# A Crowdsourcing based game for land cover validation

**Abstract:** Land cover datasets are critical environmental information which are becoming increasingly available nowadays as open data. However, the problem is the accuracy of these freely available data for the particular application. The process of comparing one land cover dataset to another is considered one of the methods to validate the land cover data. But during this comparison, disagreement emerges between those data products. In order to validate the disagreement region between two datasets, an innovative game was built using a Human Computation technique named Game with a Purpose. The game was played during the FOSS4G Europe 2015 event by the conference participants. This paper introduces the issue of land cover validation and presents the data elaboration, the game design and the experimental results obtained.

**Keywords:** Land cover validation; Citizen Science; Human Computation; Game with a Purpose.

---

**1. Introduction**

Land cover is often a valuable information in many researches, including climate modelling, biodiversity monitoring, environmental and sustainable development. However, the usage of such data in many applications cannot neglect its validation and its classification accuracy.

Among the available land cover validation techniques, new methods involving Citizen Science are increasingly growing. The Land Cover Geo-Wiki project represents a well-known example. This Web-based application shows the disagreement areas between three global land cover maps and asks volunteers to identify the correct land cover class according to Google Earth images and their local knowledge [1,2]. Bastin et al. (2013) [3] also proposed an application to visually assess the uncertainty on land cover information at various levels: from a general rating of its confidence to the quantification of the proportions of land-cover types within a reference area.

Over the recent years Web-based crowdsourcing/gamified applications have emerged in land cover validation. The gaming approach allures players by providing them awards and gifts and the competition among the players increases their participation in the game. These reasons make the gaming approach more vital for decision making.

Among the existing applications, *Cropland Capture* [4] and *Picture Pile* [5] games proposed by Geo-Wiki project can be mentioned as the most notable examples. The C*ropland Capture* is a simple game in which players have to determine the presence of cropland on satellite imagery. The P*icture Pile* is another example in which players are asked to analyze two images of the same area related to different time periods and highlight the evidence of possible tree loss. These games are designed with scoreboards which displays the scores obtained during the gameplay. Every week the top three players in the leaderboard will be chosen as the weekly winners. At the end of the game, from these weekly winners, three people will be drawn randomly and will be awarded with gifts. Another recent example is the FotoQuest Europe project [6] which has been launched with the aim to obtain in-situ data on land cover and land use change for specific European point locations defined in LUCAS (Land Use and Coverage Area frame Survey) survey [7]. This initiative, which is an extension of FotoQuest Austria project [8], provides citizens with an app for taking pictures and collecting land cover information at specific point locations according to well-defined rules. During summer 2016, a survey campaign based on awards was launched to motivate students from European institutions to collect information for as many point locations as possible.

The present project fits within this context and proposes an innovative game based on the Game with a Purpose (GWAP) approach for land cover maps validation. This game has the objective to handle the disagreement pixels between land cover maps and validate them through the involvement of citizens. In a nutshell, it allows participants to visualize aerial photos on the disagreement areas and validate them pixel by pixel. The platform was tested during the FOSS4G Europe 2015 conference and allowed participants to evaluate the disagreement areas resulting from the comparison between the GlobeLand30 (GL30) and the DUSAF (Destinazione d'Uso dei Suoli Agricoli e Forestali) land cover maps on Como City (Lombardy Region, Italy).

The remainder of the paper has four parts. In Section 2 information about the two land cover maps exploited in the game, the benchmarking analysis performed between them and the generation of the disagreement pixels sample to be evaluated in the game are provided. Section 3 focuses on the design and the technical architecture of the game, while Section 4 presents the results obtained and a discussion on the main outcomes of the study. Finally, the conclusions in Section 5 present the main results and the future directions of the study.

## 2. Land cover data comparison and processing

### *2.1. Land cover datasets*

The crowdsourcing game platform has the objective to engage citizens in evaluating the disagreement areas detected between two land cover maps. For the case study of Como City, the selected maps are GL30 and DUSAF.

GL30 is the product of the "Global Land Cover Mapping at Finer Resolution" project led by the National Geomatics Centre of China (NGCC) and funded by the Chinese government. This dataset consists of two global land cover maps at 30 m resolution resulting from the classification of Landsat (TM and ETM+) and HJ-1 satellite images according to the Pixel-Object-Knowledge (POK)-based approach [9]. The two maps, which refer to the baseline years of 2000 and 2010 respectively, have been donated to the United Nations and released for open access and non-commercial use in 2014.

The GL30 data are available in raster format in World Geodetic System 1984 (WGS84) reference system and Universal Transverse Mercator (UTM) projection; the classification scheme is based on 10 major classes: *water bodies*, *wetland*, *artificial surfaces*, *cultivated land*, *forest*, *shrubland*, *grassland*, *bareland*, *permanent snow and ice*, and *tundra*.

