Ref. Ares(2023)3769445 - 31/05/2023

HORIZON EUROPE

HORIZON-CL4-2021-DIGITAL-EMERGING-01
AI, Data and Robotics for the Green Deal (IA)

AI-powered Robotic Material Recovery in a Box

RECLAIM

D6.2: Algorithms and pipelines for Recycling Data Games

Contractual Date of Delivery:	31/05/2023
Actual Date of Delivery:	31/05/2023
Security Class:	Public
Editor:	*Antonios Liapis (UM)*
Contributors:	Konstantinos Sfikas (UM), Iro Voulgari (UM)
Quality Assurance:	*Michail Maniadakis (FORTH)*
Deliverable Status:	Final

The *RECLAIM* Consortium

Part. No.	Short Name of Participant	Participant Organization name	Country
1	FORTH	Foundation for Research and Technology Hellas	EL
2	UoM	University of Malta	MT
3	KUL	Katholieke Universiteit Leuven	BE
4	HERRCO	Hellenic Recovery Recycling Corporation	EL
5	IRIS	Iris Technology Solutions, Sociedad Limitada	SP
6	RBNS	ROBENSO PC	EL
7	AIMPLAS	AIMPLAS - Technological Institute of Plastics	SP
8	AXIA	Axia Innovation UG	DE
9	ISWA	International Solid Waste Association	NL
10	ION	Periferiakos Foreas Diaxirisis Stereon Apovliton Ionion Nison Anonimi Eteria Ton Ota	EL

Document Revisions

Version	Date	Editor	Overview
0.1	29/04/2023	Antonios Liapis (UM)	Structure and Early Draft
0.8	16/05/2023	Konstantinos Sfikas (UM), Iro Voulgari (UM)	Content additions on mini-games and educational content
0.9	17/05/2023	Antonios Liapis (UM)	Final Edits and Typesetting
0.95	29/05/2023	Michail Maniadakis (FORTH)	Comments and corrections
1.0	29/05/2023	Antonios Liapis (UM)	Corrections applied, final version

Table of Contents

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
DoA	Description of the Action
RDG	Recycling Data Game
prMRF	portable, robotic Material Recovery Facilities
RoReWo	Robotic Recycling Workers
GDPR	General Data Protection Regulation
IP	Intellectual Property

Executive Summary

RECLAIM is a Horizon Europe funded project with an objective to **develop** a portable, robotic Material Recovery Facility (**prMRF**) tailored to small-scale material recovery. RECLAIM adopts a modular multi-robot/multi-gripper approach for material recovery, based on low-cost Robotic Recycling Workers (RoReWos). An **AI module** combines imaging in the visual and infrared domain to identify, localize and **categorize recyclables**. The output of this module is used by a multi-RoReWo team that implements efficient and accurate material sorting.

Further, RECLAIM englobes a citizen science approach to increase social sensitivity to the Green Deal. This is accomplished via a **novel Recycling Data-Game** that enables and encourages citizens to participate in project RTD activities by providing annotations to be used in **deep learning** for the **re-training of the AI module**. Three different scenarios will attest its effectiveness and applicability in a broad range of locations that face material recovery challenges.

This deliverable reports on the design and development of the recycling data game (RDG), in light of its several goals (annotating waste data, increasing awareness, etc.). The report is in part capturing the game design, discussing the concepts of the intended player experience envisioned in the final version of the RDG, and in part documentation of the development efforts and working prototypes for the RDG thus far. The report also touches upon identified topics that must be integrated and made accessible to players as part of the RDG experience in terms of educational and information content. In terms of algorithms, the report lists certain methods identified as promising and feasible for upcoming developments on validating user data in terms of their accuracy. Finally, the report concludes with future work on developing several of the designed features described in the report, as well as integration work with other tasks of RECLAIM.

1. Introduction

At its core, RECLAIM proposes the development of a low cost, portable, easy to install and increased productivity prMRF that can achieve full material recovery anywhere, even in the most remote areas. The developed prMRF is expected to have a key role in developing a global, leakage-free circular economy model benefiting businesses, the society, and the environment.

However, we do not consider that a circular economy is only limited to automating material recovery. With the complementary RECLAIM pillar (PIL-4) for *Environmental gaming for social awareness and data collection*, we envision that **data can also form a positive feedback loop and be re-used in a circular fashion**. Therefore, recycling data games (RDG) are proposed as a novel approach introduced by RECLAIM to enrich collected waste data with users' own feedback and thus improve the AI algorithms. In turn, better algorithms can filter which collected data is most ambiguous and thus relevant for users' feedback, achieving a self-sustaining (assuming user engagement) cycle of data re-use. Beyond this, the RDG also aims to improve the users' own views on (and awareness of) the circular economy through play.

