

# StarBorn: Towards making in-situ land cover data generation fun with a location-based game

Manuel F. Baer<sup>1,4</sup>  | Flurina M. Wartmann<sup>2,3</sup> | Ross S. Purves<sup>1,4</sup>

<sup>1</sup>Department of Geography, University of Zurich, Zurich, Switzerland

<sup>2</sup>School of Geosciences, Institute of Geography, University of Edinburgh, Edinburgh, United Kingdom

<sup>3</sup>Swiss Federal Research Institute for Forest, Snow and Landscape, WSL, Birmensdorf, Switzerland

<sup>4</sup>University Research Priority Program: Language and Space, University of Zurich, Zurich, Switzerland

## Correspondence

Manuel F. Baer, Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.  
Email: manuel.baer@uzh.ch

## Funding information

University Research Priority Program: Language and Space, University of Zurich

## Abstract

Data contributed by a large number of non-experts is increasingly used to validate and curate land cover data, with location-based games (LBGs) developed for this purpose generating particular interest. We here present our findings on StarBorn, a novel LBG with a strong focus on game play. Users conquer game-tiles by visiting real-world locations and collecting land cover data. Within three months, StarBorn generated 13,319 land cover classifications by 84 users. Results show that data are concentrated around users' daily life spaces, agreement among users is highest for *urban* and *industry* land cover, and user-generated land cover classifications exhibit high agreement with an authoritative data set. However, we also observe low user retention rates and negative correlations between number of contributions and agreement rates with an authoritative land cover product. We recommend that future work consider not only game play, but also how motivational aspects influence behavior and data quality. We conclude that LBGs are suitable tools for generating cost-efficient in-situ land cover classifications.

## 1 | INTRODUCTION

Land cover products derived from remote sensing data are increasingly important in the analysis of diverse processes in the earth system (Congalton, Gu, Yadav, Thenkabail, & Ozdogan, 2014), ranging from monitoring biodiversity (Turner et al., 2003) and tropical deforestation detection (Achard et al., 2002) to gauging the extent

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Transactions in GIS Published by John Wiley & Sons Ltd

of urban sprawl (Sahana, Hong, & Sajjad, 2018). They are thus crucial in policy- and decision-making processes (Foody et al., 2013; Lambin et al., 2001; Sexton et al., 2015). However, between-product variability is often high when comparing different land cover products at given locations on the Earth's surface (See et al., 2013; Sexton et al., 2015). Seemingly trivial questions such as "How much forest is there?" metamorphose into complex discussions on data quality, sensors, semantics and underlying biases (Burenhult et al., 2017; Côte, Wartmann, & Purves, 2018; Sexton et al., 2015). Omnipresent in such discussions is the need for appropriate ways of assessing the quality of land cover products. Quality assessment is most often based on a gold standard validation data set, often in the form of an independently labeled data set. Such gold standards (or *ground truth*) typically take one of two forms. *Ex-situ* assessments of land cover classifications (cf. Fritz et al., 2009; See et al., 2013) are performed remotely, typically by independent assessors who label pixels using satellite imagery and an underlying classification schema. For *in-situ* assessments assessors visit defined points in a landscape and assign them to a land cover classification (cf. Bayas et al., 2016).

Ground truth data, in the form of either labeled images or in-situ measurements, are therefore a crucial input part of the process of validating land cover products (See et al., 2013). In-situ data are traditionally collected by experts or trained persons; however, these are associated with financial restrictions. For example, a maximum of 10% of the budget for the "Coordination of Information on the Environment" (CORINE) land cover product was set aside for ground truth surveys (European Environment Agency, 1994). This financial restriction leads to a major issue in land cover product validation: the scarcity of high-quality in-situ ground truth data for validating remotely sensed land cover products.

One possible approach to this challenge is to generate data sets for the validation of land cover products using citizen science and crowdsourcing. User-generated content (UGC), such as volunteered geographic information (VGI), has gained traction in validation and calibration efforts and has been identified as a potential source of ground truth data (Foody & Boyd, 2012; Fritz et al., 2009; Goodchild, 2007; See et al., 2013). The growth of interest in research focusing on the use of VGI to generate data sets for land cover product validation is partially due to advances in internet and mobile technologies, allowing users to easily contribute and share geographic information on handheld devices. However, even though technological advances have provided scientists with new accessible tools for data generation, how to motivate users to contribute data remains an issue (Crall et al., 2017). Gamification has been identified as one approach to motivating users to contribute to scientific crowdsourcing efforts (Bubalo, Zanten, & Verburg, 2019; Fritz, See, & Brovelli, 2017; Hamari & Sarsa, 2014).

Gamification has been applied to a number of research endeavors (cf. Hamari & Sarsa, 2014), with various projects incorporating one to many game elements. However, we argue that simply adding arbitrary game elements (e.g., a single rankings page or a simple reputation system) to a crowdsourcing approach does not necessarily make using the application fun or enjoyable. Game elements should be carefully chosen and combined through an underlying narrative to create an entertaining experience.

Location-based games (LBGs) are of particular interest in this respect. Users visit real-world locations to contribute geographic information, making LBGs especially interesting for in-situ data collection. LBGs also allow game elements to be implemented to increase user motivation for specific collection tasks or contributions at specific locations. LBGs are usually played on a handheld, location-enabled device, acting as a bridge between a virtual game world and the physical world. A prime example of a widely known LBG is *Pokémon GO*, where players capture virtual monsters and compete over specific virtual locations while moving through the real world (Colley et al., 2017). Since these games allow specific interactions with a virtual game world depending on criteria based on real-world location, changing position becomes an integral part of game play. These underlying game mechanics can be used to collect data from users by harnessing the users' ability to sense their immediate surroundings (cf. Celino, 2015; Wang & Ben-Arie, 1996), thus allowing in-situ data generation. Various authors have argued that LBGs have considerable potential for collecting geographic data (Celino et al., 2012; Matyas, Kiefer, & Schlieder, 2012; Winter et al., 2011). However, user retention has been identified as a key issue in citizen science efforts in

general (cf. Crall et al., 2017), and a rapid decline in active users after an initial period of interest has been observed in LBGs (cf. Andone, Blaszkiewicz, Böhmer, & Markowetz, 2017).

A number of questions arise regarding the viability of LBGs to generate in-situ land cover data. Firstly, how can such a game be designed and implemented, such that players are motivated to contribute? Secondly, what intrinsic characteristics do the data thus collected have, and how are these influenced by the behaviors of the game's players? Finally, how are the data collected comparable with an authoritative land cover product, and are they likely to be of use as high-quality in-situ data for land cover validation?

## 2 | RELATED WORK

In this section we present related work from four key areas. First, we explore the current state of the art in efforts to validate land cover data, focusing on approaches using some form of crowdsourcing. Second, we introduce aspects of user motivation in crowdsourcing efforts. We then explore key aspects of LBGs, highlighting gamification elements that go beyond rankings of user contributions. Finally, we briefly discuss user characteristics and behaviors in the general context of user-generated content.

### 2.1 | Land cover product validation

Land cover validation revolves around comparing existing land cover products with ground truth data based on either ex-situ image analysis or in-situ land cover assessment.

Foody (2002) summarizes four historical stages of land cover assessment and validation. Perhaps the most commonly accepted method for comparing data sets is using a confusion matrix (Foody, 2002), where the reported land cover classes of each pixel from two different data sets are compared, allowing various metrics to be computed (Ji & Niu, 2014). Where one product is of a known higher quality (a gold standard) these metrics describe the quality of a second product; however, lack of gold standard data is a major limitation in land cover product validation efforts. In-situ ground truth data are commonly generated by paid and trained (semi-)experts as in the Land Use/Cover Area Frame Survey (LUCAS; Bayas et al., 2016), but the cost of ground truth surveys and the accessibility of crucial locations are often considered major constraints (Bayas et al., 2016; European Environment Agency, 1994; Foody, 2002).

One potential alternative source of ground truth data, which has been the subject of considerable attention in the scientific community, is UGC, or VGI. A large body of work has explored the use of data sources such as OpenStreetMap and Flickr to extract land cover data (e.g., Fonte & Martinho, 2017; Jeawak, Jones, & Schockaert, 2017). While these approaches relied on existing data as a form of passive crowdsourcing, our focus lies in the use of ground truth data collected through active crowdsourcing. Such approaches can effectively be divided into two broad groups, focusing on either *ex-situ* or *in-situ* land cover validation efforts.

Ex-situ approaches use crowdsourcing platforms and task users with labeling imagery, often based around a particular land cover product. For instance, *Geo-Wiki* (Fritz et al., 2009; See et al., 2013) was developed specifically to crowdsource data for land cover validation. See et al. (2013) concluded that the overall quality of data was relatively high and that differences between experts and non-experts exist but vary depending on land cover class and throughout the test period. In-situ approaches seek to motivate contributors to visit locations and generate data in the field. For example, *FotoQuest Austria* successfully generated data in over 1,699 unique locations, incorporating a simplified LUCAS protocol and engaging non-expert users by offering prizes to top-scoring players (Bayas et al., 2016).

An important question with respect to such approaches is the quality of non-expert contributions, and in particular the number of non-expert user contributions per unique location needed to ensure high quality. Haklay, Basiouka, Antoniou, and Ather (2010) show the validity of Linus's law in OpenStreetMap and discuss its validity

for VGI in general. The basis for Linus's law is the assumption that the quality of user contributions increases with the quantity of contributions (Haklay et al., 2010). However, Maisonneuve and Chopard (2012) state that the overall accuracy and variability of user contributions seem to stabilize at five volunteers, meaning allocating more volunteers to a task only increases resource use while not affecting quality. Haklay et al. (2010) also found the first five contributions to have the largest impact on data quality. Goodchild and Li (2012) argue that in practice, Linus's law applies to prominent and easily identifiable geographic facts rather than less-known or niche features.