DUSAF is a detailed geographical database created in 2000-2001 by Lombardy Region's Department for Territorial Planning, Agriculture and Forests (ERSAF) with the cooperation of the Lombardy Regional Agency for the Protection of the Environment (ARPA) [10]. The database provides five vector datasets at 1:10000 scale information describing the land cover of Lombardy Region for different temporal periods (2000, 2005-2007, 2007, 2009, 2012). Data are freely available on the Lombardy Geoportal (http://www.geoportale.regione.lombardia.it/en/home) and are provided in WGS84-UTM32 reference system. The land cover is classified in a hierarchical way and it is based on five levels. The first three levels comply with the CORINE Land Cover (CLC) nomenclature and the most general one consists of 5 classes: *artificial cover*, *agricultural areas*, *forest and semi natural areas*, *wetlands*, and *water bodies*.

For the current study, the DUSAF datasets referred to 2000 (named DUSAF 1.1) and 2012 (named DUSAF 4.0) were selected as the most appropriate for the comparison with GL30 2000 and GL30 2010, respectively. To allow the comparison between the datasets, some preprocessing steps for format and legend homogenization were performed, i.e. the GL30 datasets were reclassified according to the first level of the DUSAF legend (Table 1) while the DUSAF data were rasterized according to the resolution of the GL30. Figure 1 shows the two datasets related to Como City.

**Table 1.** Correspondence between DUSAF and GlobeLand30 land cover classes according to the first level of the CORINE Land Cover nomenclature.

| DUSAF legend | GLOBELAND30 legend |
|---|---|
| Artificial surfaces | Artificial cover |
| Agricultural areas | Cultivated land |
| Forest and semi natural areas | Forest, shrubland, grassland, bareland, tundra, permanent ice or snow |
| Wetlands | Wetlands |
| Water bodies | Water bodies |

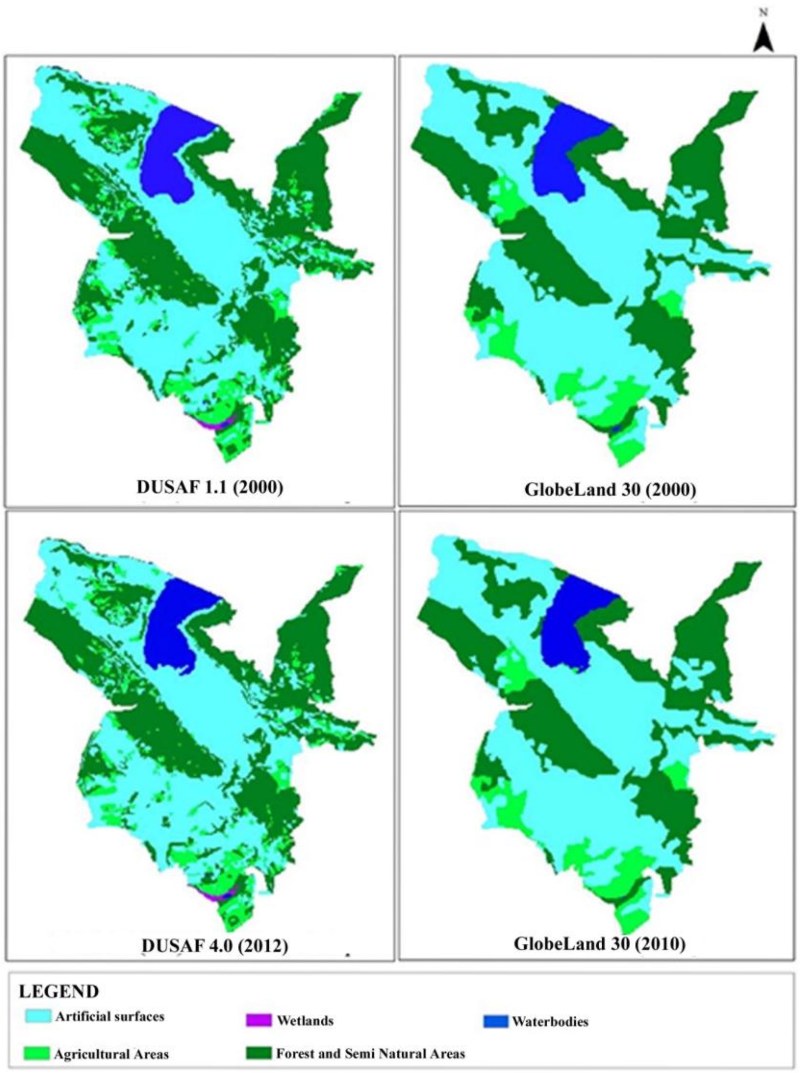


**Figure 1.** GlobeLand30 and DUSAF datasets related to Como city.

*2.2. Dataset comparison and disagreements detection*

To evaluate the accuracy of the GL30 classification a comparison with the more detailed DUSAF data was performed according to the confusion matrix or error matrix approach [11], which is suggested by Foody (2011) [12] as the "good practice" in LC classification accuracy assessment. The confusion matrix is the result of a spatial comparison between a classified map, i.e. the dataset to be evaluated (GL30), and a reference map, i.e. the ground truth (DUSAF). Many indexes assessing the classification quality can be derived from the confusion matrix; among them, the most commonly used are the overall accuracy ($OA$) and, for each LC class $i$, the user accuracy ($UA_i$) and the producer accuracy ($PA_i$). The $OA$ identifies the percentage of correctly classified pixels. The $UA_i$ is the percentage of the classified pixels that exactly match the ground truth, whereas the $PA_i$ is the percentage of pixels of the ground truth which are correctly detected in the classified map.