Due to the ambitions of PIL-4, the envisioned RDG has several goals (at times orthogonal). We list the identified goals based on both the DOA and discussions with focus groups of end-users and stakeholders conducted as part of WP2. We sort these goals based on their importance and, relatedly, based on how challenging they are to achieve:

1. Collect of user annotations for waste material, useful for AI algorithm training
2. Provide feedback to users acknowledging their contribution in scientific work
3. Implement "fun" games
4. Integrate quiz questions on the reduction of CO_2 emissions in relation to material recycling
5. Familiarize users with the recycling processes followed in European countries
6. Communicate the general principles of AI and Data Science to the public
7. Promote the results of the project to the public (as part of exploitation & innovation)

Due to the many goals of the RDG, it is neciessary to design the experience in a way that accomplishes as many of these goals as possible seamlessly. The work undergone in T6.2 and reported in this early deliverable has been multi-pronged, aiming to ensure that a refined user experience is defined before developing the final game, while also testing interface and technical concerns in functional prototypes for standalone challenges, and collecting data for content shown to players while interacting with the game. The sections below will elaborate on outcomes and current state of all these activities, and offer a roadmap for future work to integrate them in a well-designed and impactful recycling data game (in Section 6).

1.1 Intended readership

The present report is a public (PU) document. Its readership is considered to be the European Commission, the RECLAIM Project Officer, the partners involved in the RECLAIM Consortium, beneficiaries of other European funded projects, and the general public.

1.2 Relationship with other RECLAIM deliverables

The game design and first prototype mini-games described in this report are informed by the end-users feedback collected via surveys and focus group discussions under WP2. Throughout this report, we will be referencing D2.1 regarding key design decisions taken as part of that data collection process and integrated in the current game design. Since the games will expose to the public the AI algorithms' outputs, and contribute to them with new user-provided ground truth data, this deliverable is strongly linked to WP3 (Recyclable Waste Detection and Categorization). The image data used in current prototypes of this report, and used by RDGs moving forward, are collected as part of D6.1 submitted concurrently to this deliverable. Table 1 shows the main deliverables consulted (in case of past work), and impacted by (in case of future work) by this report.

Table 1: Other RECLAIM deliverables related.

Del. No	Deliverable Name	WP	Month
1.1	Data management plan and ethics/privacy manual	WP 1	M6/M36
2.1	prMRF and RDG requirements and systems specification	WP 2	M6
3.1	Material recognition based on RGB and Hyperspectral imaging	WP 3	M18
3.2	prMRF operation monitoring and repeating advancement	WP 3	M30
6.1	Waste Data for material recognition and Recycling Data Game	WP 6	M9/M18
6.1	Waste Data for material recognition and Recycling Data Game	WP 6	M9/M18
6.3	Assessment of the Recycling Data Game	WP 6	M18/M36
1.3	Final Project Report	WP 1	M36

2. Game Design

As part of the collection of requirements for RDGs, three focus group interviews were conducted with stakeholders and end-users, and a general survey was disseminated in citizen science portals for broader end-user feedback. As reported extensively in D2.1, the conclusions of the requirements analysis pointed towards an ideal RDG interaction on mobile devices which players can interact with during downtime for short periods of time (approximately 10 minutes). Among the important game mechanics or game design features that the focus groups identified were (a) a progression system, (b) rewards, (c) savegames, (d) intellectual challenges, (e) strategy elements, (f) competition/collaboration.

In light of the above requirements, and drawing inspiration from mature gamified platforms such as DuoLingo [1], a game design document has been drafted and deployed online as a living document for the upcoming development efforts. This living document is available to (and editable by) all members of the design and development team at University of Malta, and also serves as a way of communicating with other RECLAIM partners and external stakeholders on the vision and state of the RDG. Below we summarize the main concepts described in the game design document, in order to frame the aspects developed so far (Section 3).

In terms of the high concept of the Recycling Data Game, it will consist of a series of independent ***challenges*** shown as a partial ***roadmap***. Missions will be *grouped and ordered* according to the user's ***level***. There will be no experience points or resource requirements for increasing the user's level, the level increases by completing all challenges in the current roadmap. Once all challenges in the current roadmap are completed, the user will increase in level, and a new roadmap will be shown to them. The rationale of a level system is for *onboarding*, in order to show only a few simple challenges and "basic" or "vital" information content to the player in the beginning, giving them the necessary skills to complete more challenging tasks.

For example, the first level will have two challenges, one merely informing the user on what PET bottles are, a single annotation challenge (with simpler and faster interactions, such as "Detect" PET bottles) and a simple quiz regarding recycling of PET bottles for consumers. The next level will have three challenges, with a more complex "Count" challenge for PET bottles using the same data as the previous level's challenge, while expanding on information for other plastic materials (which will be annotated in future levels) and a quiz on facts for recycling PET bottles from information provided so far. This will allow the user to slowly learn how to use the interface, how to "read" the images, and facts about recycling without being overwhelmed.

In terms of rewards, the player will collect resources for completing tasks within a mission that will contribute to the user's score. The score is used for high-score boards, and potentially for giving away (real-world) awards. In addition, there will be achievements (one-time, daily or weekly) that give the same resources as above. The score will be "built" from the individual resources collected by the user as rewards for the different challenges (indicatively, green resources are rewarded for completing annotation tasks, blue resources are rewarded for engagement with information content, clear resources are provided for achievements and longitudinal interaction with the game). Which resources are spent to increase score is in the control of the player, allowing for some strategies in terms of which resources to collect.

While the game is online, interactions with other players will only be implicit and never direct. This will limit any potential for harassment and diminish the need for moderation of social spaces. For the purposes of maintaining privacy and GDPR issues, the player will be assigned a player profile complete with hard-coded pseudonym by the system on their first login. The player profile will be tied to the device where the RDG is installed. The pseudonym will be used for public notification about the user's game activities, e.g. displaying the player's score in the high-score board. Other login options will be evaluated in the future, but raise issues of privacy and security without necessarily an equivalent benefit to the player experience.