## 2.2 | User motivation

A major issue in crowdsourcing is motivating users to make constructive contributions (Brabham, 2012; Coleman, Georgiadou, & Labonte, 2009; Fritz et al., 2017; Wright, Underhill, Keene, & Knight, 2015). Several motivators for participation have been identified, including developing a certain skill (Brabham, 2012), social interactions (Wright et al., 2015), contributing to scientific endeavors (Cox et al., 2018), and engaging in an enjoyable activity (Brabham, 2012). A more extensive list of motivators for users contributing geographic information was compiled by Coleman et al. (2009) through analyzing relevant literature. The authors mention eight key motivators, including altruism, professional or personal interest, intellectual stimulation, protection or enhancement of a personal investment, social reward, enhanced personal reputation, outlet for creative and independent self-expression, and finally pride of place. Budhathoki and Haythornthwaite (2013) investigate motivation at a finer-grained level, analyzing 22 motivational factors in OpenStreetMap (cf. Gómez-Barrón, Manso-Callejo, Alcarria, & Iturrioz, 2016). The authors found altruism, perceived importance of a project goal, instrumentality of local knowledge, and fun to be the main motivators. Interestingly, fun appears to be one of the most important aspects of user motivation in the findings of Budhathoki and Haythornthwaite (2013) which is not reflected in the former list of motivators (Coleman et al., 2009). In the context of land cover product research, Fritz et al. (2009) have suggested making the participation in land cover validation efforts more desirable by incorporating elements used in computer games such as competition and entertainment.

## 2.3 | Location-based gaming

Recent rapid developments in performance and features of handheld devices with integrated global positioning system (GPS) capabilities have led to a surge in popularity of LBGs, perhaps best illustrated by the wave of popularity of the LBG *Pokémon GO* (Colley et al., 2017). An increasing number of researchers are harnessing this interest in LBGs and using them as data collection tools (Bayas et al., 2016; Celino, 2015; Davidovic & Stoimenov, 2013; Matyas, 2007; Yanenko & Schlieder, 2014). LBGs are based on only allowing interactions with a virtual (game) environment when specific real-world locations are visited. Games aimed at collecting geographic information can be distinguished based on various characteristics: (a) the structure of the game field (Matyas, 2007; Matyas et al., 2012); (b) the duration of a game; (c) the presence or absence of a narrative or storyline; and (d) whether the game is team-based or not (cf. Avouris & Yiannoutsou, 2012). There is general agreement that LBGs are viable not only as (geospatial) data acquisition tools (Matyas, 2007; Yanenko & Schlieder, 2014), but also for data verification and curation purposes (Celino, 2015; Yanenko & Schlieder, 2014). Celino (2015, p. 9) concludes that using LBGs "can bring effective tools for geospatial data curation by exploiting the physical presence of the contributors in the environment."

Game play in the real world typically also involves immersion in some form of linked virtual world (e.g., the monsters in *Pokémon GO*). Gradinar, Huck, Coulton, Lochrie, and Tseklevs (2015) argue that overemphasizing virtual immersion elements can draw a user's attention away from the physical environment. This in return reduces the users' ability to perceive and reflect on their surroundings, which is an important aspect of in-situ data generation. Underlying game mechanics and design features must therefore be carefully implemented to balance immersion between the virtual game world and the physical world (Gradinar et al., 2015; Lei & Coulton, 2011).

Many games use virtual reward systems as a key motivational mechanism to ensure player satisfaction and thus increase the time users invest in playing a game (King, Delfabbro, & Griffiths, 2010). Virtual rewards can be seen as a proxy for the time and effort a player has invested in a given task in a game and are often comparable and communicable with other players (Hoe, Goh, Pa, Pe-Than, & Lee, 2017). Relatedness, or the desire of users to feel connected with and interact with other users, has been argued to increase users' intrinsic motivation (Deci & Ryan, 2000; Hoe et al., 2017). Motivational elements in an LBG may thus not only correlate to perceived enjoyment, but also influence user behavior.

Individual user behaviors in LBGs can in return result in different spatial contribution patterns. The literature suggests two main behavioral types: users primarily playing LBGs while traveling (Bell et al., 2006; Zhang, Lee, Radhakrishnan, & Balan, 2015), leading to route-like contribution patterns, and users significantly adjusting their route to play Colley et al. (2017), potentially leading to spatially coherent contribution patterns.

Despite these advantages of LBGs, current implementations often have weak narratives and have not yet fully explored the potential of gamification as a motivational element. Given the surge of interest in crowdsourcing land cover data, we set out to investigate the potential of an LBG with strong game play and narrative elements as a source of ground truth data for land cover validation.

## 2.4 | Data quality and user characteristics

Crowdsourcing data commonly leads to a more or less diverse base of contributing users with varying characteristics and collection behaviors. Coleman et al. (2009) give an overview of users who contribute to volunteered geospatial data sets and discuss characteristics of their contributions. The authors summarize five categories of contributors: "neophytes" (e.g., interested users with time and a certain level of willingness to contribute but with no formal background); "interested amateurs" (e.g., interested users with limited background knowledge); "expert amateurs" (e.g., interested non-professional users with broad background knowledge); "expert professionals" (e.g., professional users with extensive background knowledge); "expert authority" (users who are recognized as experts in an area). In addition, users contributing to a crowdsourcing effort were characterized more broadly by their humanity as distinguishable from automated systems, the frequency, type and degree of their contributions, the quality and veracity of their contributions, and their reputation in the context of contributions.

Flanagin and Metzger (2008) dissect the terms "accuracy," "credibility," "trust," and "quality" in the domain of VGI and point out that contributions in crowdsourced projects must be assessed regarding their reliability. Bordogna, Carrara, Criscuolo, Pepe, and Rampini (2016) analyze different volunteer contribution behaviors and also conclude that the quality of the resulting data is heavily influenced by the underlying characteristics of contributing users. Comber, Mooney, Purves, Rocchini, and Walz (2016) analyzed contributions to Geo-Wiki, specifically focusing on the contributed land cover classifications based on the nationality of contributing users and found differences in the labeling of land cover classes.

All of these issues emphasize the importance of exploring not only the properties of the data contributed in crowdsourcing, but also the properties and influence of the participants themselves.

## 2.5 | Research gaps

Reviewing the presented literature highlights important research gaps which must be addressed and existing research which can be built upon. Gamification has enjoyed increasing attention in scientific research, especially in crowdsourcing and citizen science approaches. However, most efforts incorporate single arbitrary game elements calling for the implementation and analysis of a coherent LBG, including competition, progression, an underlying narrative, and entertaining game play. Games have been suggested as having the potential to motivate non-experts to participate in improving land cover products; however, research on using LBGs to generate in-situ land cover data is limited and needs further exploration. Especially questions of how to motivate users need to

be addressed in the context of LBGs as scientific data generation tools. Further, research analyzing UGC from an LBG focusing on between-user agreement rates and agreement rates between users and an authoritative data set is limited.

### 3 | METHODS AND IMPLEMENTATION

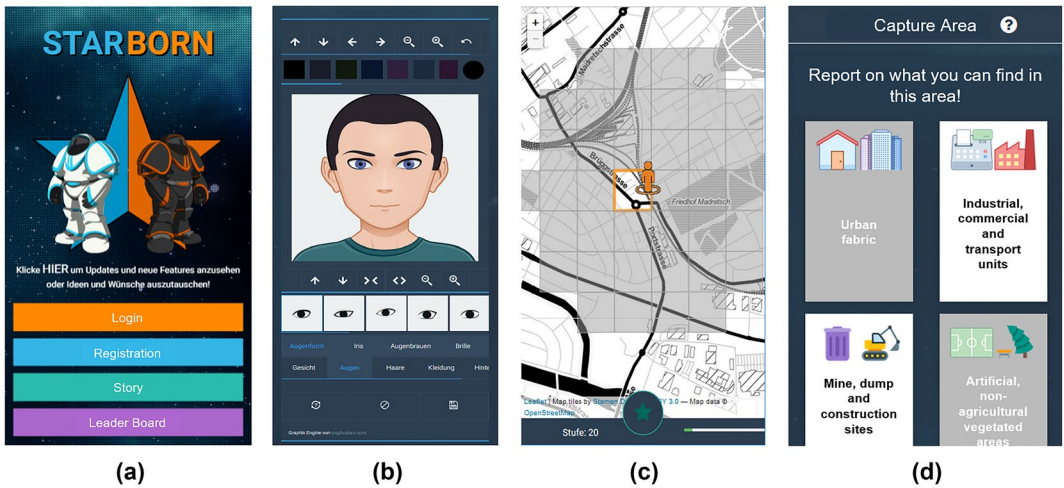
In this section we describe the key steps we took in implementing the LBG StarBorn for Switzerland and analyzing the generated data. We describe how the game was implemented, focusing on the introduction and iterative development of elements designed to motivate game players, and discuss the recruitment and retention of a community of game players. We highlight key methods for analyzing user characteristics and contributions. Finally, we present methods used to compare the in-situ land cover classifications generated through StarBorn with an authoritative land cover product.

