much prior knowledge or long-term commitment from players. They correspond to a LWPP model of crowdsourcing and fit the prevalent idea of game-based crowdsourcing as something "*light hearted*", "*easy*", "*fun*" and therefore more likely to engage volunteers (Holley, 2010).

TMW, by contrast, focuses on complex tasks that involve initiative, collaboration between players and sustained engagement to tackle complex problems related to the discovery and verification of new knowledge about collection items.

The only hint at using a game with these qualities in the context of crowdsourcing was found in Ridge's (2011b) discussion of Donald's detective puzzle as "*an experimental game that explored the design issues around longer forms of content that required some form of research or personal reference*" (ibid). Developing and evaluating two versions of Donald, Ridge (2011b) found that the concept of more advanced activities and user-generated content in games is feasible, but would benefit from matching tasks to specifically selected objects, closely defining tasks in a call to action and making available results from previous challenges available to players in order to provide them with examples of appropriate content.

Ridge's (2011b) findings are highly relevant for TMW and supports a range of fundamental design decisions in the project with empirical evidence, such as the selection of ten objects to focus on and the curator's brief for each object.

## 5.2 Game elements and reward structures

Many of the structural elements of games defined in 2.3 *Games* are inherent in crowdsourcing projects and more a matter of interpretation than design:

- a framework of agreed constraints and/or rules can be represented by the guidelines, procedures and acceptable quality standards in a project
- management of resources and/or time is an implicit necessity for volunteers who use their own resources and don't get paid for their involvement
- the pursuit of objectives or goals can be mapped to the tasks set by the project team and the overall project goals
- outcomes and feedback can be mapped to the outcomes of crowdsourcing activities and the feedback volunteers receive from the community and the project team

- representation or story can be provided by the context, problem description and support material in which tasks are framed and by the metaphors used in the project, such as for instance the concept of taking on a role on a ship in Old Weather<sup>17</sup> (Romeo and Blaser, 2011) or the concept of the iconic Ten Most Wanted list in TMW.
- emotional attachment can relate to the feeling of belonging to a community and, in the case of intrinsic or altruistic motivations, to identifying with the project goals (e.g. Bradford and Israel, 2004; Grove-White et al., 2007).

The key structural element not covered in this list is *Challenge/Opposition/Competition, either with oneself or others* (see 2.3). Citing research by Malone (1980) into the design of instructional computer games, Von Ahn and Dabbish (2008) propose that "*challenge translates into game features like timed response, score keeping, player skill level, high-score lists, and randomness*", i.e. mechanisms and reward structures commonly associated with casual games. Serious games (Abt, 1970), by contrast, often have more complex and qualitative reward structures reflecting their complex goals. This difference is also mentioned in Haythornthwaite's (2009) distinction between LWPP models of crowdsourcing, which often have quantitative reward structures reflecting the simple nature of contributions, and HWPP models of social production, which usually have qualitative reward structures reflecting the level and quality of participants' engagement.

With respect to *quantitative* reward structures, Ridge (2011b) points out that careful reward design can help to maximise the amount of data generated by participants. An important design heuristic for quantitative rewards is provided by Von Ahn and Dabbish (2008), who stress that rewards should provide a clear connection between effort, performance and outcomes (cited in Ridge, 2011b). Typical examples of quantitative reward structures include points, ranks, levels and geographic or temporal high-score lists and leader boards (e.g. Von Ahn and Dabbish, 2008; Brooklyn Museum, 2009; Ridge, 2011b; Causer and Wallace, 2012).

With regard to *qualitative* rewards, the literature offers a rich variety of approaches:

- *access to exclusive content*: Romeo and Blaser (2013) report that Solar Stormwatch<sup>18</sup> provides participants with near-real-time data from the STEREO spacecraft which enables them to detect solar storms up to three days before they reach earth. Feedback in the project suggests this feature is highly appreciated by participants.
- *badges for specific achievements*: Romeo and Blaser (2013) describe the introduction of badges earned by volunteers and displayed on their profile page on the project website to indicate extraordinary achievements.
- *publication and attribution*: Both Solar Stormwatch and Stardust@home<sup>19</sup> credit involved participants in publication relating to the discovery of solar storms or cosmic particles respectively (Mendez, 2008; Romeo and Blaser, 2013). The strong link between attribution and potential indirect financial rewards such as for instance improved employment has been pointed out by Elkin-Koren (2011).