For the current study the comparison between GL30 and DUSAF datasets was performed pixel by pixel for both the two considered temporal periods. As reported in Table 2, in both cases the $OA$ index highlights an agreement degree slightly higher than 80%; the number of disagreement pixels is about 8000 out of a total of about 41000 pixels. Table 3 shows $UA_i$ and $PA_i$ values for each considered LC class $i$. The agreement degree is particularly high for the *water bodies* class where $UA_i$ and $PA_i$ values are higher than 90%; neglecting the *wetland* class, which represent a very small percentage (0.25%) compared to the total area of interest, the most significant problems are detected for the *cropland* areas, where $UA_i$ and $PA_i$ values are equal or lower than 50%. The spatial distribution of the disagreement pixels on Como city is shown in Figure 2.

**Table 2.** Comparison between GlobeLand30 and DUSAF datasets: number of agreement and disagreement pixels, overall accuracy ($OA$) and disagreement percentages (for years 2000 and 2010).

| | **Agreement pixels** | **Disagreement pixels** | **OA [%]** | **Disagreement [%]** |
|---|---|---|---|---|
| GL30 2000 - DUSAF 1.1 | 33416 | 8022 | 80.60 | 19.40 |
| GL30 2010 - DUSAF 4.0 | 33318 | 8120 | 80.40 | 19.60 |

**Table 3.** Comparison between GlobeLand30 and DUSAF datasets: user accuracy ($UA_i$) and producer accuracy ($PA_i$) values for years 2000 and 2010.

| | **GL30 2000 - DUSAF 1.1** | | **GL30 2010 - DUSAF 4.0** | |
|---|---|---|---|---|
| | **$UA_i$ [%]** | **$PA_i$ [%]** | **$UA_i$ [%]** | **$PA_i$ [%]** |
| Artificial cover | 78.76 | 90.88 | 80.23 | 88.55 |
| Cropland | 50.34 | 43.34 | 47.36 | 43.05 |
| Forest and semi natural areas | 87.01 | 76.55 | 85.72 | 78.56 |
| Wetland | 0.00 | 0.00 | 0.00 | 0.00 |
| Water bodies | 97.32 | 94.89 | 97.41 | 92.27 |

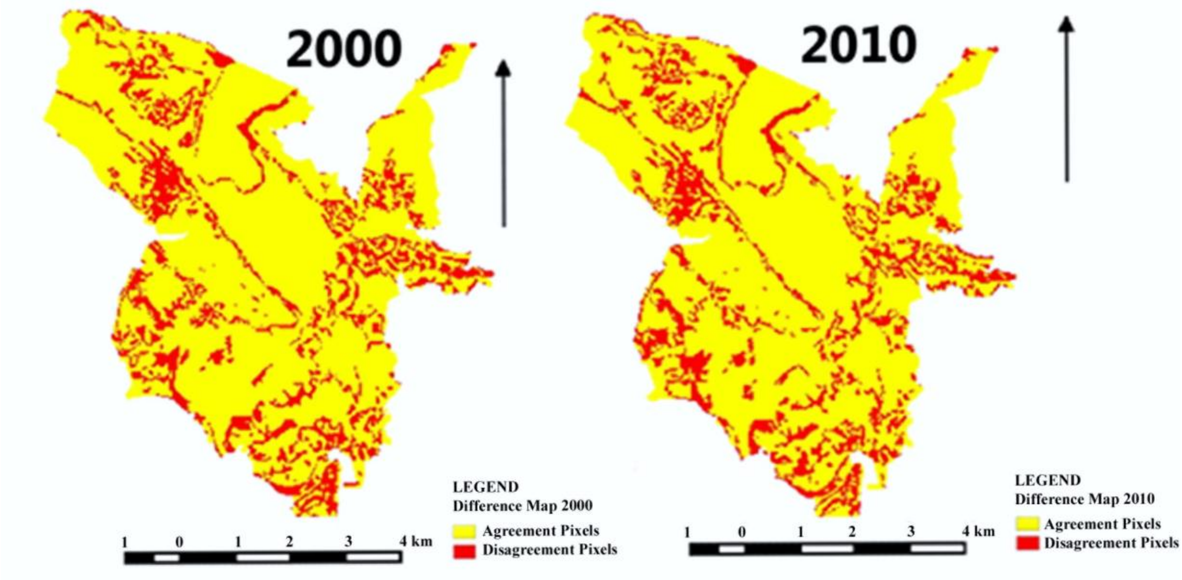


**Figure 2**. Agreement (in yellow) and disagreement (in red) pixels between GlobeLand30 and DUSAF datasets for the two considered temporal periods.

Although DUSAF is certainly the most appropriate dataset to be selected as reference, it derives from the classification of aerial photos and it may also contain errors. For this reason, a more in-depth analysis on the disagreement pixels has been performed in order to identify which among the two datasets contains the correct LC classification. To solve this task, the authors decided to ask volunteers to validate the disagreement pixels by promoting the Land Cover Validation Game.