#### 3.1 | Key game play and implemented game elements

Game elements are the fundamental building blocks of any game and are chosen or omitted depending on intended game play (Aldemir, Celik, & Kaplan, 2018). A game designer can therefore use game elements to guide users through various stages of a game, foster motivation, and tailor user experience.

Users interested in playing StarBorn (Figure 1a) were first encouraged to read a back story portraying opposing extraterrestrial teams (*Blue Dwarves* versus *Red Giants*) in a struggle for world dominance, serving as the underlying theme of the game. Users signed up for the game and were prompted to submit limited demographic information (age and gender). We did not collect information about the academic level or occupation of the participants to make the sign-up process as easy as possible while retaining some basic demographic information. Following email confirmation, users were invited to create an avatar, a self-designed icon representing the game player (Figure 1b) as a means of identity building and motivation (cf. King et al., 2010). Users then affiliated themselves with one of the two teams, purposefully implemented as a competitive element (opposing teams) as well as a collaborative element (common goals within teams) to increase user motivation (cf. Söbke, Hauge, & Andreea Stefan, 2017). This further stimulated interest and introduced users to the competitive elements of the game. We did not explicitly mention land cover data collection as being the main aim of StarBorn, but did inform users that contributed data would be used in a research context.

The implemented LBG encouraged users to conquer virtual game areas corresponding to fixed real-world extents of 200×200 m in Switzerland through continuous real-time game play (Figure 1c). Conquering real-world extents in a virtual game world was the main activity of the implemented LBG. To further encourage users to visit and conquer as many tiles as possible, users were rewarded with different *titles* (short texts indicating a player has earned a certain achievement) and *badges* (small images with increasing rarity corresponding to the mentioned titles) depending on their number of contributions (cf. Hamari, 2017; Hoe et al., 2017). In addition, a rankings page was implemented showing a ranked list of users by the number of respective contributions. All interactions related to conquering an area required the users' real-world position to be within the virtual game-tile. Conquering an area for their team thus required users to physically visit a location and select one or more land cover classes represented by a title and an icon (Figure 1d) from a predefined list of classes based on CORINE land cover data (European Environment Agency, 1994). Table 1 shows the CORINE land cover class titles and their respective abbreviations used in the remainder of this article. CORINE does not clearly distinguish between land cover—"biophysical attributes of the earth's surface" (Lambin et al., 2001, p. 262)—and land use—"human purpose or intent applied to these attributes" (Lambin et al., 2001, p. 262). The classification scheme we implemented thus incorporated both land cover (e.g., urban, forest) and land use (e.g., industry, pasture) classes. For simplicity, we will refer to all classes as land cover classes in the remainder of this paper.



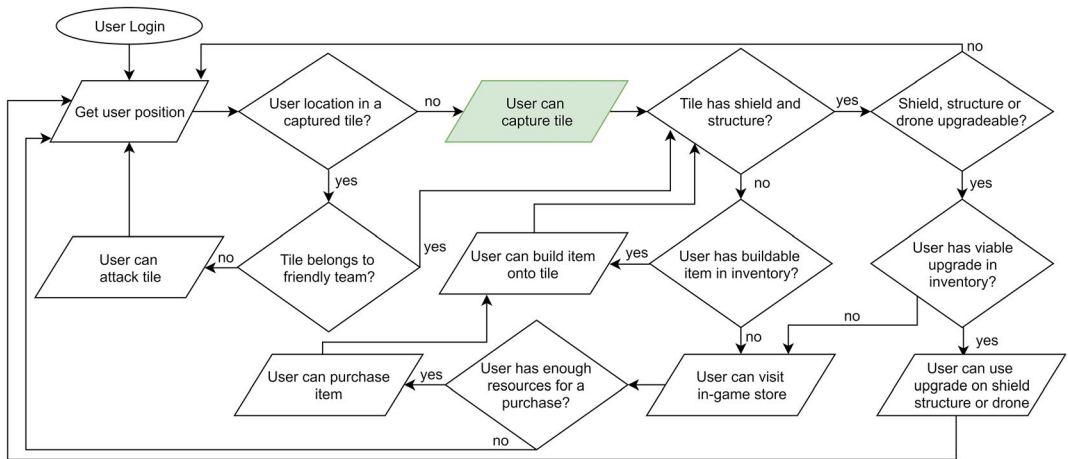
**FIGURE 1** Screenshots of game play showing (a) the homepage where users could login and register, view the current rankings page and read the back story of the game; (b) the character creation page where users could create their individual avatar; (c) the main game map window showing the user's location as a marker, the surrounding game-tiles colored by team affiliation and, finally; and (d) the capture screen where users report on one or more perceived land cover classes to capture a game-tile

**TABLE 1** Table of land cover class abbreviations as used in this article and their respective CORINE titles

Land cover class abbreviation	CORINE land cover class title
<i>urban</i>	Urban fabric
<i>industry</i>	Industrial, commercial, and transport units
<i>forest</i>	Forests
<i>arable</i>	Arable land
<i>pasture</i>	Pastures
<i>greenarea</i>	Artificial, non-agricultural vegetated areas
<i>shrub</i>	Scrub and/or herbaceous vegetation associations
<i>water</i>	Inland waters
<i>novveg</i>	Open spaces with little or no vegetation
<i>agriculture</i>	Heterogeneous agricultural areas
<i>permacrop</i>	Permanent crops
<i>wetland</i>	Inland wetlands
<i>mine</i>	Mine, dump and construction sites

Conquering a location was rewarded with in-game resources (e.g., “stardust”). In-game resources could be spent on in-game items, including various defensive structures, which could be built onto game-tiles belonging to one's own team, making it more difficult for the other team to take over a conquered area. We were interested in analyzing between-user agreement rates by comparing consecutive classifications of the same area and thus implemented an *attack and reconquer* feature. For instance, players from the Red Giants could take over game-tiles held by the Blue Dwarves by attacking the game-tile's defensive structures and recapturing the respective tile by resubmitting the perceived land cover classes. This resulted in tiles being classified multiple times and thus consecutive agreement rates could be analyzed. However, given the added needed effort of first attacking an





**FIGURE 2** Flowchart of main user decisions during game play

area when reconquering an enemy area, users were implicitly directed towards conquering areas not yet classified. These features were implemented to ensure a large spatial coverage while allowing a subset of tiles to have multiple captures. All major user decisions are summarized in Figure 2.

### 3.2 | Infrastructure decisions

The defined goal of StarBorn was to generate land cover data from users by exploiting their ability to sense their immediate surroundings (Celino, 2015; Wang & Ben-Arie, 1996). Therefore, we implemented the game as a browser-based web application, using modern browsers' built-in location capabilities (accessing GPS) and only allowing users to contribute land cover classifications of their immediate real-world surroundings. The virtual game field comprised a raster of game-tiles, stored using a national equal-area projection coordinate system for Switzerland and covering the whole area of Switzerland. A raster cell resolution of 200×200 m was chosen based on four criteria. First, the tiles were large enough that players could conquer relatively large areas and to encourage competition to reconquer tiles. Second, moving within the game was rapid on foot, encouraging players to capture more than one tile in relatively rapid succession. Third, this resolution is of the same order of magnitude as the data we wished to compare (CORINE with a resolution of 100 m). However, we judge 100 m to be too fine a resolution for our first criterion, and thus settled on a compromise of 200 m. Finally, the chosen resolution allows public access to most tiles, minimizing potential access issues in a Swiss context.

### 3.3 | Pilot study and iterative game design

Before members of the public were allowed to play the game, a selected group were invited to test it and give feedback on their user experience. Key improvements and new game elements were introduced based on the feedback of test users. With respect to competition, improvements were implemented including introducing a rankings page to foster competition and motivate players to contribute more data as a means of increasing their ranking (Hoe et al., 2017) and introducing player progression elements such as experience points and a nonlinear level system to increase user satisfaction (King et al., 2010; Wang & Sun, 2011). To increase the spatial coverage of contributions we added a treasure hunt system to motivate users to contribute data in remote areas (Bell et al., 2006; Colley et al., 2017; Foody, 2002). Further improvements included reducing the number of obtainable resources from six to two to reduce complexity (Iyengar & Lepper, 2000). Special in-game events were introduced



at specific times with higher rewards received for capturing tiles or greater damage inflicted on opposing players so as to motivate users to continue playing (Cantallops & Sicilia, 2016).

### 3.4 | User recruitment

Motivating new users to join and play a game is paramount to its success. In particular, competitive elements of a game (in our case used to ensure that we could collect multiple observations of land cover for the same location) rely upon game play and a sufficient number of players. It is generally agreed that content creator–consumer or consumer–consumer interactions can create a sense of connectedness and can facilitate peer feedback, with a positive impact on user motivation (cf. Brabham, 2012).

We recruited our first users by word of mouth. These were mostly friends, family and students at various higher education institutions in Switzerland. In a second step, we used social media platforms for recruitment. StarBorn was mostly promoted in large online community groups revolving around similar LBGs on Google+. We also created a Google+ community to interact with interested users and give the users a platform to exchange ideas, report bugs and get information about upcoming in-game events.

### 3.5 | Analyzing spatial patterns of user behaviors

Since previous work has identified typical patterns of behavior in both citizen science projects generally (Haklay, 2013; Tulloch & Szabo, 2012) and LBGs specifically (e.g., Lund, Lochrie, & Coulton, 2010) we initially visualized patterns of individual users' behaviors to explore whether behavioral patterns were visible. However, unlike Lund et al. (2010), we did not focus on temporal contribution patterns, but identified three spatial behavioral characteristic patterns:

- *route-based* play, where it appeared that players were taking part in the game while traveling, with linear patterns, often following public transport routes;
- *area-based* play, where players appeared to aim to capture a large number of tiles within a given spatial extent; and
- *random* play, where a small number of more or less randomly located tiles were captured.