---

<sup>17</sup> Old Weather Project. <http://www.oldweather.org/>

<sup>18</sup> Solar Stormwatch. <http://www.solarstormwatch.com/>

<sup>19</sup> Stardust@home, <http://stardustathome.ssl.berkeley.edu/>

- *naming privileges*: Stardust@home already gives participants the highly coveted privilege of naming cosmic particles they found (Mendez, 2008). Solar Stormwatch plans to let participants name solar storms in the future (Romeo and Blaser, 2013).
- *roles*: In the Old Weather<sup>20</sup> project, volunteers who complete the most transcriptions for a specific ship are promoted to captain that vessel (Romeo and Blaser, 2013).
- *progress visualisations*: Rome and Blaser (2013) report that progress for each ship in Old Weather was visualised on a map, which inspired volunteers to set their own milestones in terms of reaching a specific port by a specific time.
- *personal thank you notes*: Oomen and Arroyo (2011) report that participants in the British Library's UK Soundmap<sup>21</sup> project received a personal thank you message every time they uploaded a recording.

Many of these rewards are relevant to TMW and can either be implemented directly, such as for instance personal thank you notes, or adapted to its specific context.

## 6 Data quality and validation

### 6.1 Professional v Amateur content creation

One of the key advantages of crowdsourcing is that it combines audience engagement with the production of useful outcomes. In the context to TMW, this translates to a sustainable model of maintaining and extending collections by delegating aspects of curatorial work to members of the public.

A potential downside is that the public usually lacks the expert knowledge and skills of professional curators. While Brabham (2008) postulates that crowdsourcing can lead to solutions superior in quality and quantity to professional efforts, there are widespread concerns among professionals about data quality. Some of these concerns are highlighted in Eveleigh's (2012) discussion of participatory archives:

*"User participation initiatives in archives are haunted by a fear that a contributor might be wrong, or that descriptive data might be pulled out of archival context, and that researchers using collaboratively authored resources might somehow swallow all of this without question or substantiation."* Eveleigh (2012)

From a curator's perspective, data quality and verification are critical to avoid compromising quality standards for the collection as a whole. Introducing invalid data would not only impact on the collection's value as a research resource but also undermine the organisation's authority, which is a distinguishing quality specifically for heritage organisations (Oomen and Arroyo, 2011).

Dunn and Hedges (2012) point out that data quality is also important from the perspective of the community, which wants to be reassured that the outputs of their efforts are useful and academically valid. It is important therefore that projects are clear about what their *"quality*

<sup>20</sup> Old Weather Project. <http://www.oldweather.org/>

<sup>21</sup> UK Soundmap. <http://sounds.bl.uk/Sound-Maps/UK-Soundmap>

requirements and processes are (and adhere to them), and also make the QA process evident" (ibid).

## 6.2 Approaches to increase rigour

Oomen and Arroyo (2011) suggest that a combination of technological, psychological and organisational measures are needed to support heritage organisations' transition from "*the old in-situ culture based on controlled authority [to] the new in-vivo reality based on the wisdom of the crowd*". These include establishing behavioural norms, setting quality targets by example and filtering or correcting invalid content (ibid).

Other measures suggested in the literature to improve data quality in crowdsourcing projects can be broadly grouped into four approaches:

1. *Make the task easier*. Holley (2009) suggests that increasing the quality of the materials volunteers work with makes errors less likely. This is a specific form of the more general concept of making the task easier, which is a key idea at the root of crowdsourcing: breaking down complex problems into small, simple tasks that do not require any specialist knowledge.
2. *Train/inform volunteers*. Cohn (2008) suggests training volunteers in order to give them a better understanding of professional standards and practices. A more lightweight approach might just inform participants of the organisation's needs: Kidd (2013) describes how citizen journalists during the Arab Spring met the requirements of broadcasters by using establishing shots to verify their positions and timings.
3. *Crowdsource quality control*. Raddick et al. (2010) describe how user-generated classifications of galaxies in the GalaxyZoo project are "*written into a database and compared with the findings of other volunteers*". This approach can also be made explicit: Brooklyn Museum's Freeze Tag game (Bernstein, 2009) involves players in the clean-up of user-generated tags created in another crowdsourcing game.
4. *Professional quality control*. Eveleigh (2012) points out that curators play the role of gatekeepers when user-generated content is integrated into collections. While professional quality control has led in some cases to allegations of censorship (Eveleigh, 2012), most users accept the organisation's decisions as guided by professional expertise.

Many projects combine several different approaches to meet their specific needs, in what Eveleigh (2012) has described as a "*metadata-processing assembly line*" involving several steps in a "*hierarchical chain of command*" to arrive at valid high-quality data (ibid).

Professional quality control is usually the final step in this process.

