

# Crowd Sourcing: a Toolkit-based Approach

Chris Greenhalgh, Richard Mortier,  
Robert Houghton, Catherine Mulligan,  
Alexa Spence

Horizon Digital Economy Research,  
The University of Nottingham  
Sir Colin Campbell Building  
University of Nottingham Innovation Park  
Nottingham NG7 2TU, UK

*firstname.lastname@nottingham.ac.uk*

Ronald Carter, Svenja Adolphs  
School of English Studies,  
The University of Nottingham  
University Park  
Nottingham NG7 2RD, UK

*firstname.lastname@nottingham.ac.uk*

## ABSTRACT

We propose a software toolkit – initially comprising cloud-hosted services with web APIs and web and mobile clients – to support a range of crowd-sourcing activities based on the provision of information. As well as supporting essentially stand-alone activities, this toolkit will have the option of linking to the Personal Container(s) being developed within the Horizon Hub, which are a software infrastructure for managing and utilizing an individual's personal data. The link to personal containers allows crowd-sourcing of already-collected data (such as information from my personal journey logs, store transactions or medical history). Initial areas of application include: journey data; personal and community history; and contextually-appropriate use of spoken English for non-native speakers. As well as technical challenges of ease of use, scalability, privacy-preservation and filtering, work on crowd-sourcing also unavoidably opens up questions of motivation, reward, intellectual property, safety and policy.

## Keywords

Crowd-sourcing.

## 1. INTRODUCTION

Crowdsourcing is an extremely topical idea with a very diverse range of applications (e.g. [1]). Wikipedia, the crowdsourced online encyclopedia, defines crowdsourcing as:

*a neologistic compound of "crowd" and "outsourcing" for the act of outsourcing tasks, traditionally performed by an employee or contractor to a large group of people or community (a crowd), through an open call.*  
<http://en.wikipedia.org/wiki/Crowdsourcing> 2010-04-15

Although defined with roots in business (specifically outsourcing) the term is used more generally, for example where there is no outsourcing organization as such or where the task being crowdsourced would not otherwise have been undertaken. For example, Wikipedia would not exist except as a community effort.

Horizon's work on Personal Container [2] (aka Personal Data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Digital Futures'10*, October 11 – 12, 2010, Nottingham, UK

Store) provides a fresh perspective on crowdsourcing and a range of new opportunities. First, it supports a broad spectrum of scenarios in which individuals can contribute carefully managed fragments of the information known to their (distributed, federated) Personal Container in response to a crowdsourcing call. For example, my journey information, purchasing decisions, energy consumption data and entertainment reviews might all be routinely available through my Personal Container for my own benefit, but may also be of benefit to different crowdsourcing-based applications and services. Second, the same concepts and technical facilities for data management and controlled use are also applicable to situations involving new and explicit crowdsourcing tasks and activities. For example, I may contribute to a citizen science activity by reporting sightings of particular wildlife. Doing this in concert with my Personal Container would allow me to use a set of then-familiar tools and concepts to trace and manage my participation. This might also lower the barriers to initial participation in other activities.

Some of the fundamental challenges of crowdsourcing are: identifying the tasks and challenges for which crowdsourcing is an appropriate solution; defining, operating, supporting and ending a crowdsourcing activity; identifying and creating technical means of participation that minimize barriers to use; establishing and maintaining participation through appropriate incentives; ensuring appropriate privacy and safety for the contributors (e.g. when individual contributors might be identifiable and/or locatable); maximizing the quality and benefit of the outcome (e.g. example through filtering, rating, cross-checking or peer or expert moderation).

## 2. APPROACH

The focus of this first phase (one year) of work is primarily technical, i.e. the development and initial application of a prototype crowdsourcing toolkit. The scope of crowdsourcing tasks is limited to information gathering, but will explicitly include test cases of both implicit tasks (such as collecting journey traces) and explicit tasks (such as rating or proposing phrases for a phrase-book). This project will provide a resource to support and inform research in crowdsourcing in the human and innovation challenges within Horizon.

The toolkit logically comprises the software elements required to define, establish, manage, participate in and end a crowdsourcing activity. It supports four main user perspectives:

- The initiator or commissioner of the crowdsourcing activity, who will specify the task(s) to be performed (in this case, the

information desired) and oversee the use of contributions (e.g. derivation and exploitation of contributions).