To distinguish between these behaviors we used the DBScan algorithm to cluster data based on a specified search radius ( $\epsilon$ ) and the number of neighbors needed (*minPts*) (cf. Ester, Kriegel, Sander, & Xu, 1996). Route-based play was identified by using a large radius and a small number of neighbors (effectively identifying long, thin clusters) and area-based play behaviors by using a smaller radius with more neighbors (identifying compact clusters of coherent spatial extents). We identified parameter values of  $\epsilon=500$  m and *minPts*=10 as being most appropriate for area-based play, and  $\epsilon=900$  m and *minPts*=3 for route-based play, with all other user contributions being classified as random behavior (in practice, these were typically players who only classified a small number of tiles). We categorized all users according to their predominant user behavior. To compare behaviors we tested for correlations of the number of contributions of each land cover class using Spearman's rank correlation test and plotted the number of classifications of each land cover class for area and route behaviors.

### 3.6 | Agreement rates of users classifying the same location

Uncertainty is an influential characteristic of user-generated data, but how can we better quantify the uncertainty in our data set? One obvious approach is to compare classifications of different users for the same game-tile and thus measure between-subject agreement. We analyzed the agreement rates of consecutive classifications of each game-tile, and explored variation in agreement rates as a function of both how often a tile was classified,

and the classes to which a tile was allocated. By exploring consecutive agreement rates of the same tile, we can explore whether, for example, land cover classes or user behavior influence uncertainty.

We first explored relative agreement between consecutive classifications for the same tile submitted by multiple users. We created a subset containing only tiles classified by at least two users. For each tile we ordered contributions according to the submission time-stamp. We assessed agreement rate for a game-tile by comparing every classification a user submitted with the consecutive classification. Land cover classes reported in both classifications counted as agreements and land cover classes only present in one of the two contributions as disagreements. The agreement rate for consecutive contributions is calculated as the number of agreements divided by the sum of all land cover classes submitted for a tile.

We hypothesized that a positive correlation exists between the number of classifications and the corresponding agreement rates of consecutive classifications; in other words, the more a location is classified, the more agreement we expect between users. This assumption is based on the applicability of Linus's law in user-generated spatial data, as suggested by Haklay et al. (2010). To explore this hypothesis we generated box-and-whisker plots of agreement rates compared to the number of classifications and calculated Spearman's rank correlation values between mentioned agreement rates and number of classifications.

### 3.7 | Comparing user-generated land cover classifications with an authoritative land cover data set

We compared data collected with StarBorn with CORINE classifications to explore agreement between the two data sets. In doing so we implicitly assume that CORINE is of sufficient quality to tell us something about the quality of the data we collected, but at the same time we expect to identify patterns of disagreement which might help us better understand uncertainties or disagreements between CORINE and other real-world applications.

In this analysis we analyzed all tiles irrespective of the number of users who had classified them. We first calculated the overall agreement rate using all game-tiles with one or more reported land cover classifications. We defined agreement where any of the user reported land cover classes for a tile (e.g., *forest, industry, water*) corresponded to the land cover reported in the authoritative data set (e.g., *industry*). We further calculated average agreement rates dependent on the behavioral group of the contributing users to shed light on potential accuracy differences between the different behaviors.

In a second step, we created a confusion matrix using only hard user-generated land cover classifications (tiles where only a single land cover class was reported by users). We calculated measures of precision and recall following Fawcett (2006):

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (2)$$

where *TP* denotes true positives, *FP* false positives, and *FN* false negatives.

Furthermore, we calculated the *F*-score (the harmonic mean of precision and recall) (cf. Fawcett, 2006) as:

$$F = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}. \quad (3)$$

An *F*-score of 1 consequently indicates complete agreement between the user-generated land cover classifications and the authoritative land cover data set, while an *F*-score of 0 indicates no agreement. We first calculated the *F*-score

for each individual land cover class and then compared the  $F$ -score with the number of user-generated classifications of the respective land cover class.

Finally, we calculated the overall agreement rate, defined as the sum of the diagonal of the confusion matrix divided by the total number of contributions (cf. Brovelli, Molinari, Hussein, Chen, & Li, 2015). To identify potential differences depending on hypothesized user behaviors, the overall agreement rate of hard classifications was additionally calculated again for *area* and *route* contribution behavior groups.

## 4 | RESULTS AND INTERPRETATION

In this section we present results with respect to the data collected during the approximately 3 months that StarBorn was live. In doing so, we first briefly summarize the number and variety of land cover tiles contributed by players, before exploring reported user demographics. We then present the nature of contributions as a function of different patterns of user behavior. Finally, we analyze agreement of consecutive classifications of same locations and agreement rates between the user-generated land cover classifications and an authoritative data set.

### 4.1 | User contributions

From the game going live (November 9, 2016) until promotion was stopped and data were analyzed (February 23, 2017), players classified a total of 13,319 tiles distributed across 11,364 unique locations. Given the resolution of 200×200 m, this corresponds to a total classified tile area of 533 km<sup>2</sup>. The majority of tiles were classified as containing *urban* ( $n=8,010$ ) or *industry* ( $n=7,724$ ) land cover classes. This is unsurprising and reflects spontaneous game play in locations where players are most likely to find themselves. The next most frequent classifications were of *pasture* ( $n=2,716$ ), *forest* ( $n=2,418$ ), *arable* ( $n=2,221$ ), *shrub* ( $n=1,631$ ), *water* ( $n=1,166$ ) and *greenarea* ( $n=1,065$ ). In total 28,633 land cover classes were reported, giving an average of 2.15 land cover classes reported per captured tile.

Our game design aimed to encourage multiple users to classify the same tile (by, for example, attacking and (re) capturing tiles). However, of the total of 11,364 game-tiles captured, only around 10% (936 tiles) were captured multiple times. These results highlight users' desire to conquer "empty" tiles, potentially due to the additional effort needed to recapture areas belonging to the enemy team. Figure 3 shows for an area around the city of Zurich the pattern of contributions and their frequency where players were most active. Notable are the linear patterns of regularly captured tiles, often associated with transport corridors.

### 4.2 | User demographics

In the period during which the game was online, 138 users registered, of whom 84 captured at least one tile. The rate of registering users can be divided into three distinct periods: beta-testing, live with promotion and live without promotion. Figure 4 highlights the total number of registered users over time and thus the rate of player registration. During the open beta period the game was shared among a small, select audience to test server stability and ease of use and a correspondingly small uptake is visible. Between November 9, 2016 and December 8, 2016 a considerable increase in users can be observed. In this period, the game was actively shared and promoted. In the promotion period, the game was also advertised in various lectures. After December 8, 2016, the game was no longer actively promoted and a rapid decline in the number of new players registering is observed. These curves point to the importance of promotion in building an active and large community of players.

We collected basic demographic information about age and gender. However, although players were required to register, these data should be treated with care. For example, user-reported years of birth were mostly between 1987 and 2003 (min = 1954; max = 2016; mean = 1989), reflecting a largely student population, but spiked in 2000 (the default value). Gender distribution of active players was heavily biased towards males: 62 players reported