The general design of TMW provides opportunities to incorporate most of these measures for increasing rigor and verifying the validity of contributed information. While the project cannot make the tasks easier, it can provide training and make verification part of the game. It can also develop suitable workflows that ensure professional quality control when integrating user-generated content into the collection.

## 7 Intellectual property

### 7.1 Copyright

In the context of curation, intellectual property primarily refers to copyright, which regulates the relationships between content owners and non-owners. While much of the popular discussion about copyright focuses on the challenge of enforcing copyright in a networked, digital environment where the ease of copying content contrasts with the difficulties of identifying and prosecuting copyright infringement, Elkin-Koren (2011) points out a more fundamental challenge related to the social production of content in crowdsourcing.

At the core of this challenge is the fact that current copyright is primarily designed to safeguard the rights of a single, sovereign author/owner, whereas social production involves multiple authors whose individual contributions depend on each other, vary in quality and quantity and often cannot be clearly attributed. This adds an additional layer of complexity to current copyright because it "*requires us to articulate a matrix of relationships between the individual, the facilitating platform and the communities and crowds involved in social production*" (Elkin-Koren, 2011).

To address these issues, Elkin-Koren (2011) suggests that copyright should be amended to better support social production models by:

- regulating credit and attribution, which have been identified as key motivators in social production and "*foundational to the modern economy*" (Fisk, 2006).
- establishing clear criteria defining commercial and non-commercial use and providing clear mechanisms for switching between the two in order to support the effective exploitation of the products of social production.
- balancing the interests of the individual with the interest of others and the group as a whole. Current copyright enables individuals to claim their contributions and thereby pull the rug from under others' and the community's efforts.

While changes to copyright law might take a long time to implement, the discussion of its shortcomings in the context of social production identifies critical aspects that need explicit attention in crowdsourcing projects.

### 7.2 Alternative frameworks

Where current copyright fails to adequately address key aspects of editing, transforming, managing, transferring and monetising the outputs of social production (Elkin-Koren, 2011), it leaves the door open for alternative frameworks such as the Gnu Public Licence (GPL)<sup>22</sup> or Creative Commons<sup>23</sup>. While such "*private ordering arrangements*" (Elkin-Koren, 2011) have the advantage that they can be tailored to communities' specific needs, a key disadvantage is that they require explicit consent from participants in the community to be enforceable. Copyright, by contrast, is an "*automatic right*", meaning that content creators do not need to register works for copyright to apply (IPO, 2013b).

---

<sup>22</sup> GNU General Public License. <http://opensource.org/licenses/gpl-license>

<sup>23</sup> Creative Commons. <http://creativecommons.org/>

A related disadvantage is that private licensing arrangements are difficult to change once they are in place as changes, again, require consent from every participant in the community to come into force (Elkin-Koren, 2011). Well-documented examples in this context include the transition from GPL 2.0 to GPL 3.0 and Wikipedia's<sup>24</sup> migration from GFDL<sup>25</sup> to Creative Commons<sup>26</sup>, both of which involved lengthy consultations before changes took effect.

The fact that alternative legal frameworks for social production require informed consent from participants in order to be enforceable is a critical aspect for TMW. As obtaining consent involves focusing participants' attention on legal issues, this creates an obvious design tension with regard to appealing to volunteers' altruistic motivations and making participation fun (see 4.1 *Motivation*).

### 7.3 Current practice

There is a general tendency in the Cultural Heritage sector towards making resources openly available to maximise their public utility. The *NMC Horizon Report: 2012 Museum Edition* makes a strong case for free and open content, stating that "*it is now the mark - and social responsibility - of world-class institutions to develop and share free cultural and educational resources*" (Johnson et al., 2012). The report explicitly mentions Creative Commons as a framework for museums to license open content. However, the discussion focuses mainly on the museum's copyright relating to content in their collections and does not extend to the intricacies of regulating social production in a crowdsourcing context.

Surprisingly, reports and reviews of crowdsourcing projects rarely mention copyright and licensing issues, and the few that do (e.g. Haklay and Weber, 2008; Oomen et al. 2010) do not provide much detail on what aspects and outputs in projects are covered by licensing agreements or how consent was obtained from participants.

The main outputs of TMW can be grouped into user-generated content, the developed software platform and documents such as reports, templates and best-practice guidelines:

- with respect to user-generated content, some reports explicitly mention the use of Creative Commons licenses (Haklay and Weber, 2008; Oomen et al. 2010), while some others implicitly adopt a licensing framework through their choice of platform, e.g. Flickr Commons<sup>27</sup> has been used by a range of crowdsourcing initiatives (Oomen et al., 2010; Dunn and Hedges, 2012; Hagon, 2013).
- with respect to software development, there is a rich variety of licensing models available (for an overview, see Rosen, 2004). While it is common practice in publicly funded research to release software under open source terms (e.g. JISC, 2012; RCUK, 2013), research reports rarely mention such details. One exception in this context is Trant (2009) who points out that software created in *steve.museum*<sup>28</sup> is released under a GPL license.