Since the external gameplay loop will consist of completing challenges in a pre-ordained sequence, extensive design effort has been placed on the types of challenges. We identify four types of challenges:

- **annotation**: where players provide data on material identification, localization and categorization
- **content knowledge**: where players are informed about materials, impact of recycling, and practical aspects of recycling.
- **content testing**: where players test their knowledge about content they have been exposed to in past content knowledge challenges.
- **decompression**: where players engage in short mini-games without any goal except to relax in-between taxing challenges. This would classify as "fun" challenges with no purpose except entertainment.

The current development efforts have focused on annotation challenges, which are described in Section 3, while content for content knowledge and content testing challenges is being drafted as presented in Section 5. The decompression challenges will be refined later as they have few design concerns, borrowing mechanics from more traditional mobile games and simplified for short interactions.

3. RDG Mini-Games

The main development effort in these first months of T6.2 (M3-M9) revolved around the design and development of standalone mini-games that solicit and collect user annotation of waste. Following PIL-1 of RECLAIM (*Advanced AI for material identification, localization and categorization*), the designed mini-games attempt to cover all three annotation tasks: identification, localization, categorization. By developing standalone mini-games, rather than a mission structure as described in Section 2, we can showcase and test individual functionalities quickly. This is important for internal testing, to ensure that the graphic user interface, mobile device interactions (e.g. touch screen controls), and game loop work before they are integrated in a larger (but slower to test) game. Moreover, this modularity allows us to communicate with stakeholders, especially task leaders in T6.1 and WP3, to ensure that the interfaces work with their data and for their purposes.

Based on the above rationale, we have developed and report here seven mini-games with their core gameplay loop and user interface, focusing on user annotation tasks for material identification, localization, categorization. Titles of these mini-games follow a "verb" format based on their game mechanic [Sicart2008], borrowing from the *WarioWare* (Nintendo, 2003) style of mini-games:

1. "Paint" mini-game
2. "Detect" mini-game
3. "Count" mini-game
4. "Outline" mini-game
5. "Locate" mini-game
6. "Choose" mini-game
7. "Categorize" mini-game

The following sections present the gameplay loop and user interface of each mini-game in more detail.

3.1 "Paint" Mini-Game

In the "Paint" mini-game the user is presented with a sequence of images and asked to highlight all items of a specific material, in each one of them, using a virtual paint-brush.

The introduction screen (Fig. 1) explains the task and provides a brief description of the material in question. As soon as the user is ready, they can tap on the Start button to begin the interactive session. A similar introduction screen exists for every mini-game, but is omitted from the other examples for the sake of brevity.

During the interactive session, the user is presented with a sequence of reference images, displayed in the layout seen in Fig. 2. The reference image is shown in the middle, a reminder of the current task is shown at the top, and a tool-box is on the right. Using the paint-brush tool, they can use their finger (or stylus) to freely draw on top of the picture, in a manner that resembles a Photoshop brush. The highlighted regions remain semi-transparent, to allow the user to inspect their input and the underlying material. The user also has access to an eraser tool, which they can use to make corrections. Finally, they can clear their input and start from scratch, by using the Clear button. As soon as they have finished highlighting the items of their choice, they can hit the “accept” button and move on to the next one.

As soon as the session is over (after 5 images), the user lands on the outro screen (seen in Fig. 3), where they get confirmation that their input has been received.