Being the validation of more than 8000 pixels a tedious and time consuming process, a sampling design was performed to select a significant subset of pixels on which to perform the evaluation. Choosing a sampling design requires a consideration of the specific objectives of the accuracy assessment and a prioritized list of desirable design criteria. The most critical recommendation is that the sampling design should be a probability sampling design. For the present study, the random stratified sampling proposed by Cochran (1977) [13] was selected. The sample size *n* is computed according to the formula proposed in Equation 1:

$$n = \frac{(\sum W_i S_i)^2}{[S(O)]^2 + (\frac{1}{N}) \sum W_i S_i^2} \tag{1}$$

where *N* is the total number of pixels in the area of interest, *S(O)* is the standard error of the estimated overall accuracy we would like to achieve. Since we expect a confidence level of 99%, *S(O)* has been set to 0.01. $W_i$ is the mapped proportion of area of class *i* (it is the proportion between the number of pixels in the *i-th* class and the total number of pixels), while $S_i$ is the standard deviation of stratum (class) *i* and it is computed from the user accuracy $UA_i$ as shown in Equation 2:

$$S_i = \sqrt{(UA_i * (1 - UA_i))} \tag{2}$$

According to the selected sampling design, the computed samples sizes *n* for the years 2000 and 2010 are 1340 and 1208 pixels, respectively. As will be explained in the following section, the validation process of each pixel requests the fulfillment of specific requirements that in some cases could not be reached; for this reason, the sample size was incremented of about 20% in such a way to obtain the validation of the required number *n* of pixels and complete the game. The subset was randomly extracted from the disagreement pixels' original dataset.

## 3. Gaming for Citizen Science

It was observed that people all over the world spend billions of hours in playing computer games each year [14]. Suppose people, while playing computer games get to solve real world problems on the basis of their knowledge and capabilities. There are indeed examples of problems which cannot yet be effectively solved by computers, like for example people identification in pictures. What if we utilize people as "human processors" in a distributed system in which each individual accomplishes a small part of a massive computation? This kind of approach is named Human Computation [15].

Scientific endeavours demand data of very high quality. Prestopnik & Crowston (2011) [16] demonstrated that Citizen Science efforts often produce high quality data, but the open question remains on how Citizen Science projects can attract participants and make them stay involved. Indeed, people may need some sort of encouragement or motivation to be involved and execute tasks. In literature, we can distinguish between extrinsic rewards (e.g. payments, tangible prizes) and intrinsic rewards (e.g. personal recognition and satisfaction, passion). Among intangible rewards, games emerged as a method for attracting people to participate in solving a task and thus using "brain power" to address an open problem.

The research work presented in this paper refers to a specific Human Computation technique called Game with a purpose (or GWAP) which relies on the human desire to be entertained [14]. A GWAP is a fully-fledged gaming application whose users are players motivated by the fun and enjoyment of the game itself (e.g., points, badges, leaderboards). The computational task that a GWAP players are asked to solve is usually "hidden" or embedded within the gameplay. Of course, the task should be simple enough to be made part of a fun challenge.

### *3.1. Design of the game*

The Land Cover Validation Game (Figure 3) is a simple casual game in which each player is displayed with a blue square box (a "pixel" of 30 m resolution) which is placed above high resolution aerial photos. The pixels correspond to non-coherent pixels (i.e.) which are disagreements between two classifications (DUSAF and GL30). The players are asked to choose an appropriate category of LC classification from the five categories displayed for the given pixel (Figure 3). The five categories are the first level of CLC nomenclature, thus for each pixel they include both the result of DUSAF classification and the result of GL30 classification.