- The operator or provider of the crowdsourcing facility (who may or may not be the same as the initiator), who will deploy and manage any necessary resources (e.g. services, applications), help to recruit and support participants, and manage the activity as a whole.
- The contributor or crowd-member, who will (implicitly or explicitly) provide the information, as well as joining and leaving the activity (including giving consent) and may also monitor and change their own involvement and the way their contribution(s) are used.
- The consumer or beneficiary of (the information from) the crowdsourcing activity, who may also be the commissioner, a contributor or another party entirely.

Architecturally, the heart of crowdsourcing activity will be a cloud-based service (e.g. hosted on Google App Engine or Amazon EC2) which can host multiple concurrent crowdsourcing activities and which supports the initiator and operator interfaces, and links to personal containers as appropriate. The contributor and consumer interfaces in particular will be defined as web APIs, facilitating the integration of the service into other applications and services. Specific user interfaces may include: the standard personal container interface(s), application-specific web interfaces and/or application-specific thick clients (e.g. to allow off-line collection of task-specific information on mobile devices). The precise contributor and consumer interfaces developed within the project will depend on the emerging requirements from the application areas.

We believe that one of the main software challenges lie in striking a useful balance between very well-defined data structures, data transformation operations and user interfaces with relatively narrow applicability on the one hand, and very flexible and/or extensible (complex) data structures, transformations and user interfaces with broad applicability on the other hand. Scalability in terms of numbers of contributors and number/size of contributions is also a fundamental challenge, requiring careful consideration of the underlying data management, computation and communication facilities.

The toolkit will be developed iteratively, working with the application projects (see below). All source will be released under the AGPL v3 Open Source license (a strong copy-left license), with the option of simultaneous licensing under other licenses. Initial mobile client support will target the Android platform, while Google App Engine is the default cloud platform for those service elements that fit within its constraints.

### 3. APPLICATIONS

This project will link to crowdsourcing elements that exist and have emerged within other Horizon projects, in particular: the collection of journey information within the car sharing/socially connected journey project; and the collection of personal and community history within the community history, pervasive monuments and urban games projects.

In addition the toolkit will support a further complementary pilot application in the domain of context-specific English Language usage and learning (as a second language). This will explore the potential of crowdsourcing in relation to: gathering information on contextual usage of English in various settings and locations (e.g. in a grocers vs a bank vs a train station); and presenting, filtering and reviewing suggested phrases and words (to non-native speakers) in such settings and locations. As an illustrative use of this kind of data, imagine a non-native English speaking visitor going into a bank in Nottingham. The electronic phrase book application on their mobile device determines that they are in a bank (e.g. by reading a visual glyph, RFID tag or IR beacon) and automatically sorts the words and phrases that it presents according to how likely they are to be useful in this particular context (based on previously contributed ratings and feedback).

We already envisage several possible uses of crowdsourcing to support such an application. The various locations (banks, shops, etc.) may accept the crowdsourced task of identifying themselves and suggested words and phrases that are typical or appropriate in their setting (they have a vested interest in supporting visitors in this way). Local citizens or other visitors may accept the task of recording the language that they use or would use in such settings. Users of the application may also return ratings and other annotations of the phrases they actually use (or choose to avoid), or may simply allow the application to monitor their interaction much as Google monitors user's click-through behaviours to determine link rankings in search results. This information might also be mined for patterns of usage to support the preparation of additional teaching and learning resources.

### 4. STATUS AND NEXT STEPS

An initial workshop has been held to begin to identify research and experimental approaches to tackling the psychology, innovation and human factors challenges around the proposed work. The next phase of activity involves the detailed characterizing of the initial applications, and in particular the task(s) to be crowdsourced, the specific information to be gathered, how it will be filtered, how it will be combined and how the results will be accessed and portrayed. Initial interfaces to the Personal Container will be developed as part of the concurrent development of the next Personal Container prototype. The first release of the crowd-sourcing toolkit is expected in early 2011. This will be available for other projects and activities to make use of (and potentially contribute to!).

### 5. ACKNOWLEDGEMENTS

We gratefully acknowledge the support of RCUK through grant EP/G065802/1 Horizon Digital Economy Research.

### 6. REFERENCES

- [1] Daren C. Brabham. (2008). "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", *Convergence: The International Journal of Research into New Media Technologies*, 14(1), pp. 75-90.
- [2] Richard Mortier, Chris Greenhalgh, Derek McAuley et al., "The Personal Container, or Your Life in Bits", *Digital Futures '10*, October 11–12, 2010, Nottingham, UK