---

<sup>24</sup> Wikipedia. <http://www.wikipedia.org>

<sup>25</sup> GFDL is an acronym for the Gnu Free Documentation License. <http://www.gnu.org/copyleft/fdl.html>

<sup>26</sup> Creative Commons. <http://creativecommons.org/>

<sup>27</sup> Flickr Commons. <http://www.flickr.com/commons>

<sup>28</sup> *steve.museum*. <http://www.steve.museum/>

- with respect to other outputs, such as reports, support materials and best-practice guidelines, no instances were found in the literature discussing licensing terms. However, given that many projects are publicly funded it can be assumed that their outputs are released under open licenses.

The lack of detailed information in the literature about the licensing terms and mechanisms in crowdsourcing projects seems to justify the research aim in TMW to develop a licensing framework that can be reused in similar crowdsourcing projects in the UK Cultural Heritage sector.

## **8 Summary and conclusions**

Drawing on a wide range of literature on crowdsourcing and related concepts, this review has situated TMW in a wider research context and discussed how it can inform relevant design aspects in the project concerning sustained volunteer engagement, game elements and reward structures, data quality and intellectual property.

The review started by clarifying relevant terminology. It contrasted crowdsourcing with peer production and citizen science, finding substantial overlaps between these concepts. It then synthesised a working definition of games from the literature, which provided a useful foundation for the discussion of game elements and rewards, and identified problematic aspects of current copyright, which framed the discussion of IP issues in the context of social production.

In order to situate TMW in a wider research context, the review related the project to various classification systems for crowdsourcing projects. It found that TMW marks new territory in particular with respect to Ridge's (2012) activity categories for game-based crowdsourcing and Dunn and Hedges (2012) crowdsourcing typology as it examines crowdsourcing games in a HWPP context as opposed to previous projects which primarily involved LWPP patterns (Haythornthwaite, 2009).

The review looked in detail at engagement as a critical aspect concerning the sustainability of crowdsourcing projects, focusing in particular on volunteers' motivations and measures to encourage initial engagement and sustain engagement over longer periods. It found that volunteers have a range of intrinsic, altruistic and extrinsic motivations that can be leveraged in TMW and that it is important to make participation fun and convey this aspect to current and potential volunteers. It discussed the benefits of a broad dissemination strategy involving a wide range of activities and materials, and in particular singled out the role of broadcast media in attracting new volunteers. With respect to sustaining engagement, the review discussed a wide range of aspects ranging from communication to rewards structures, attribution, governance and training which can help to keep participants interested and actively engaged in the project.

With respect to game elements and reward structures, the review discussed differences between casual games and serious games, and found qualitative reward structures to be more suitable in the context of HWPP models of crowdsourcing. It described a wide range of

qualitative rewards mentioned in research reports of other crowdsourcing projects. The review found that many of these rewards are relevant to TMW and can be either implemented directly or adapted to its specific context.

Exploring issues around data quality and verification, the review found deep-seated fears among Heritage professionals with regard to incorporating user-generated content into professionally curated collections. Looking towards addressing these fears, it identified a range of measures to increase data quality, including training of volunteers, crowd-sourcing verification and developing workflows that ensure professional quality control when integrating user-generated content into the collection.

The review finally looked at Intellectual Property issues, exploring conflicts between current copyright and social production and looking at alternative frameworks that address these conflicts. It discussed the relationships between copyright, attribution and motivation in social production and uncovered design tensions when obtaining informed consent from participants while at the same time appealing to their altruistic motivations and making participation seem fun.

The literature review has provided insights into a number of critical aspects of TMW, including the viability of using serious, complex, collaborative games for crowdsourcing, the challenge of sustaining participant engagement over prolonged periods, and unresolved questions around data quality and intellectual property. While it has not provided clear answers to these challenges, it has helped to refine research questions and formulate a set of recommendations to inform the design of TMW (see below *Recommendations for TMW*).

## **9 Acknowledgements**

Supported by the Digital R&D Fund for the Arts - Nesta, Arts & Humanities Research Council and public funding by the National Lottery through Arts Council England.