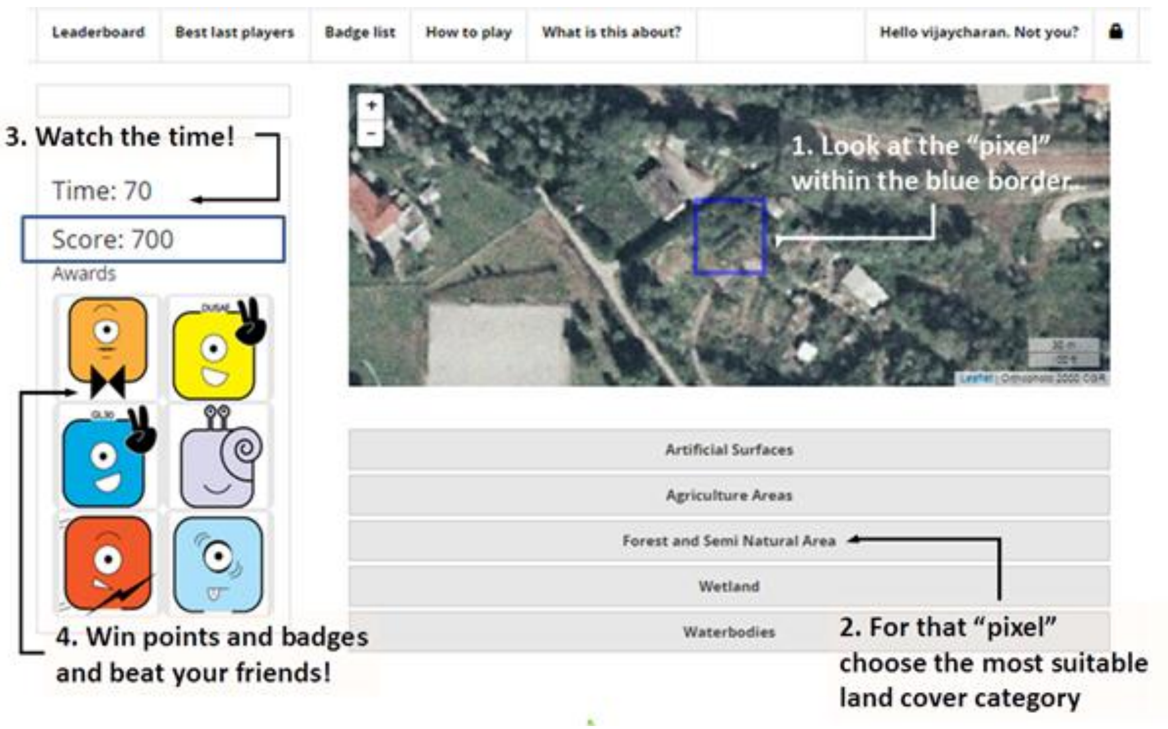


**Figure 3.** Land Cover Validation Game screenshot explaining the gameplay.

The rationale is that while DUSAF and GL30 were produced from the automatic processing of aerial photographs through machine vision techniques that may fail, human users looking at the same photographs can rely on their cognitive capabilities to classify land cover and thus they can help in validating those pixels for which the two automatic classification produced non-coherent results.

In the gameplay, players score if they agree with one of the existing classification (DUSAF or GL30) and do not score if they choose any of the other three categories. The players also get more points for consecutive agreements.

The responses from the players are then processed and aggregated [17] in order to achieve the purpose of validating the non-coherent pixels; intuitively, the more players agree on a category, the more probable that the pixel actually belongs to that land cover category. More specifically, each pixel-category pair has an associated confidence score which is incremented every time a player selects that category for that pixel during gameplay (or decremented if another category is selected). Once the confidence score overcomes a specific threshold, the pixel-category pair is considered correct and it is no more given to players.

In mathematical terms, the player responses aggregation works as follows. Each pixel $p$ has a score for each land cover category $c$, so that the score is $\sigma(p,c) \in [0,1]$, where zero means that the association between the pixel and the category is false, while one means that the association between the pixel and the category is true. At the beginning of the gameplay, we do not know which pixel-category link is "true", therefore the corresponding link scores are initialized with a value in the range of possible scores.

Whenever a player selects a land cover category for a pixel (say, pixel $p_i$ is associated to category $c_j$), the pixel-category links are updated according to the following Equations (3) and (4):

$$\sigma(p_i, c_j)_{t+1} = \sigma(p_i, c_j)_t + \Delta_{inc} \cdot \rho_{player} \quad (3)$$

$$\sigma(p_i, c_k)_{t+1} = \sigma(p_i, c_k)_t - \Delta_{decr} \cdot \rho_{player}, \text{for each } k \neq j \tag{4}$$

where the scores are updated at time *t+1* by taking the respective values at time *t* and incrementing or decrementing them on the basis of user selection: the association with the selected category is increased, while the associations with the other land cover categories are decreased. In the above formulas, $\Delta_{inc}$ and $\Delta_{decr}$ are respectively the increment and decrement factors, while $\rho_{player}$ is the player's reliability, i.e. the reputation mechanisms that takes into account the fact that some users may be more accurate than others; reliability is computed on the basis of the number of "mistakes" they make during a game round, where a mistake happens whenever a player does not select the DUSAF category not the GL30 category, as in the following Equation (5) (cf. also [16]):

$$\rho_{player} = e^{-mistakes/2} \tag{5}$$

Since in each game round, a player is asked to classify six pixels, the number of mistakes is in [0,6], thus the player's reliability varies between 1 (no mistakes, perfect reputation) to 0.05 (only mistakes, worst reputation). In this way, we both take into account the fact that a player is not reliable because he is just giving random answers and the possible inherent difficulty of the pixel classification.

The pixel-category link scores are therefore updated over time according to players' choices; once a score overcomes a given threshold $\sigma(p_i, c_j) \geq \tau$ we consider the pixel "validated", i.e. the pixel $p_i$ is of category $c_j$ according to the game players. The values of $\Delta_{inc}$, $\Delta_{decr}$ and $\tau$ are empirically estimated so that it needs at least three players with perfect reputation to validate a pixel. As we will discuss in the following sections, given the varying reliability of players, in most cases three players were enough for validation.

*3.2. Technical architecture of the game*

Figure 4 illustrates the Land Cover Validation Game architecture. The two different repositories are used to store base map layer and data products. The first repository (Geoserver) contains the high resolution aerial photos of Como City which were kindly provided by Blom CGR S.p.a for the two baseline years 2000 and 2010. These datasets are displayed in the game through a Web Map Service (WMS) by making use of the Leaflet JavaScript library. The second repository (MySQL Database) contains non-coherent pixels coordinates and information about players, their contributions, scores and badges.