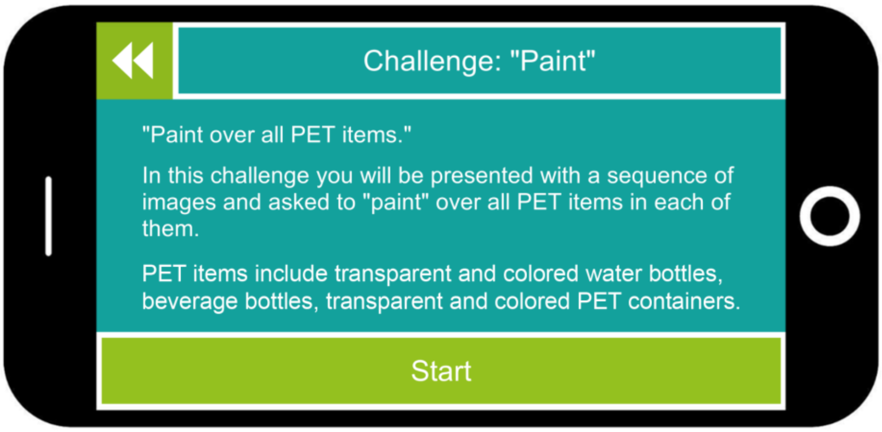


Fig. 1: Paint Mini-Game, Introduction Screen

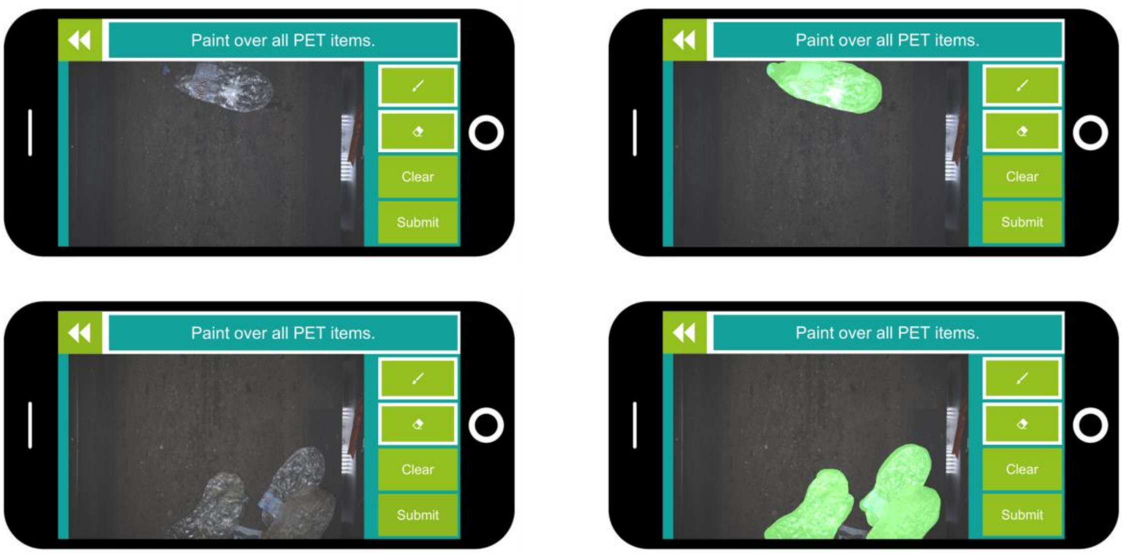


Fig. 2: Paint Mini-Game, Screenshots from an interactive session. The left row shows the reference images, while on the right the user-input is overlaid.

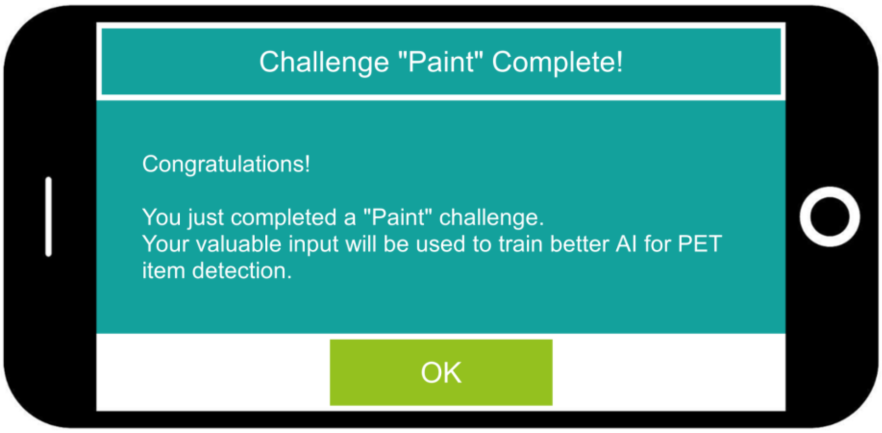


Fig. 3: Paint Mini-Game, "Challenge Complete" Screen.

3.2 "Detect" Mini-Game

In the "Detect" mini-game, the user is presented with a sequence of images and asked whether they can detect any item of a specific material.

During the interactive session, shown in Figure 4, they provide their response by simply tapping the "Yes" or "No" buttons, at the right of the layout. As soon as they provide their input (Yes / No), the next image is displayed.

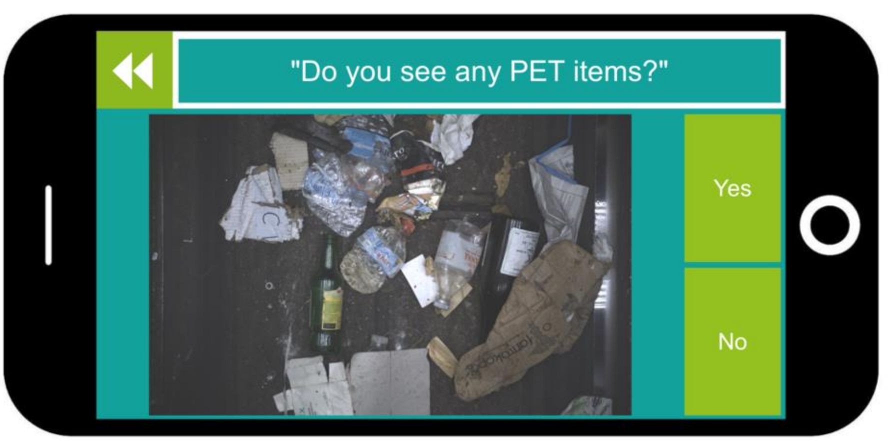


Fig. 4: Screenshot from an interactive session of the "Detect" mini-game.

3.3 “Count” Mini-Game

In the “Count” mini-game, the user is presented with a sequence of images and asked to count how many items of a specific material they can spot in the picture.

During an interactive session, shown in Fig. 5, they can adjust their response using the plus (+) and minus (-) buttons which increase and decrease the counter shown in-between them. The counter’s lower limit is zero (0). As soon as they are happy with their response, they can tap the “OK” button and move on to the next image.