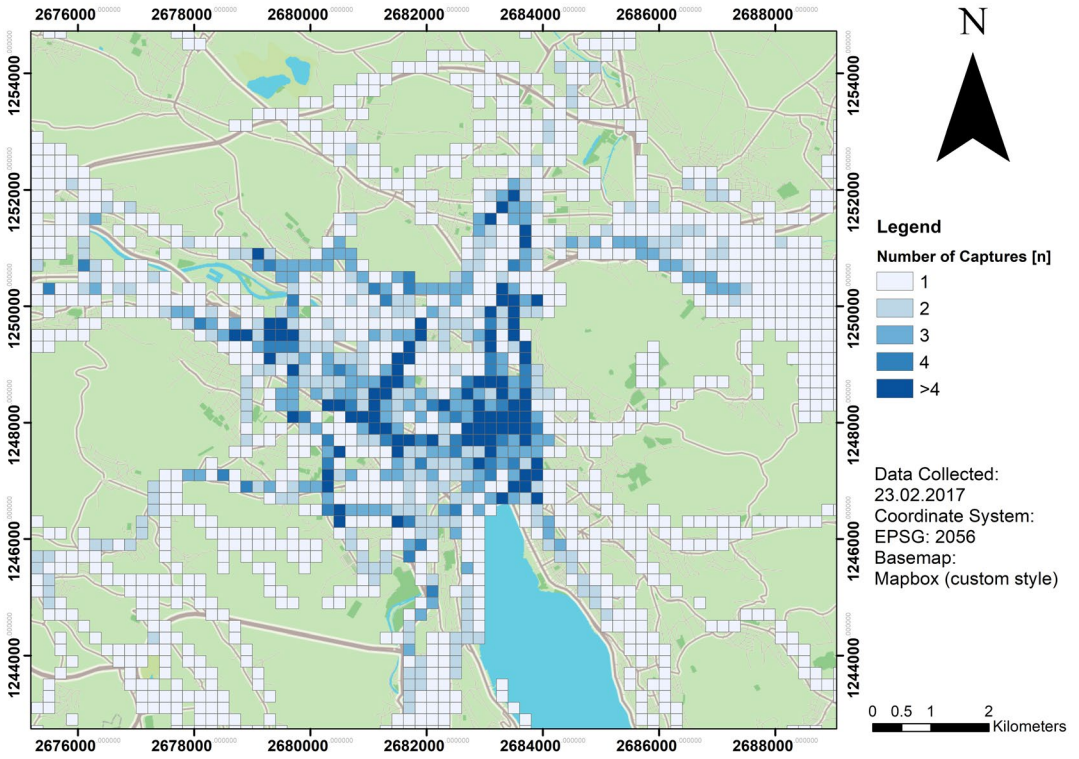


FIGURE 3 Tile capture count around the city of Zurich

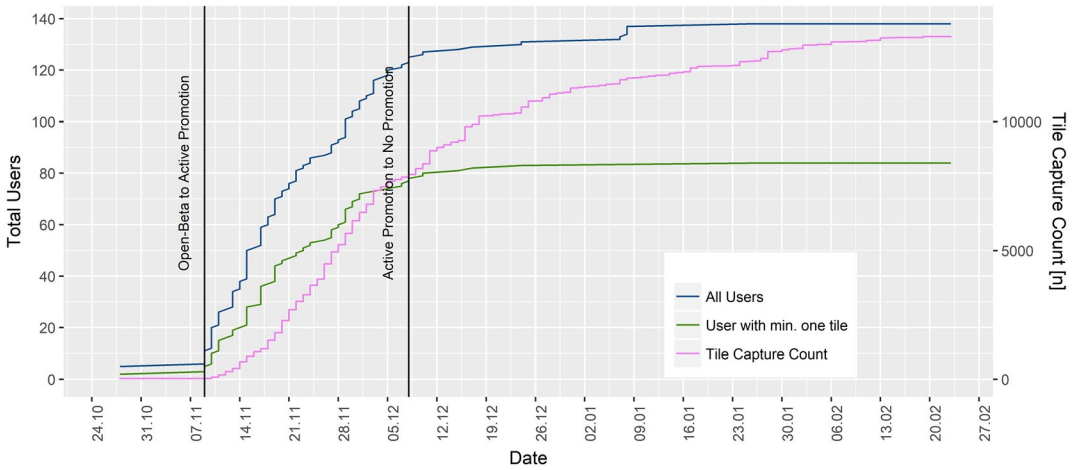


FIGURE 4 Total number of registered users and captured game-tiles over the extent of the game period. Two vertical lines indicate key transitional moments

being male, 16 female, and 6 chose not to report gender. We observed a slightly higher average number of contributions from reported female users, with highest average number of contributions coming from those who did not report gender (mean = 360.4; median = 350, n=6), followed by females (mean = 163.8; median = 55.5; n=16) and males (mean = 142.3; median = 29.5; n=62).

### 4.3 | User behaviors

We analyzed users not only with regard to their individual demographic characteristics but also according to their in-game behavior. Users belonging to the random behavior class ( $n=25$ ) captured a negligible number of tiles (51) and we do not report on their behavior in detail. Fifteen users were classified as belonging to the area behavior class and captured 3,987 tiles in total, while the largest subgroup belonged to the route class ( $n=44$ ) and captured 9,281 tiles in total. Players classified as exhibiting area behavior captured statistically significantly more tiles on average (mean = 265.8, median = 156) than those playing with route-type behaviors (mean = 210.9, median = 67; Wilcoxon rank test  $W=216.5$ ,  $p<.05$ ). The large difference between the median and the mean indicates that the results are particularly influenced by outliers (e.g., single users contributing considerably more data than others). Although individual players belonging to the area class contributed more tiles, the larger number of players playing with route behavior contributed overall more tiles (cf. Figure 5). This observation gave rise to the question whether user behavior influences the types of land cover contributed.

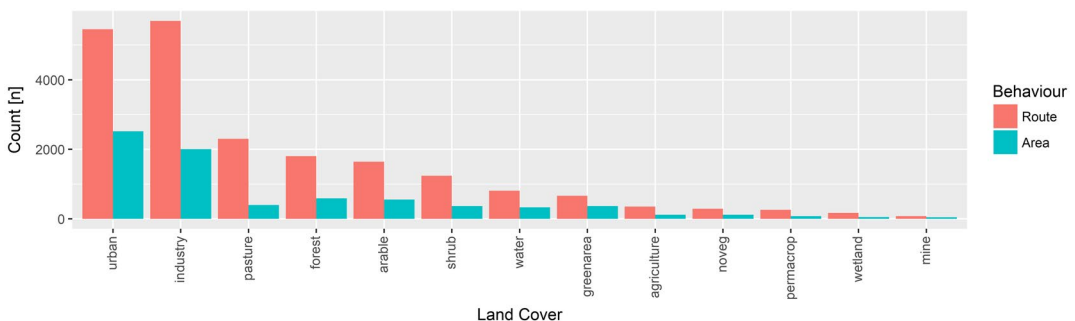
We found a strong correlation between the number of contributions for each land cover class between the area and route behavior groups (Pearson test  $r = 0.97$ ,  $p < 0.01$ ). In Figure 6 we plot the resulting fit, with 95% confidence intervals to identify outliers with respect to this correlation. Three land cover classes lie outside these confidence intervals: *urban* and *greenarea* are associated more than expected with area behavior and *pasture* more than expected with route behavior. Since two of these classes are ranked first and third overall by frequency, the influence of behavior on their categorization is potentially of importance.

### 4.4 | Location-specific agreement rates between users

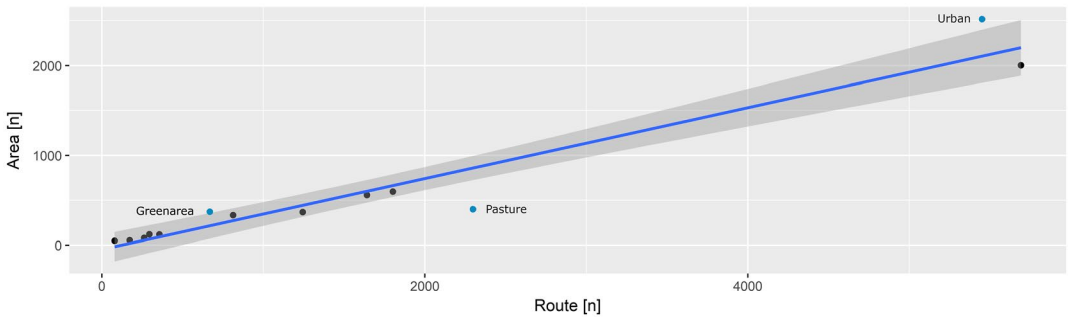
Given the relatively short time-span of data collection (approximately 3 months), no explicit temporal analysis was conducted. We did, however, implicitly include temporal aspects in our analyses by investigating consecutive classifications where classifications are ordered by time of contribution and by analyzing correlations with user contribution counts which are a proxy for the time respective users spent playing the game.

We analyzed user agreement rates for game-tiles consecutively captured by different users and identified a total of 936 tiles with multiple classifications. Taking all reported land cover types into account, the overall consecutive agreement rate was 42.2%.

When analyzing the number of tiles according to the number of classifications by different users, we observe a highly skewed distribution towards many tiles having few classifications, which is common for crowdsourced data (Stewart, Lubensky, & Huerta, 2010). This typical skewed distribution, along with visual inspection of Figure 3, shows that the distribution of game-tiles with multiple classifications is highly heterogeneous and spatially clustered. This begs the question whether or not users agree more on the reported land cover classes in areas with many contributions. The box-and-whisker plots comparing classification counts with agreement rates (Figure 7) show increasing agreement rates in consecutive classifications with increasing number of classifications. The



**FIGURE 5** Count of reported land cover classes by user behavior



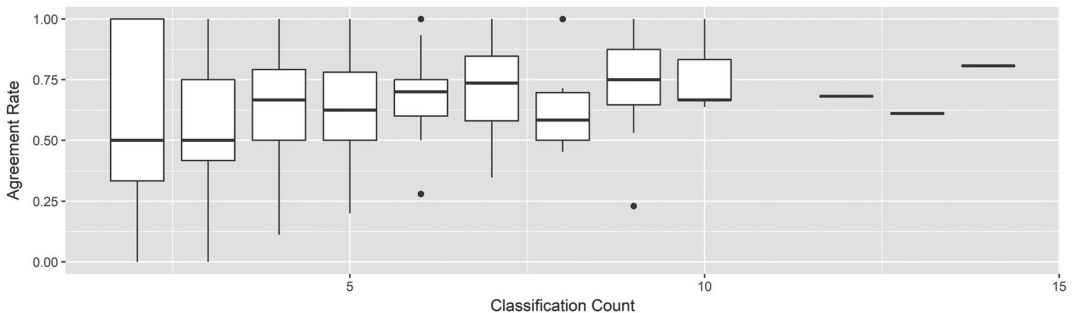
**FIGURE 6** Comparison between total number of contributed classifications per land cover class for area and route behaviors. Trend-line and smoothing: method = l m, level = 0.95

correlation between the number of classifications and agreement rates of consecutive classifications is statistically significant both for the median (Spearman rank test  $\rho = 0.67, p < 0.01$ ) and mean (Spearman rank test  $\rho = 0.73, p < 0.01$ ) agreement rates. We argue this result indicates the validity of Linus's law in our data set of user-generated land cover classifications, as more classifications correlate with higher agreement rates among users. In addition, we interpret this result as an effect of the fact that game-tiles with many classifications are likely to be easily accessible and therefore also more familiar (e.g., urban locations).