The server on which the game is deployed runs on Ubuntu machine and makes use of open source software. The Apache Web server (version 2.2.22) hosts the PHP services to access the database and the WMS service. The client application is a single-page application based on the AngularJS framework. Each time a player classifies one particular pixel, the client application displays the following pixel on top of the aerial photo.

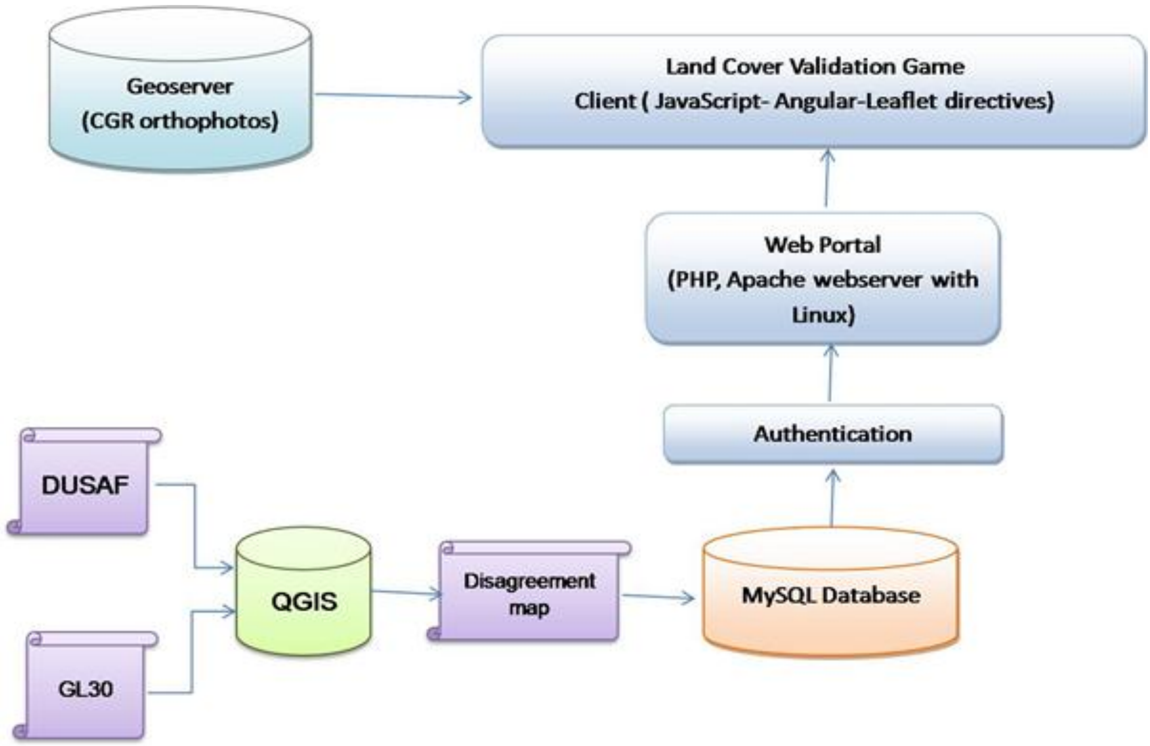


**Figure 4.** Architecture of the Land Cover Validation Game.

## 4. Results and Discussions

The Land Cover Validation Game was one among the four mapping parties during the FOSS4G Europe 2015 conference. It was introduced on the first day of the conference (15th July) to the participants and was played during the whole event with the aim to evaluate the disagreement between GL30 2000 and DUSAF 1.1. At the end of the conference 75% of the game was completed, so the game was made available online for some more days for the players to play till the game reaches 100% of completion. Table 4 illustrates the main statistics of the game, explained in the following.

**Table 4.** Evaluation of Land Cover Validation Game.

| Quantity | Value | Unit of measure |
|---|---|---|
| N° Of Players | 68 | |
| Total Played Time | 0 20:17:08 | d HH:mm:ss |
| N° Of Validated Pixels | 1600 | # |
| N° Of Played Pixels | 1600 | # |
| Total N° Of Pixels | 1600 | # |
| Completion rate | 100 | % |
| Throughput | 78.92 | solved tasks/hour |
| Average Life Play (ALP) | 17.9 | minutes/player |
| Expected contribution | 23.54 | solved tasks/player |
| DUSAF Agreements | 86.82 | % |
| GL30 Agreements | 11.87 | % |
| Disagreements | 1.31 | % |

In order to evaluate the game's enjoyability and effectiveness, the main GWAP metrics were computed as defined in Von Ahn (2006) [13]. Throughput is the average number of problem instances solved per human-hour which gives the effectiveness of the game. The Average Lifetime Play (ALP) gives the enjoyability of the game and is computed as the overall amount of time the game is played by all players divided by the total number of players who played it. Expected contribution is the concise measure of GWAP effectiveness and is computed by multiplying throughput and ALP.

For the Land Cover Validation Game, the throughput is 78.92 solved tasks per hour, which is a good result. ALP is 17.90 minutes per player. The expected contribution is 23.54 solved tasks per player which means that on average each player helped in validating 23.54 pixels.