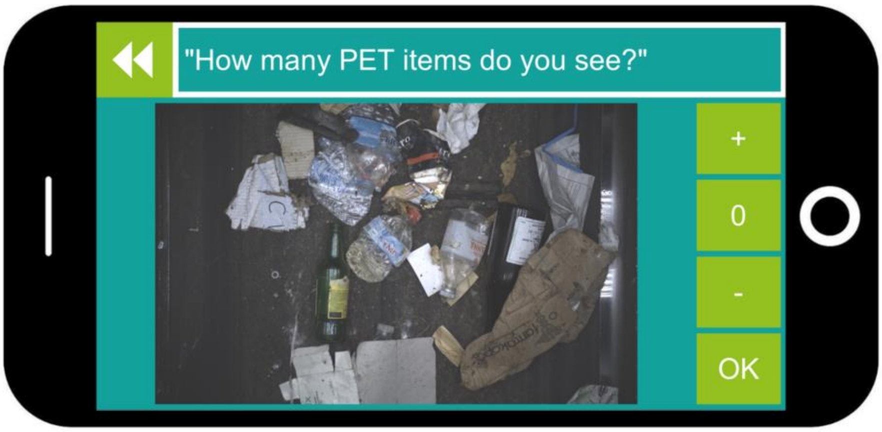


Fig. 5: Screenshot from an interactive session of the “Count” Mini-Game.

3.4 “Outline” Mini-Game

In the “Outline” mini-game, the user is presented with a sequence of images and asked to draw the bounding rectangle of a single item of a specific material.

They can draw a bounding rectangle on the screen (as shown in Fig. 7) by simply dragging their finger on the screen. In doing so, the two opposite corners of the rectangle are defined, and the rectangle is drawn on screen. If they are not happy with their selected region, they can use the “Reset” button to clear it and start over. As soon as they are happy with their response they can hit the “OK” button and submit it. In case they can spot no items of the specified material in the reference image, they can also submit an “empty” response.

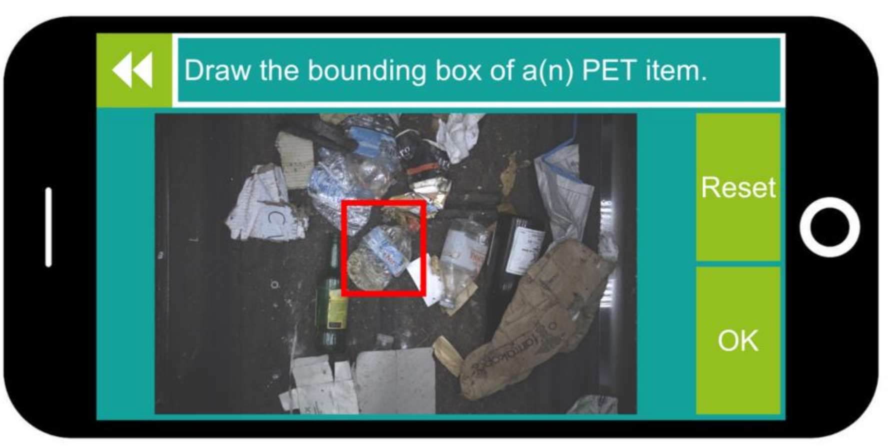


Fig. 7: Screenshot from an interactive session of the "Outline" mini-game.

3.5 "Locate" Mini-Game

In the "Locate" mini-game, the user is presented with a sequence of images and asked to locate the center of a single item of a specific material.

During the interactive session, shown in Fig. 6, they can tap on the reference image to indicate their selection. As soon as they do, a "target" icon appears, indicating their selected location. They can use the "Reset" button to clear their current selection and try again. As soon as they are ready, they may hit the OK button and submit their response. If the user cannot spot any items of the specified type, they can also submit an "empty" response.

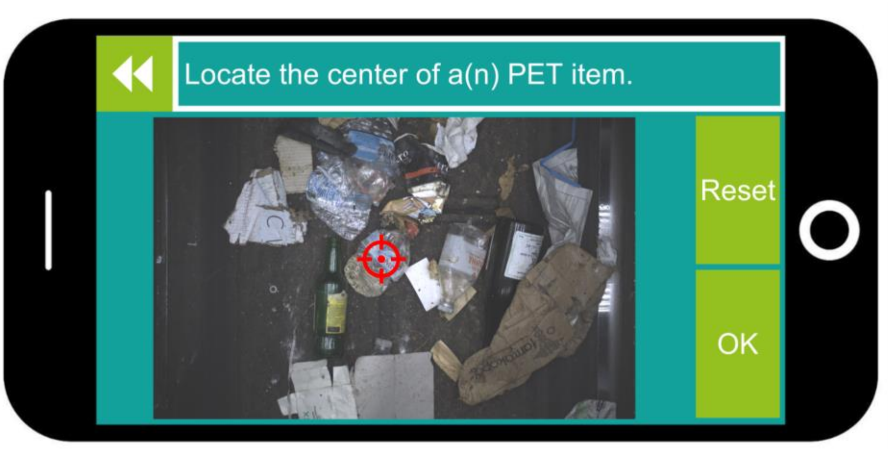


Fig. 6: Screenshot from an interactive session of the "Locate" mini-game.