#### 4.5 | Agreements between user classifications and CORINE classes

We are interested whether user-generated land cover classifications show potential to be used in land cover product validation efforts. We therefore compared our user-generated data set with an authoritative data set (the CORINE 2012 land cover product) to reveal key characteristics, uncertainties and differences between the two data sets. Users contributed a total of 13,319 classifications, of which 10,157 (76.3%) classifications showed an agreement between the authoritative land cover data set in one user-reported class. We identified noteworthy differences between the agreement rates of behavioral groups of users with the authoritative land cover data set. We found agreement rates of 82.3% for area-based user contributions (agreement in 3,282 out of 3,987 tiles) and 73.7% for route-based user contributions (agreement in 6,836 out of 9,281 tiles) between the authoritative data set and the user-generated data set. This implies that area-based contributors were either playing the game more seriously than route-based users, or that area-based players played the game in areas where the land cover class was easy to identify (e.g., *urban, greenarea*).

As we found a statistically significant correlation between the number of times an area was classified (by different users) and the between-user agreement rates, we were especially interested to see if the agreement rate with CORINE classes also correlated with the number of contributions. We found a statistically significant negative correlation between the number of classifications of a user and their respective overall agreement rate with



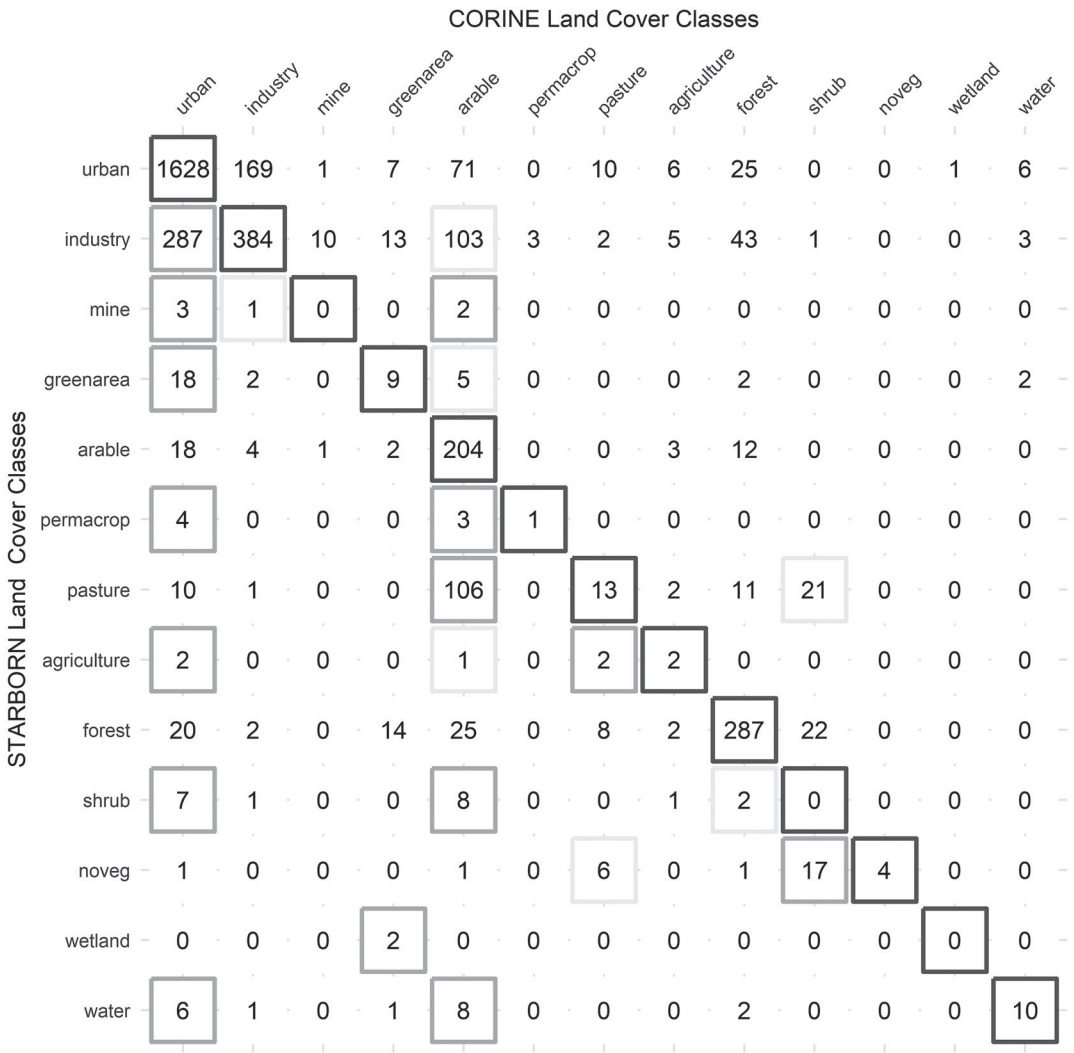
**FIGURE 7** Box-and-whisker plots of between-user agreement rates versus classification count

the authoritative land cover data set ( $\rho = -0.34, p < 0.01$ ). In other words, users with a lower number of contributions have (somewhat) higher agreement rates with the authoritative data. We interpret this result as demonstrating that, contrary to some arguments, in-game performance does not increase with experience. Users with more contributions may be classifying tiles quickly to maximize in-game performance, making their contributions potentially less accurate (since there is no penalty to game play associated with disagreeing land cover classifications).

### 4.6 | Confusion matrix between user classifications and CORINE classes

In this subsection we report the results for the confusion matrix between land cover classifications from users of StarBorn and the CORINE data set. The confusion matrix only includes classifications where a single class was reported, constituting 3,704 tiles (compared to 13,319 total classified tiles).

The confusion matrix in Figure 8 sheds light on major disagreements between individual land cover classes. The results show, for example, that 64.6% ( $n=106$ ) of tiles classified as being *pasture* in the user-generated data set



**FIGURE 8** Absolute confusion matrix showing the number of confusions between individual land cover classes of the UGC and the authoritative data set. The diagonal shows agreements between StarBorn and CORINE (dark gray, at least 20% of user-generated data; light gray, at least 10% of user-generated data)



were classified as *arable* in the authoritative land cover product. Other major confusions include *novveg* with *shrub* (56.7%,  $n=17$ ), *industry* with *urban* (33.6%,  $n=287$ ) and *pasture* with *shrub* (12.8%,  $n=21$ ). The results arguably show that similar land cover concepts seem to have a major negative influence on overall agreement rates and imply that land cover pairs showing high confusion rates are also hard to differentiate semantically.

The confusion matrix shows an overall accuracy of 68.6%. The rather high overall agreement rate can be attributed to high agreement rates between user-generated land cover classifications and the authoritative data set in land cover classes which were reported frequently. These were *urban* ( $F = 0.8289$ ), *industry* ( $F = 0.5412$ ), *arable* ( $F = 0.5224$ ) and *forest* ( $F = 0.7503$ ). Regarding different user behavior groups, we found an agreement rate of 74.6% in the area behavior group and 64.5% in the route behavior group.

## 5 | DISCUSSION

In the introduction we set out three research questions which we wished to address, relating to: the design and implementation of LBGs; the characteristics of game players, including the influence of different behavioral characteristics on the data collected; and comparison with an authoritative data set. In this section we briefly return to each of these questions, setting out the strengths and limitations of our approach, and comparing it with previous work.

Key to our implementation was going beyond the state of the art in many attempts at gamification in GIScience, and focusing on developing an LBG to motivate players to contribute land cover data. What sets our game apart from previous attempts is the incorporation of a fantasy-based narrative as an underlying story (cf. Kenny & Gunter, 2007), along with a multitude of game elements. We argue that simply having a rankings page does not suffice for an application to be deemed an enjoyable game. We thus made a special effort to surpass many LBGs implemented in GIScience by including a myriad of game elements focusing on entertainment and competition such as an underlying narrative, team-based play, a rankings page and an attack-and-reconquer feature. Many implementation decisions with respect to game play were made in response to feedback from active users. These included (during the beta-testing phase) performance improvements of database queries to improve responsiveness of the game, the implementation of new features to increase user motivation (e.g., level and experience points systems) and introducing new game elements (e.g., treasure hunt system). These improvements crucially not only increased game playability and enjoyment, but also, according to individual user feedback, fostered a sense of community and thus increased individual motivation and overall contributions (cf. Crall et al., 2017; Hoe et al., 2017). Beyond the implementation of the game itself, promotion focused on use of existing contacts and networks, and, unlike other similar work, we did not use, for example, media channels (cf. Bayas et al., 2016). However, we note that the number of contributions generated was comparable, providing an indication of the importance of well-targeted campaigns, rather than blanket efforts in the media (cf. Crall et al., 2017).

The total number of user contributions over time showed a characteristic sigmoid curve, with the rate of new user contributions decreasing significantly towards the end of the collection period. This decline in interest is also observed in highly popular commercial LBGs including *Pokémon GO* (cf. Andone et al., 2017). Long-term low retention rates suggest that after a short period of initial engagement, users are not sufficiently motivated and lose interest. Since user retention has been identified as key for successful crowdsourcing or citizen science campaigns (Crall et al., 2017), approaches to retaining users are important and resources should be allocated accordingly.