Regarding the actual resulting land cover classification, in 86.82% of cases players agreed with the DUSAF category, in 11.87% of the cases with GL30 category and only in 1.31% of the cases with neither of them.

An additional analysis was performed to understand how many classifications were completed during the mapping party in FOSS4G and after that, as displayed in Figure 5. It is clear that 72% of the classification were achieved during the event.

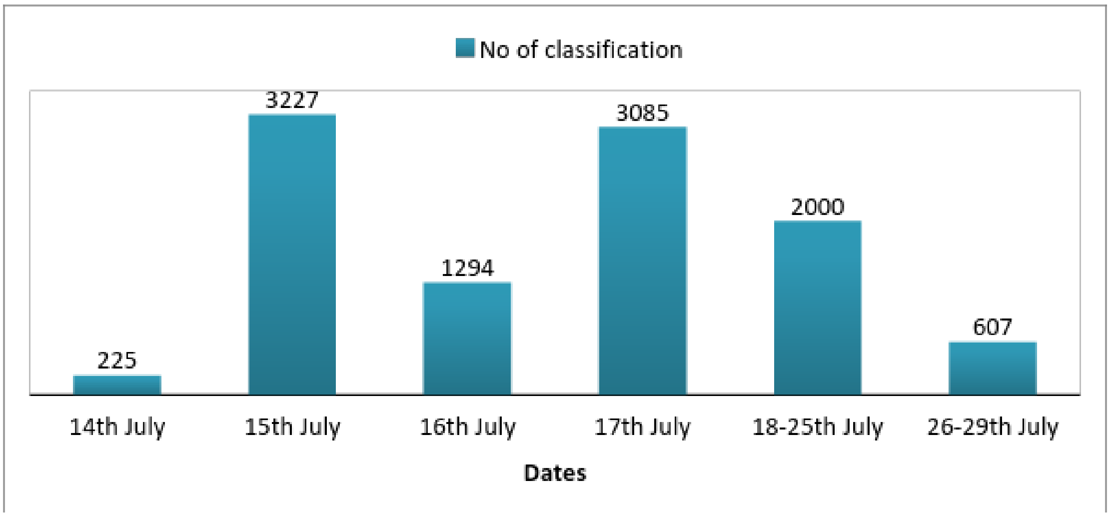


**Figure 5.** Number of classification for different dates.

Figure 6 shows the number of times each pixel was classified; intuitively, this corresponds to the difficulty of the classification task. In order to consolidate a pixel-category pair - i.e. making its confidence score overcome a threshold - a minimum of three players with 100% reputation were required by design; the more players required to consolidate the land cover, the more controversial the pixel classification. It is clear that 1300 pixels (>80% of the total) required a limited number of players, thus demonstrating that this validation task is actually amenable of being performed by human "processors".

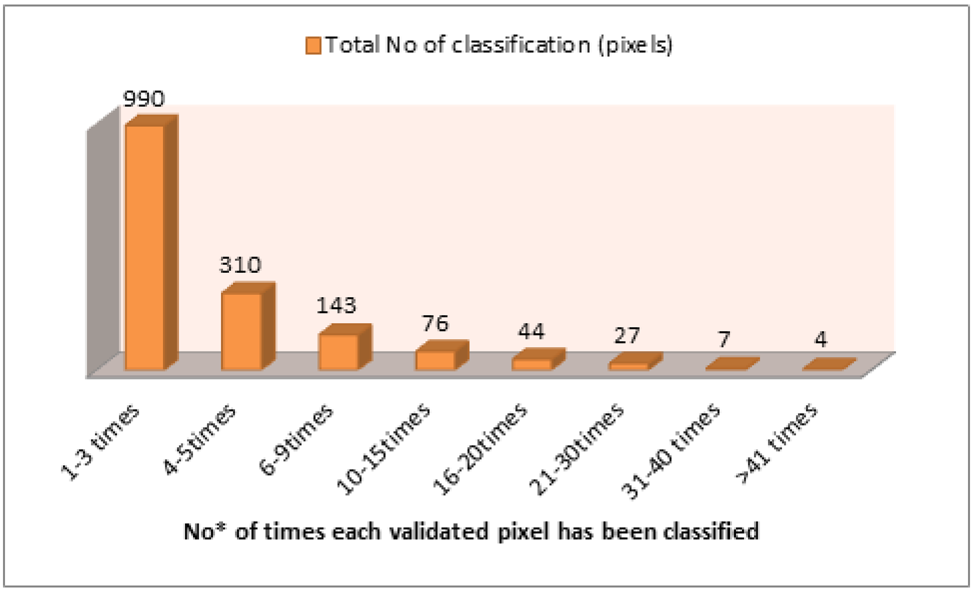


**Figure 6.** Number of times each validated pixel has been classified.

## 5. Conclusion

This research work illustrated an experiment to classify land cover by involving human users (with little background knowledge and no training) in visually elaborating fine-resolution aerial photos in a Web-based game. In this paper, we presented the design choices, both in terms of data pre-processing and in terms of game design, and the experimental results in validating the disagreement of two land cover classifications – DUSAF and GL30 – in a defined area. By showing the results obtained in quite a short time frame, we also demonstrated the feasibility of our approach and the potentiality of gaming and entertaining incentives in user engagement: indeed, we proposed an innovative method to attract ordinary citizens in a land cover validation campaign. We also explained how we addressed the issue of varying reliability of participant player in the contributions' aggregation. The output of the Land Cover Validation Game is a manually-annotated dataset with the land cover for each pixel, as identified by game players and aggregated through our cross-validation algorithm, and a set of evaluation metrics.