3.6 "Choose" Mini-Game

In this mini-game, the user is presented with a sequence of images and asked to indicate the best AI-generated mask that captures all items of a specific material. The number of alternative masks can vary, with a maximum of four options to choose from, as shown in Fig. 8. The user can select any of the displayed masks, by tapping on it, and their current selection is indicated by a red bounding rectangle. They can clear their selection, using the reset button and they can also submit an "empty" response, in case they can't decide which mask is best, or if all masks are equally good or bad.

Fig. 8: Screenshot from an interactive session of the "Choose" mini-game.

3.7 "Categorize" Mini-Game

In this mini-game, the user is presented with a sequence of reference images, overlaid with a set of object outlines. Their assigned task, in this case, is to classify each one of the object outlines as belonging to a specific type of material.

As shown in Fig. 9, the reference image and the overlaid item boundaries are shown at the center, while a set of available materials is shown at the right. The user can tap on any of the available materials, thus selecting it, and then tap on any of the item boundaries, assigning this material to this item. At any point, they can use the "Clear" button to clear their selections and start from scratch. If they are unsure about an item, they can leave it blank. As soon as they are ready to submit their categorization, they can hit the OK button and move on to the next picture.

Fig. 9: Screenshot from an interactive session of the "Categorize" mini-game.

4. Methods for User Data Validation

A core challenge with citizen science games, and RDG in particular, is that user data is collected in lieu of expert annotations. Therefore, there is no ground truth on what a correct user response is. Based on current practices in crowdsourced data collection tasks (such as CAPTCHA) and the particularities of the RDG datasets and mini-games developed so far [Arnab2019; Strobl2019; Sumner2020], the following methodology has been designed for validating user data. Note that the current methods have not been integrated yet into the mini-games of Section 3, as they require a database of user data which is planned for development in upcoming months in collaboration with other WP leaders (see Section 6).

4.1 Types of data collected

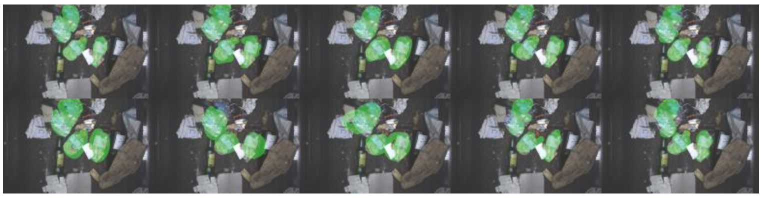
Each mini-game presented in Section 3 has been designed to solicit a specific type of user-generated data that can be later-on used for training (or retraining) specialized AI models for object detection. The following paragraphs describe the type of data collected from each of the mini-games, and the ways in which they can be used in the context of machine-learning.

4.1.1 Data collected via the "Paint" mini-game

As described in Section 3.1, the "Paint" mini-game asks the user to highlight the regions of an image that include items of a specific material. Every time the user annotates a reference image in this manner, they essentially generate a binary mask, in the form of a bitmap (as seen in Fig. 10), which classifies each pixel of the reference image as belonging to the reference material, or not (binary classification per pixel).

Obtaining a large number of such masks, from different users, can be beneficial for generating a dataset for AI-based image classification and segmentation. In contrast to the ideal case, where an expert can annotate each image with high fidelity and precision, gathering data in a crowd-sourced manner is bound to be noisier and less precise. If we were to compare each annotation of Figure 10 or Figure 11 with an expert-generated annotation, we would probably find that none of them is as especially precise or correct. Non-expert users may not be as skilled for (1) correctly identifying the items, (2) properly drawing over the correct regions. However, by obtaining many annotations (from many non-expert users) for a single reference image, we can easily calculate an average annotation (as shown in Figure 11). This average of all user annotations can be used as a probabilistic annotation, where each pixel is now assigned with a probability of belonging to the reference material. The idea here is that the regions (pixels) where more users' annotations coincide are more likely to be correct. Of course this assumption may not hold in all cases. For example, there may exist reference images that are misleading for the larger number of non-expert users. However, we tentatively accept that for most examples of this type of annotation, the previously mentioned assumption is safe enough to make.

(1) Reference image

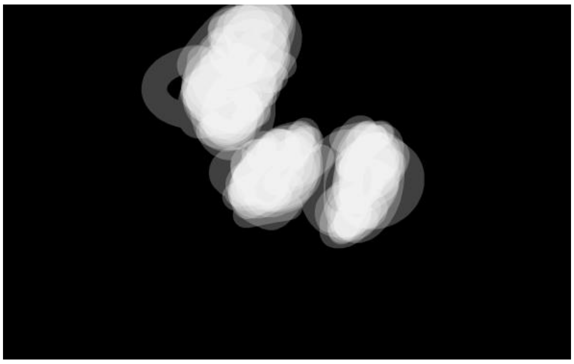


(2) Annotations of PET bottles from 10 different users, for the same reference image.

(3) Binary classification masks (per pixel), generated from the user annotations.

Fig. 10: A reference image (1), where the user is asked to highlight all PET bottles. In (2), a number of annotations from different users and in (3), their generated binary masks.

(1) Reference image, where the users are asked to highlight all PET bottles.



(2) Averaged user annotations, resulting in a probabilistic classification "heatmap".

Fig. 11: Overlaid (averaged) user input masks collected via the "Paint" mini-game. Such overlaid data can be used to estimate the average probability for the existence of items of a specific material, within the image.