## References

- Abt, C. (1970). *Serious Games*. New York: The Viking Press.
- Benkler, Y. (2000). From Consumers to Users: Shifting the Deeper Structures of Regulation Towards Sustainable Commons and User Access. *Yale Law Journal*, 52, 561-579.
- Benkler, Y. (2002). Coase's Penguin, or, Linux and The Nature of the Firm. *Yale Law Journal*, 112, 369-446.
- Benkler, Y. (2004). Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production. *Yale Law Journal*, 114, 273-278.
- Bernstein, S. (2009). Crowdsourcing the Clean-Up with Freeze Tag! Brooklyn Museum. Available <http://www.brooklynmuseum.org/community/blogosphere/2009/05/21/crowdsourcing-the-clean-up-with-freeze-tag/>
- Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1), 75-90. doi:10.1177/1354856507084420
- Bradford, B.M. and Israel, G.D. (2004). Evaluating Volunteer Motivation for Sea Turtle Conservation in Florida. Technical Report AEC 372. Institute of Agricultural and Food Sciences, University of Florida. Available <http://ufdc.ufl.edu/IR00002485/00001>
- Brooklyn Museum (2009). Tag It! You're it. Available [http://www.brooklynmuseum.org/opencollection/tag\\_game/start.php](http://www.brooklynmuseum.org/opencollection/tag_game/start.php)
- Budhathoki, N. R. and Haythornthwaite, C. (2012). Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist*, 57(5), 548-575. doi:10.1177/0002764212469364
- Causser, T. and Wallace, V. (2012). Building A Volunteer Community : Results and Findings from Transcribe Bentham. *Digital Humanities Quarterly*, 6(2).
- Cohn, J.P. (2008). Citizen Science: Can Volunteers Do Real Research? *BioScience*, 58(3), 192-197.
- Costikyan, G. (2002). I Have No Words & I Must Design. *Proceedings of Computer Games and Digital Cultures*, Tampere University Press, pp. 9-33.
- Crawford, C. (2003). *Chris Crawford on Game Design*. Thousand Oaks, CA: New Riders Publishing.
- Creative Commons (2013). *The Licenses*. Available <http://creativecommons.org/licenses/>
- Deci, E. (1975). *Intrinsic motivation*. New York: Plenum Press.
- Dempsey, J.V., Lucassen, B.A., Haynes, L.L. and Casey, M.S. (1996). Instructional applications of computer games. Paper presented to the American Educational Research Association, 8-12 April 1996, New York. ERIC Document Reproduction Service No. ED 394 500.
- Dunn, S. and Hedges, M. (2012). *Engaging the Crowd with Humanities. A scoping study*. Research Centre for e-Research , Department of Digital Humanities. King's College London.

Available <http://stuardunn.files.wordpress.com/2013/04/crowdsourcing-connected-communities.pdf>

- Dunn, S. (2013). More than a business model: crowdsourcing and impact in the humanities. Blogpost on Impact of Social Sciences, London School of Economics. Available <http://blogs.lse.ac.uk/impactofsocialsciences/2013/03/21/more-than-a-business-model-crowdsourcing-and-impact-in-the-humanities/>
- Elkin-Koren, N. (2011). Tailoring copyright to social production. *Theoretical Inquiries in Law*, 12(1).
- Estelles-Arolas, E. and Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
- Eveleigh, A. (2012). Welcoming the World: An Exploration of Participatory Archives. Presented at International Council on Archives (ICA) Conference, ICA 2012, Brisbane, Australia, 20-24 August 2012 (pp. 1–10).
- Fabricatore, C. (2000). Learning and videogames: an unexploited synergy. [online] [www.learndev.org/dl/FabricatoreAECT2000.pdf](http://www.learndev.org/dl/FabricatoreAECT2000.pdf). [retrieved] 3 June 2009.
- Fisk, C. L. (2006). Credit Where It's Due: The Law and Norms of Attribution. *Geo. LJ*, 95, 49.
- Frasca, G. (2003). Ludologists love stories, too: notes from a debate that never took place. *Proceedings of International DiGRA Conference* (pp. 1–8).
- Grove-White, R., Waterton, C., Ellis, R., Vogel, J., Stevens, G. & Peacock, B. (2007). Amateurs as experts: harnessing new networks for biodiversity. Lancaster University, Lancaster. Available <http://csec.lancs.ac.uk/docs/Amateurs%20as%20Experts%20Final%20Report.pdf>
- Hagon, P. (2013). Trove crowdsourcing behaviour. National Library of Australia. Available [http://www.information-online.com.au/pdf/Tuesday\\_Concurrent\\_2\\_1125\\_Hagon.pdf](http://www.information-online.com.au/pdf/Tuesday_Concurrent_2_1125_Hagon.pdf)
- Haklay, M. and Weber, P. (2008). Open Street Map: User-generated street maps. *Pervasive Computing, IEEE*, 2008 (October–December), 12–18.
- Hars, A. and Ou, S. (2001). Working for free? Motivations of participating in open source projects. *International Journal of Electronic Commerce*, 6(3), 25-39
- Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. *Proceedings of the Hawaii International Conference On System Sciences* (pp. 1–11)
- Hinske, S., Lampe, M., Magerkurth, C., and Röcker, C. (2007). Classifying Pervasive Games: On Pervasive Computing and Mixed Reality. In Magerkurth, C., Röcker, C. (eds). *Concepts and technologies for Pervasive Games - A Reader for Pervasive Gaming Research*. Vol. 1, Shaker.
- Holley, R. (2009). Many Hands Make Light Work : Public Collaborative OCR Text Correction in Australian Historic Newspapers. National Library of Australia. Retrieved from: <http://www.nla.gov.au/content/many-hands-make-light-work-public-collaborative-ocr-text-correction-in-australian-historic>.
- Holley, R. (2010). Crowdsourcing : How and why should libraries do it ? *DLib*, (March/April), 1–21.