Analyses of Geo-Wiki, the crowdsourced land cover validation project, for example, show a rapid increase in user contributions towards the end of the data collection phase (See et al., 2013), arguably due to users with the most points at the end of the data collection period having a chance to win Amazon vouchers or paper co-authorships (See et al., 2013). This suggests that incorporating some form of reward for high-contributing users at fixed moments in time could increase competition and ultimately user retention. Another strategy suggested is active and continuous community engagement, including regular updates on social media platforms, communicating

project progress, acknowledging contributions, and incorporating user feedback (e.g., incorporating user requests), while taking care not to overburden users (Crall et al., 2017).

As is typical in many crowdsourcing efforts, we found both an age and gender bias in our data. Gender bias may demonstrate some gender-specific affinity towards gaming in general (Willemse, Waller, & Süß, 2016). Whether the predominantly young players illustrate a similar affinity to video games or an effect of our promotion efforts is unclear, but it is important to note that these biases are likely to influence properties of the data collected. In terms of user-specific influence on the data generated, we also investigated the contributing users' behaviors. The group who appeared to play along routes captured some 70% of all tiles and likely consists primarily of those playing as a secondary activity while engaged in other activities such as traveling or commuting (cf. Bell et al., 2006; Zhang et al., 2015). Traveling using forms of heightened mobility (e.g., car, bus, train) can increase the users' in-game performance if not addressed through game mechanics, as a larger in-game area can be covered compared to walking. This hypothesis corresponds to the assumptions of Bell et al., who analyzed user behavior in another LBG, stating: "Journeys may have been good times to play, as players naturally move through different locations" (Bell et al., 2006, p. 422). This appears to have been the case in StarBorn, as such players contributed more tiles in total. Players exhibiting area contribution patterns were most likely playing the game as their primary activity. These users appear to have moved around to maximize in-game performance and contribute areas of coherent extent. Colley et al. (2017) and Bell et al. (2006) both reported users actively and substantially changing their route between given locations in order to play an LBG. Users belonging to the area behavioral group showed an overall higher average number of contributions and a lower spread in average contributions per land cover class per user. This can be interpreted as a more consistent data-contributing behavior. We suggest that identifying such types of game play is important when considering ways to incentivize players. Those playing the game as a secondary activity while traveling are unlikely to take different routes or be motivated to visit locations off the beaten track. This points to the most important weaknesses of our approach: although we collected a large volume of data, it is spatially heterogeneous and concentrated in areas which our players were likely to visit during their daily lives. Finding ways to motivate players to visit other locations through the game narrative is important if we wish to collect notably more non-urban tiles. Two diversification approaches (treasure hunt system, in-game events) were implemented, but the effects on contribution behavior were not explicitly analyzed. A potential further approach might be to emphasize the importance of visiting new or unusual places by adding an element of discovery to game play, similar to that documented in GeoCaching (O'Hara, 2008). However, a number of tiles are difficult or impossible to reach for most individuals (e.g., on cliffs, in lakes, dense forests) (cf. Bayas et al., 2016). Even though areas with limited access can contain interesting and often neglected information, crowdsourced data-generation efforts must not encourage dangerous or illegal behavior, further biasing results towards easily accessible locations.

Our results show large variations in the general agreement rates of consecutive classifications within individual tile locations as well as in the different land cover classes. The results also highlight the heterogeneity of users in terms of agreement rates. Interestingly, we found a statistically significant negative correlation between the number of contributions of a user and the overall agreement rate of the respective user with the authoritative data set. These findings seem to contradict the findings of See et al. (2013) who state that non-expert users tend to improve at a land cover classification task the longer they participate. Our findings could thus be an important characteristic specifically of data generated by an LBG. We believe that this highlights one of the biggest potential weaknesses of using an LBG to generate data: users with unusually large contributions may be interested only in increasing in-game performance and thus neglect the quality of the underlying classification task. In other words, users with a large number of classifications were arguably more interested in entertainment and virtual rewards than in contributing high-quality data. In addition, we argue that new users were potentially more careful in classifying a given extent since they were not yet aware of the lack of in-game repercussions from random classifications. This links to questions of immersion and arguably shows a shift towards virtual world immersion on the continuum (physical world–virtual world) of immersion. Users making less contributions show higher agreement rates, suggesting a more balanced immersion in both the physical world (e.g., when capturing an area and thus

contributing in-situ data) and the virtual world. This suggests that even though the implemented LBG successfully immersed players in the physical world as well as the virtual world, it may not have been successful in maintaining a balance on the immersion continuum.

We were able to identify statistically significant positive correlations between classification counts of individual tiles and between-user agreement rates. This suggests the validity of Linus's law in the user-generated data set and highlights the desirability of multiple classifications of the same geographic extents from different users when generating in-situ land cover data to achieve optimal quality.

We identify potential uncertainties in the user-generated as well as the authoritative data set using a confusion matrix. Our confusion matrices comparing StarBorn land cover classifications with CORINE showed that overall agreement was relatively high, but only for very common land cover classes. Mismatches between CORINE and StarBorn were common and often showed clear patterns, for example in the classification of tiles as *pasture* in StarBorn and *arable* in CORINE. This observation has been made before; for instance, Comber et al. (2016, p. 16) state that "it is important to consider and test for potential variations in the way that landscape features are labelled and conceptualised by different groups of contributors when analysing crowdsourced data." The data collected in StarBorn are thus not directly applicable as ground truth for CORINE, but rather illustrate how CORINE classes were interpreted on the ground by untrained observers. We argue that this has two important implications for those seeking to use crowdsourced data as an alternative to expensive, ground truth campaigns. First, if we wish to collect a gold standard, training and definitions with respect to the schema in use are necessary. However, such training does not fit well within the game-playing environment. Second, LBGs allow us to collect data that contribute to our understanding of how humans describe and partition land cover. Arguably, by comparing to authoritative data, we identify potential areas of semantic disagreement, which may in turn have implications for ways in which these authoritative data are used. We therefore suggest that such data may not, at least in the first instance, be best used in validation, but rather to consider whether existing products and definitions reflect common understandings of land cover (Egenhofer & Mark, 1995).

## 6 | CONCLUSIONS AND FURTHER WORK

In this article we demonstrated how an LBG can effectively be used to collect land cover data. Through the use of carefully implemented game elements, players can be motivated to provide land cover classifications of predefined spatial extents. Understanding patterns of contribution and factors influencing these are a first step towards collecting and analyzing crowdsourced land cover data. Although the usual caveats of crowdsourcing apply to LBGs, we believe imaginatively implemented, well-designed LBGs can complement and extend other existing efforts in crowdsourcing, and importantly can generate rich in-situ data. We particularly highlight that a game implies more than a rankings page and that an underlying narrative can be a useful element to guide user experience and foster motivation. In future work we intend to explore not only how LBGs can be used in validation efforts for land cover data, but also how they can be used to source natural language descriptions of landscape which might play a role in better understanding how land cover is perceived by non-experts.

## ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their comments and expertise. We would also like to thank the Geocomputation Group of the University of Zurich for their invaluable input. Finally, we would like to extend our gratitude to all persons who actively played the location based game, without you this work would not have been possible.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## ORCID

Manuel F. Baer  <https://orcid.org/0000-0002-9474-3299>

## REFERENCES

- Achard, F., Eva, H. D., Stibig, H.-J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.-P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, 297, 999–1002.
- Aldemir, T., Celik, B., & Kaplan, G. (2018). A qualitative investigation of student perceptions of game elements in a gamified course. *Computers in Human Behavior*, 78, 235–254.
- Andone, I., Blaszkiewicz, K., Böhmer, M., & Markowetz, A. (2017). Impact of location-based games on phone usage and movement: A case study on Pokémon GO. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 102:1–102:8). Vienna, Austria.
- Avouris, N., & Yiannoutsou, N. (2012). A review of mobile location-based games for learning across physical and virtual spaces. *Journal of Universal Computer Science*, 18, 2120–2142.
- Bayas, J. C. L., See, L., Fritz, S., Sturn, T., Perger, C., Dürauer, M., ... McCallum, I. (2016). Crowdsourcing in-situ data on land cover and land use using gamification and mobile technology. *Remote Sensing*, 8, 1–18.
- Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., ... Benford, S. (2006). Interweaving mobile games with everyday life. In *Proceedings of the 2006 SIGCHI Conference on Human Factors in Computing Systems* (pp. 417–426). New York, NY: ACM.
- Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2016). On predicting and improving the quality of volunteer geographic information projects. *International Journal of Digital Earth*, 9, 134–155.
- Brabham, D. C. (2012). Motivations for participation in a crowdsourcing application to improve public engagement in transit planning. *Journal of Applied Communication Research*, 40, 307–328.
- Brovelli, M. A., Molinari, M. E., Hussein, E., Chen, J., & Li, R. (2015). The first comprehensive accuracy assessment of Globeland30 at a national level: Methodology and results. *Remote Sensing*, 7, 4191–4212.
- Bubalo, M., Zanten, B. T. V., & Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landscape and Urban Planning*, 184, 101–111.
- Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, 57, 548–575.
- Burenhult, N., Hill, C., Huber, J., van Putten, S., Rybka, K., & San Roque, L. (2017). Forests: The cross-linguistic perspective. *Geographica Helvetica*, 72, 455–464.
- Cantalops, M. M., & Sicilia, M.-A. (2016). Motivations to read and learn in videogame lore: The case of league of legends. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 585–591). New York, NY: ACM.
- Celino, I. (2015). Geospatial dataset curation through a location-based game: Description of the Urbanopoly linked datasets. *Semantic Web*, 6, 121–130.
- Celino, I., Cerizza, D., Contessa, S., Corubolo, M., Dellaglio, D., Valle, E. D., & Fumeo, S. (2012). Urbanopoly: A social and location-based game with a purpose to crowdsource your urban data. In *Proceedings of the 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and the 2012 ASE/IEEE International Conference on Social Computing* (pp. 910–913). Los Alamitos, CA: IEEE Computer Society.
- Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4, 332–358.
- Colley, A., Thebault-Spieker, J., Lin, A. Y., Degraen, D., Fischman, B., Häkkinä, J., & Schöning, J. (2017). The geography of Pokémon GO: Beneficial and problematic effects on places and movement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1179–1192). New York, NY: ACM.
- Comber, A., Mooney, P., Purves, R. S., Rocchini, D., & Walz, A. (2016). Crowdsourcing: It matters who the crowd are. The impacts of between group variations in recording land cover. *PLoS ONE*, 11(7), e0158329.
- Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., & Ozdogan, M. (2014). Global land cover mapping: A review and uncertainty analysis. *Remote Sensing*, 6, 12070–12093.
- Côte, M., Wartmann, F., & Purves, R. (2018). Introduction: The trouble with forest: Definitions, values and boundaries. *Geographica Helvetica*, 73, 253–260.