4.1.2 Data collected via the "Detect" mini-game

In the "Detect" mini-game the user is asked whether they can detect any items of a specified material in a reference image. In doing so, they provide a binary (True / False) annotation for the reference image.

Since the annotators are not expected to be experts, even such a simple annotation task should be expected to be imprecise at times, either simply due to user mistakes or due to certain images being somehow hard to analyze or misleading. One way to mitigate this is to acquire many annotations (from many different users) for each reference image. A high degree of agreement between annotators (for example: 95% True vs 5% False) could be treated as an indication for a correct classification. Images with a low degree of agreement between annotators, on the other hand (for example: 55% True vs 45% False), could be flagged as "problematic" and be assigned to experts (among RECLAIM participating institutions such as FORTH) for further analysis, or simply omitted from the dataset.

In any case, this process can help create a dataset of annotated images that can then be used to train AI models for binary image classification.

4.1.3 Data collected via the "Count" mini-game

In the "Count" mini-game, the user is asked how many items of a specified material they can spot in a reference image. Their numeric answer is a form of multi-class annotation for the reference image, where the available classes are the integers between 0 and a maximum value.

Similar to the previous case, the users' responses may be in high agreement (for example: 95% of users responded with "2 items"), or not (for example: 50% of users responded with "1 item" and 50% responded with "3 items"). Images with a high degree of agreement could be directly accepted as part of the dataset, while images with a low degree of agreement could be assigned to experts to resolve, or simply omitted from the dataset.

In any case, this process can be used to create a dataset of annotated images for training AI models that can perform multi-class image classification, being able to detect the number of items per requested material in a reference image.

4.1.4 Data collected via the "Outline" mini-game

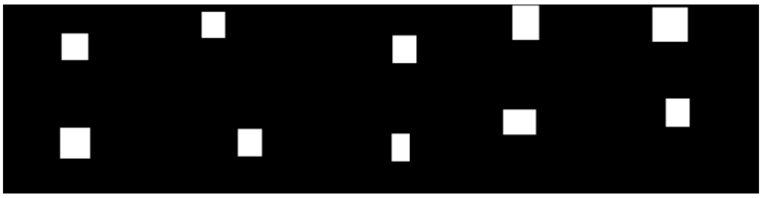
In the "Outline" mini-game, the user is asked to draw a bounding rectangle, surrounding an item of a specified material on a reference image.

In doing so, they define a region in the image, within which an item of the specified material is supposed to exist. Their input can easily be converted into a binary classification (per pixel) mask, similar to the ones shown in Figure 12. Finally, many inputs (from many different users) for a single image can be overlaid, creating a probabilistic binary classification map, like the one shown in Fig. 13. Alternatively, the center of the bounding box can be used and treated instead as a point, processed in the same way as the "Locate" minigame in Section 4.1.5.

Each one of these probabilistic maps can be used as a single annotation in the final dataset, aiming to train segmentation models, similar to how data points are treated in the case of the "Paint" mini-game.

(1) Reference Image

(2) 10 different user inputs, detecting PET items in the reference image, acquired via the RDG



(3) Binary classification masks generated from the user annotations

Fig. 12: (1) A reference image where the user is asked to spot PET items. (2) Different user annotations, as they appear within the RDG. (3) Extracted binary classification maps.

Fig. 13: Overlaid binary classification maps, resulting in an average, probabilistic map of PET items in the reference image.

4.1.5 Data collected via the "Locate" mini-game.

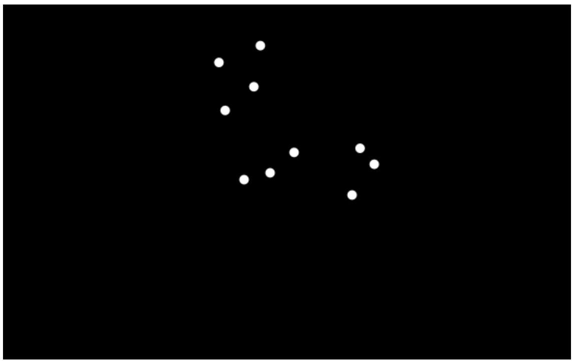
In the "Locate" mini-game, the user is asked to indicate the center of an item of a specified item, on a reference image.

Their input is stored as a single 2D point, representing the image-coordinates where the user clicked. In Fig. 14-2, we can see a cloud of points, representing the input of multiple users, detecting the centre of PET items in a reference image. Those points can be isolated, as shown in Fig. 14-3, and then a probability estimation method like Kernel Density Estimation (KDE) [Botev2007] can be used to estimate the regions of the image which are more densely populated (as shown in Fig. 14-4). These regions can be treated as more probable to include pixels belonging to a PET item.

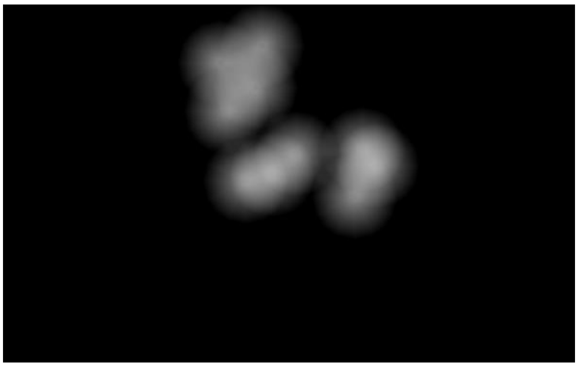
Every one of these probability maps can be used as a single data-point in a dataset, with the purpose of training AI models.