- Howe, J. (2006a). The Rise of Crowdsourcing. *Wired*, 14(6), (June 2006). Available <http://www.wired.com/wired/archive/14.06/crowds.html>
- Howe, J. (2006b). Crowdsourcing: A Definition. Blog post from June 02, 2006. Available [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html)
- Huizinga, J. (1950). *Homo Ludens: A Study of the Play Element in Culture*. Boston: The Beacon Press.
- IPO (2013a). About copyright. UK Intellectual Property Office. Available <http://www.ipo.gov.uk/types/copy/c-about.htm>
- IPO (2013b). Automatic right. UK Intellectual Property Office. Available <http://www.ipo.gov.uk/types/copy/c-about/c-auto.htm>
- JISC (2012). Policy on open source software for Jisc projects and services. Available <http://www.jisc.ac.uk/fundingopportunities/opensourcepolicy.aspx>
- Johnson, L., Adams Becker, S., Witchey, H., Cummins, M., Estrada V., Freeman, A., and Ludgate, H., (2012). *The NMC Horizon Report: 2012 Museum Edition*. Austin, Texas: The New Media Consortium.
- Juul, J. (2003). The Game, the Player, the World: Looking for a Heart of Gameness. In Copier, M. and Raessens, J. (eds.) *Level Up: Digital Games Research Conference Proceedings*. Utrecht: Universiteit Utrecht, 2003, pp. 30-45.
- Kidd, J. (2013). Visitor Generated Content (VGC) and Ethics - what we might learn from the media and journalism. Presentation at the iSay: State of Things conference, 31 Jan-2Feb, Leicester (pp. 1–8).
- Klabbers, J. (2003). The gaming landscape: a taxonomy for classifying games and simulations. In Copier, M. and Raessens, J. (eds) *Level Up: Digital Games Research Conference Proceedings*. Utrecht: Universiteit Utrecht, 2003, pp. 54-67.
- Lave, J. (1988). *Cognition in Practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.
- Lindley, C. A. (2002). The gameplay Gestalt, narrative and interactive storytelling. *Proceedings of Computer Games and Digital Cultures*, Tampere University Press, pp. 1-11.
- Malone, T.M. (1980). What makes things fun to learn? Heuristics for designing instructional computer games. *Proceedings of the Third ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems*, Palo Alto, CA, Sept. 18–19, 1980, 162–169
- Mason, W. and Watts, D. (2010). Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2), 100–108.
- Mendez, B. J. H. (2008). SpaceScience@Home: Authentic Research Projects that Use Citizen Scientists. In C. Garmony, M. G. Gibbs, and J. W. Moody (Eds) *EPO and a Changing World: Creating Linkages and Expanding Partnerships*, ASP Conference Series, 385, San Francisco.
- Nissenbaum, H. (2005). Values in technical design. In C. Mitcham (ed) *Encyclopedia of Science, Technology and Society*, MacMillan, New York.
- Norman, D. (1988). *The Design of Everyday Things*. MIT Press.