Looking at the empirical classification results obtained for the comparison between DUSAF 1.1 and GL30 2000, we noticed that game players were generally more in agreement with the DUSAF classification, rather than with the GL30 classification. Since the DUSAF is characterized by a greater level of detail, this result was expected. Furthermore, the land cover categories used in the game correspond to the first level of CLC nomenclature, i.e. the classification adopted by DUSAF, therefore the result could have been influenced by a non-optimal matching between GL30 categories and the CORINE land Cover classes; finally, the DUSAF classification is a "local" effort by the public bodies of the selected region, while GL30 is an international effort, thus the difference in classification could have been motivated by a different background knowledge included in the production of the two disagreeing classifications. Naturally, the diversity of our game classification can also be motivated by the inherent complexity of the classification task for human users; still, the very low share of game classifications that do not match either DUSAF not GL30 (less than 1.5%) can be interpreted as a sign of the effectiveness of our approach, besides the possible subjectivity of human classification.

Currently, the validation of disagreement pixels between DUSAF 4.0 and GL30 2010 is still in progress. Interested readers can participate at the game at the following link http://landcover.como.polimi.it/landcover/#/

**Acknowledgments:** This work was partially funded by Lombardy Region as part of the PROACTIVE project (Id 40723101, Lombardy Region POR-FESR 2007-13). The authors would also like to thank Blom CGR S.p.a for providing the high resolution photos used in the game.

## References

1. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; Grillmayer, R.; Achard, F.; Kraxner, F.; Obersteiner, M. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sens.* 2009, 1 (3), 345–354.
2. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; See, L.; Schepaschenko, D.; Velde, M. V.D.; Kraxner, F.; Obersteiner, M. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software* 2012, 31, 110–123.
3. Bastin, L.; Buchanan, G.; Beresford, A.; Pekel, J.F.; Dubois, G. Open-source mapping and services for Web-based land-cover validation. *Ecological Informatics* 2013, 14, 9-16.
4. Salk, C.F.; Sturn, T.; See, L.; Fritz, S.; Perger, C. Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game. *International Journal of Digital Earth* 2016, 9(4), 410-426.
5. Engaging Citizens in Environmental Monitoring, Picture Pile. Available Online: http://www.geo-wiki.org/games/picturepile/ (accessed on 30 July 2016).
6. FotoQuest Europe. Available Online: http://www.fotoquest-europe.com/ (accessed on 30 July 2016).
7. Delincé J. A European approach to area frame survey. Proceedings of the Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR), Rome, Italy, 5-7 June 2001.
8. Laso-Bayas, J.C.; See, L.; Fritz, S.; Sturn T.; Karner, M.; Perger, C.; Dürauer, M.; Mondel, T.; Domian, D.; Moorthy, I.; et al. Assessing the quality of crowdsourced in-situ land-use and land cover data from FotoQuest Austria application. Proceedings of the European Geosciences Union (EGU) General Assembly, Vienna, Austria, 17-22 April 2016.
9. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogram. Remote Sens*. 2015, 103, 7-27.
10. Credali, M.; Fasolini, D.; Minnella, L.; Pedrazzini, L.; Peggion, M.; Pezzoli, S. Tools for territorial knowledge and government. In Land Cover Changes in Lombardy over the Last 50 Years; Fasolini, D., Pezzoli, S., Sale, V.M., Cesca, M., Coffani, S., Brenna, S., Eds.; ERSAF-Lombardy Region: Milan, Italy, 2011; pp. 17-19.
11. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; Lewis Publishers: Boca Raton, FL, USA, 1999; p. 137.
12. Foody, G.M. Classification accuracy assessment. *IEEE Geosci. Remote Sens. Soc. Newsl*. 2011, 159, 8-14.
13. Cochran, W. G. *Sampling techniques*, 3rd ed. New York: John Wiley & Sons, USA, 1977; p. 448.
14. Von Ahn, L. Games with a purpose. *Computer*, 39(6), 92-94.
15. Law, E.; Von Ahn, L. Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*; Brachman, R., Cohen, W.W., Stone, P., Eds.; Morgan & Claypool Publishers, 2011, 5(3), 1-121.
16. Prestopnik, N. R.; Crowston, K. Gaming for (citizen) science: exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. Proceedings of the 7th IEEE International Conference on e-Science - "Computing for Citizen Science" Workshop, Stockholm, Sweden, 5-8 December 2011.
17. Celino, I.; Contessa, S.; Corubolo, M.; Dell'Aglio, D.; Della Valle, E.; Fumeo, S.; Krüger, T. Linking smart cities datasets with human computation–the case of urbanmatch. Proceedings of the 11th International Semantic Web Conference (ISWC), Boston, MA, USA, 10-15 November 2016.