- Cox, J., Oh, E. Y., Simmons, B., Graham, G., Greenhill, A., Lintott, C., & Woodcock, J. (2018). Doing good online: The changing relationships between motivations, activity, and retention among online volunteers. *Nonprofit and Voluntary Sector Quarterly*, 47(5), 1031–1056.
- Crall, A., Kosmala, M., Cheng, R., Brier, J., Cavalier, D., Henderson, S., & Richardson, A. (2017). Volunteer recruitment and retention in online citizen science projects using marketing strategies: Lessons from Season Spotter. *Journal of Science Communication*, 16(01), A01.
- Davidovic, N., & Stoimenov, L. (2013). Using location based game map signs to motivate VGI data collection related to traffic signs. In P. Mooney & K. Rehl (Eds.), *International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY)* (pp. 31–36). Leuven, Belgium.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268.
- Egenhofer, M. J., & Mark, D. M. (1995). Naive geography. In A. U. Frank & W. Kuhn (Eds.), *Spatial information theory: A theoretical basis for GIS. Lecture notes in computer science* (Vol. 988, pp. 1–15). Berlin, Germany: Springer.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). Menlo Park, CA: AAAI Press.
- European Environment Agency. (1994). CORINE documentation (Technical report). Retrieved from <http://www.eea.europa.eu/publications/CORO-landcover/page001.html>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874.
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137–148.
- Fonte, C. C., & Martinho, N. (2017). Assessing the applicability of OpenStreetMap data to assist the validation of land use/land cover maps. *International Journal of Geographical Information Science*, 31, 2382–2400.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185–201.
- Foody, G. M., & Boyd, D. S. (2012). Using volunteered data in land cover map validation: Mapping tropical forests across West Africa. In *Proceedings of the International Geoscience and Remote Sensing Symposium* (pp. 6207–6208). Piscataway, NJ: IEEE.
- Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS*, 17, 847–860.
- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., & Obersteiner, M. (2009). Geo-wiki.org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1, 345–354.
- Fritz, S., See, L., & Brovelli, M. (2017). Motivating and sustaining participation in VGI. In G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. Costa Fonte, & V. Antoniou (Eds.), *Mapping and the citizen sensor* (pp. 93–117). London, UK: Ubiquity Press.
- Gómez-Barrón, J.-P., Manso-Callejo, M.-A., Alcarria, R., & Iturrioz, T. (2016). Volunteered geographic information system design: Project and participation guidelines. *ISPRS International Journal of Geo-Information*, 5, 108.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69, 211–221.
- Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120.
- Gradinar, A., Huck, J., Coulton, P., Lochrie, M., & Tseklevs, E. (2015). Designing for the dichotomy of immersion in location-based games. In *Proceedings of the 10th International Conference on the Foundations of Digital Games*. Society for the Advancement of the Science of Digital Games.
- Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In D. Sui, S. Elwood, & M. F. Goodchild (Eds.), *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice* (pp. 105–122). Dordrecht, Netherlands: Springer.
- Haklay, M. M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' Law to volunteered geographic information. *Cartographic Journal*, 47, 315–322.
- Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, 71, 469–478.
- Hamari, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *Proceedings of the 47th Annual Hawaii International Conference on System Sciences* (pp. 3025–3034). Piscataway, NJ: IEEE.
- Hoe, D., Goh, L., Pa, E., Pe-Tham, P., & Lee, C. S. (2017). Perceptions of virtual reward systems in crowdsourcing games. *Computers in Human Behavior*, 70, 365–374.
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995–1006.
- Jeawak, S. S., Jones, C. B., & Schockaert, S. (2017). Using Flickr for characterizing the environment: An exploratory analysis. In *Proceedings of the 13th International Conference on Spatial Information Theory* (pp. 21.1–21.13). Wadern, Germany: Schloss Dagstuhl—Leibniz-Zentrum für Informatik.

- Ji, X., & Niu, X. (2014). The attribute accuracy assessment of land cover data in the National Geographic Conditions Survey. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2(4), 35–40.
- Kenny, R. F., & Gunter, G. A. (2007). Endogenous fantasy-based serious games: Intrinsic motivation and learning. *International Journal of Social Sciences*, 1, 591–596.
- King, D., Delfabbro, P., & Griffiths, M. (2010). Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health & Addiction*, 8, 90–106.
- Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., & Xu, J. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, 11, 261–269.
- Lei, Z., & Coulton, P. (2011). Using deliberate ambiguity of the information economy in the design of a mobile location based game. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 33–36). New York, NY: ACM.
- Lund, K., Lochrie, M., & Coulton, P. (2010). Enabling emergent behaviour in location-based games. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 78–85). New York, NY: ACM.
- Maisonneuve, N., & Chopard, B. (2012). Crowdsourcing satellite imagery analysis: Study of parallel and iterative models. In N. Xiao, M.-P. Kwan, M. F. Goodchild, & S. Shekhar (Eds.), *Geographic information science. Lecture notes in computer science* (Vol. 7478, pp. 116–131). Berlin, Germany: Springer.
- Matyas, S. (2007). Playful geospatial data acquisition by location-based gaming communities. *International Journal of Virtual Reality*, 6, 1–10.
- Matyas, S., Kiefer, P., & Schlieder, C. (2012). Changing the rules: Acquiring quality assured geospatial data with location-based games. In *Proceedings of the Fourth Annual Asian GAME-ON Conference on Simulation and AI in Computer Games (GAME-ON ASIA'2012)*, Eurosis. Kyoto, Japan. ISBN: 978-9077381-68-7.
- O'Hara, K. (2008). Understanding geocaching practices and motivations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1177–1186). New York, NY: ACM.
- Sahana, M., Hong, H., & Sajjad, H. (2018). Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India. *Science of the Total Environment*, 628–629, 1557–1566.
- See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., & Obersteiner, M. (2013). Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE*, 8(7), e69958.
- Sexton, J. O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., & Townshend, J. R. (2015). Conservation policy and the measurement of forests. *Nature Climate Change*, 6, 192.
- Stewart, O., Lubensky, D., & Huerta, J. M. (2010). Crowdsourcing participation inequality: A SCOUT model for the enterprise domain. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 30–33). New York, NY: ACM.
- Söbke, H., Hauge, J., & Andreea Stefan, I. (2017). Prime example ingress reframing the pervasive game design framework (PGDF). *International Journal of Serious Games*, 4(2), 182.
- Tulloch, A. I. T., & Szabo, J. K. (2012). A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu—Austral Ornithology*, 112, 313–325.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18, 306–314.
- Wang, H., & Sun, C.-T. (2011). Game reward systems: Gaming experiences and social meanings. In *Proceedings of the 2011 DiGRA International Conference: Think Design Play*. Hilversum, Netherlands: DiGRA / Utrecht School of the Arts.
- Wang, Z., & Ben-Arie, J. (1996). Conveying visual information with spatial auditory patterns. *IEEE Transactions on Speech and Audio Processing*, 4, 446–455.
- Willemse, I., Waller, G., & Süss, D. (2016). *JAMES: Jugend, Aktivitäten, Medien—Erhebung Schweiz*. Winterthur, Switzerland: ZHAW Zürcher Hochschule für Angewandte Wissenschaften.
- Winter, S., Richter, K.-F., Baldwin, T., Cavedon, L., Stirling, L., Duckham, M., & Rajabifard, A. (2011). *Location-based mobile games for spatial knowledge acquisition*, CEUR Workshop Proceedings. Retrieved from <http://hdl.handle.net/11343/32660>
- Wright, D. R., Underhill, L. G., Keene, M., & Knight, A. T. (2015). Understanding the motivations and satisfactions of volunteers to improve the effectiveness of citizen science programs. *Society & Natural Resources*, 28, 1013–1029.
- Yanenko, O., & Schlieder, C. (2014). Game principles for enhancing the quality of user-generated data collections. In *Proceedings of the AGILE Geogames and Geoplay Workshop*. Castellon, Spain.
- Zhang, N., Lee, Y., Radhakrishnan, M., & Balan, R. K. (2015). GameOn: P2P gaming on public transport. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (pp. 105–119). New York, NY: ACM.

**How to cite this article:** Baer MF, Wartmann FM, Purves RS. StarBorn: Towards making in-situ land cover data generation fun with a location-based game. *Transaction in GIS*. 2019;23:1008–1028.

<https://doi.org/10.1111/tgis.12543>