- Nov, O., Arazy, O. and Anderson, A. (2011). Dusting for science: motivation and participation of digital citizen science volunteers. In Proceedings of the 2011 iConference (iConference '11). ACM, New York, NY, USA, 68-74.
- O'Brien, L., Townsend, M. and Ebdon, M. (2008). Environmental volunteering: motivations, barriers and benefits. Report to the Scottish Forestry Trust and Forestry Commission. Available [http://www.forestry.gov.uk/pdf/Env\\_Volunteering\\_Full\\_Report.pdf](http://www.forestry.gov.uk/pdf/Env_Volunteering_Full_Report.pdf)
- Oomen, J. and Brinkerink, M. (2010). Emerging Institutional Practices: Reflections on Crowdsourcing and Collaborative Storytelling. Proceedings of Museum and the Web 2010. Retrieved from:  
<http://www.museumsandtheweb.com/mw2010/papers/oomen/oomen.html>
- Oomen, J. and Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. Proceedings of the 5th International Conference on Communities and Technologies (pp. 138–149).
- Orwant, J. (2000). EGG: The Extensible Graphical Game Generator. PhD Thesis, MIT. [online] <http://www.orwant.com/eggg.html>. [retrieved] 18 June 2009.
- Parlett, D. (1999). The Oxford History of Board Games. New York: Oxford University Press.
- Peddibhotla, N.B. and Subramani, M.R. (2007). Contributing to Public Document Repositories: A Critical Mass Theory Perspective. Organization Studies, 28(3). Sage Publications.
- Prensky, M. (2001). Digital game-based learning. New York: McGraw-Hill.
- Prestopnik, N. R. and Crowston K. (2011). Gaming for (citizen) science: Exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. Proceedings of the IEEE eScience Conference, Stockholm 2011, pp. 28-33. Available  
[http://crowston.syr.edu/sites/crowston.syr.edu/files/gamingforcitizenscience\\_ver6.pdf](http://crowston.syr.edu/sites/crowston.syr.edu/files/gamingforcitizenscience_ver6.pdf)
- Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S. and Vandenberg, J. (2010). Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. Astronomy Education Review, 9(1), 1-18.
- Rafaeli, S., Raban, D. R. and Ravid, G. (2007). How social motivation enhances economic activity and incentives in the Google Answers knowledge sharing market. International Journal of Knowledge and Learning, 3, 1–11
- Rafaeli, S., and Ariel, Y. (2008). Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia. In A. Barak (Ed.), Psychological Aspects of Cyberspace: Theory, Research, Applications. (pp. 243–267). Cambridge, UK: Cambridge University Press.
- RCUK (2013). Guidance for the RCUK Policy on Access to Research Outputs. Available  
[http://www.rcuk.ac.uk/documents/documents/Guidance\\_for\\_the\\_RCUK\\_policy\\_on\\_Access\\_to\\_Research\\_Output.pdf](http://www.rcuk.ac.uk/documents/documents/Guidance_for_the_RCUK_policy_on_Access_to_Research_Output.pdf)
- Ridge, M. (2011a). Playing with Difficult Objects – Game Designs to Improve Museum Collections. In J. Trant & D. Bearman (Eds.), Proceedings of Museums and the Web 2011. Retrieved from:  
[http://www.museumsandtheweb.com/mw2011/papers/playing\\_with\\_difficult\\_objects\\_game\\_designs\\_to](http://www.museumsandtheweb.com/mw2011/papers/playing_with_difficult_objects_game_designs_to)

- Ridge, M. (2011b). Playing with difficult objects: game designs for crowdsourcing museum metadata. MSc Dissertation, City University London, School of Informatics. Available [http://www.miaridge.com/wp-content/uploads/2013/01/MiaRidge\\_MSc\\_Dissertation\\_OpenAccess.pdf](http://www.miaridge.com/wp-content/uploads/2013/01/MiaRidge_MSc_Dissertation_OpenAccess.pdf)
- Ridge, M. (2012). Frequently Asked Questions about crowdsourcing in cultural heritage. Retrieved from <http://openobjects.blogspot.co.uk/2012/06/frequently-asked-questions-about.html>
- Rogoff, B. (1982). Integrating context and cognitive development. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol 2). Hillsdale NJ: Erlbaum.
- Romeo, F. and Blaser, L. (2013). Bringing Citizen Scientists and Historians Together. In J. Trant & D. Bearman (Eds.), *Proceedings of Museums and the Web 2011*. Retrieved from [http://conference.archimuse.com/mw2011/papers/bringing\\_citizen\\_scientists\\_historians\\_together](http://conference.archimuse.com/mw2011/papers/bringing_citizen_scientists_historians_together)
- Rosen, L. (2004). *Open source licensing*. Prentice Hall PTR.
- Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. and Robinson, L.D. (2012). *Understanding Citizen Science & Environmental Monitoring*. Final Report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum. Retrieved from: <http://www.ukeof.org.uk/documents/understanding-citizen-science.pdf>
- Salen, K. and Zimmerman, E. (2003). *Rules of play: the fundamentals of game design*. Cambridge MA: MIT Press.
- Shirky, C. (2009). *Here Comes Everybody: How change happens when people come together*. Penguin UK.
- Trant, J. (2009). Tagging, Folksonomy and Art Museums: Results of steve.museum ' s research. *Archives & Museum Informatics*. Retrieved from: [http://conference.archimuse.com/jtrants/stevemuseum\\_research\\_report\\_available](http://conference.archimuse.com/jtrants/stevemuseum_research_report_available)
- Von Ahn, L. (2006). Games with a Purpose. *Computer*, June 2006, pp. 96-98.
- Von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, 51(8), 58-67.
- Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.
- Wiggins, A. and Crowston, K. (2011). From Conservation to Crowdsourcing: A Typology of Citizen Science. 2011 44th Hawaii International Conference on System Sciences (pp. 1–10). Ieee. doi:10.1109/HICSS.2011.207
- Yuen, M.-C., King, I. and Leung, K.-S. (2011). A Survey of Crowdsourcing Systems. *Proceedings of IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and IEEE Third Int'l Conference on Social Computing*, 766–773.



## **Recommendations for TMW**

The following design recommendations are based on findings of the above literature review. While they are formulated with reference to the TMW project, most of the recommendations will equally apply to other game-based crowdsourcing projects using HWPP\* patterns involving deep, sustained engagement (as opposed to LWPP\* patterns involving casual engagement).

---

\* For a discussion of Light and heavyweight models of peer production (LWPP/HWPP) see Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. Proc. Hawaii International Conference On System Sciences.

### **Tapping into volunteers' motivations**

1. Appeal to intrinsic motivations, e.g. a genuine interest in design or plastic
2. Appeal to altruistic motivations, e.g. support MoDiP as a small organisation
3. Appeal to extrinsic motivations, e.g. supportive community, incentives, rewards
4. Make participation fun for participants and communicate that fun to others

### **Encouraging initial engagement**

5. Write press releases and get broadcast media interested in the project
6. Develop a broad range of dissemination materials and activities:
  - a. project website
  - b. presence in a various social media outlets
  - c. mailing lists and online forums
  - d. leaflets, posters, stickers, badges
  - e. presentations at seminars, conferences and workshops
  - f. educational resources of relevance to the curriculum
  - g. contact individual museums, archives, and educational bodies
  - h. organise and publicise site visits by actual and potential volunteers

### **Sustaining participant engagement**

7. Support communication
  - a. between project team and volunteers
  - b. between participants themselves as they shape their community
8. Use social media to convey activity in the project to outsiders
9. Organise face-to-face meetings where volunteers meet the project team and each other
10. Establish lasting connections between participants and the data they contribute, e.g. by showing finds on a map and linking to participants' profiles
11. Respect participants' views on project governance, tools and practices and provide ways for co-determination, e.g. through meetings and co-design sessions
12. Recognise participants' contributions through attribution and credits in publications
13. Provide training opportunities for participants to develop their skills e.g. workshops and online training materials

### **Serious game elements**

14. Select a number of objects to focus on
15. Match tasks closely to specific objects
16. Provide a brief explaining tasks in a call to action
17. Make available results from previous challenges to provide examples
18. Provide qualitative rewards that enhance participants' profile, e.g.
  - a. access to exclusive content
  - b. badges for specific achievements
  - c. publication and attribution
  - d. assignment of roles and privileges
  - e. public progress visualisations
  - f. personal thank you notes
19. Frame guidelines, procedures and acceptable quality standards as rules of the game
20. Use language consistent with the TMW theme to create an overall game narrative
21. Report on the outcomes of participants' efforts and give timely feedback

### **Data quality**

22. Be open about quality standards and processes
23. Provide training opportunities for participants
24. Reassure participants that their contributions are useful and academically viable
25. Involve participants in verification and validation (e.g. as part of the task/game)
26. Develop suitable workflows that ensure professional quality control when integrating user-generated content into the collection

### **Intellectual property**

27. Establish clear criteria for
  - a. credit and attribution
  - b. commercial and non-commercial use
  - c. responsibilities of individuals to each other and the community
28. Adopt a licensing framework for contributed content that meets participants' expectations and is compatible with MoDiP's own licensing terms
29. Design a suitable mechanism for participants to agree to the licensing framework without unduly impacting on altruistic motivations and the fun aspect of TMW
30. Open-source the developed crowdsourcing platform
31. Release other outputs, such as reports, support materials and guidelines under a suitable Creative Commons license

---