(1) Reference Image, where the users are asked to indicate the centre of PET items.

(2) Indicative input of multiple users, as shown in the RDG.



(3) Resulting point-cloud, based on the input of multiple users.



(4) KDE of the point-cloud, representing the probability distribution of PET items in the image.

Fig. 14: Using point-clouds to estimate the location of PET items.

4.1.6 Data collected via the “Choose” mini-game

In the “Choose” mini-game, the user is asked to compare a set of pre-computed masks, and select the one that better represents the regions of a reference image that include a specific material. The presented masks represent the boundaries of all items of a specific material, within the reference image, and the user’s comparison criterion has to do with the correctness and precision of presented boundaries.

All masks presented to the user have been generated by pre-trained AI models that perform image segmentation. Using the same process as the "Count" mini-game (Section 4.1.3) we can collect the frequency that each of the four options was picked. The pre-trained AI model with the most popular mask(s) has better hyperparameters which can then inform the AI developers to improve upon this model in future iterations.

4.1.7 Data collected via the “Categorize” mini-game

In the “Categorize” mini-game, the user is presented with a single image, and within it, the boundaries of a number of objects. Then, they are asked to classify each one of these presented regions as belonging to one of a set of possible materials. In contrast to the previous mini-game, here the emphasis is not so much on the precision of each boundary, but rather on its correct classification into the proper material.

Similar to the previous case, user-generated data here can be used to improve the existing AI-models that perform image segmentation and classification. The data collected is processed similarly to Section 4.1.6 and Section 4.1.3 as the frequency that each boundary was assigned a specific material. Therefore multiple probabilities (one per mask per material) will be produced from processing the user dataset.

4.2 Detecting Agreement

An important aspect for the gamification of the RDG (see Section 2) is rewarding the player for their annotation tasks. Since there is no ground truth of what a correct annotation is, a corpus of "definitive" user annotation needs to be provided before one can assess whether their annotations are good or not. Therefore, for the first 10 user annotations on one image in one mini[game, the reward will be uniform to all users as a participation "early bird" award. If the first 10 annotations have strong disagreements (e.g. in the "Detect" mini-game 55% of users claim it is True vs 45% claim it is False), the early bird award system will continue to be in effect until enough annotators agree above some threshold (e.g. 75%). This threshold will be ad-hoc designed per mini-game based on preliminary tests. Moreover, what constitutes agreement will be discussed below.

After a corpus of user annotations per image per mini-game is collected, then new user annotations can be awarded based on how close they are to the consensus. For simple classification tasks (i.e. "Detect", "Count", "Choose" and "Categorize") detecting agreement is easy: if the user selects the most popular class (e.g. True in "Detect") then they should be rewarded according to how definitive this preference is among other users. Each annotation will be assigned a maximum reward as a number of resources (see Section 2) in the edge-case scenario where a user selects the class that has received 100% of other players’ votes; then this number will be adjusted proportionally to how far from 100% the other player’s votes are for this class. For example, in the "Choose" mini-game with four AI-generated masks, if the users so far have chosen Mask A at 0%, Mask B at 20%, Mask C at 75% and Mask D at 5%, the user could be awarded 20% of the maximum number of resources if they choose Mask B, but

75% of the maximum number of resources if they choose Mask C. Additional testing and adjustments on how resources will be split based on ratios will be performed in future iterations to ensure that this acts both as a motivator for further engagement (e.g. via participation rewards) but also is a significant drive for players to make an effort in annotating to the best of their ability.

For mini-games where the users provide freeform data (“Paint”, “Locate” and “Detect”), the agreement will be calculated in a similar way as the error rate in the AI algorithms are calculated during training in WP3. Specifically, a new annotation (that needs to receive a reward) will be compared with each previous annotation individually via e.g. the ratio between the area occupied by the intersection of both masks over the area occupied by the union of the two masks for “Paint” or “Locate”. For “Detect”, a similar algorithm is possible based on the KDE plot or based on the distance between the points. Averaging the overlap values of comparisons between the new annotation and all past annotations gives an estimate of how close this new annotation is to the consensus of past players. Resources will be rewarded based on the ratio from a maximum number of resources per annotation similarly to how it was presented in the above paragraph.

It is worth noting that this will not be the only reward provided to players, as establishing the ground truth (while critical to the AI algorithms) is of lesser interest to the players (see CSG players’ motivations in D2.1). Therefore, other resources will reward continued engagement and participation in order to motivate the player to keep engaging with the RDG. Therefore, incentives to improve the users’ performance in annotation tasks will be orthogonal to other incentives to continue playing the RDG and participating in the research effort longitudinally. Details of the reward structure based on ratio of agreement (and how lenient or harsh this may be) will be refined with preliminary user tests and take into account recommendations from the literature. Indicatively, [Foody2014] warns that "introducing overly burdensome structures to ensure quality could damage the potential contributions from related socially-conscious and citizen-focused data collection and mapping efforts."

5. Content needs and format for the RDG

Based on the user requirements as part of WP2 and reported in D2.1, we identified a number of critical information that the RDGs should convey. This information should address misconceptions about recycling, convey the impact of recycling on the environment (both financial and societal), links to other societal and environmental issues, allow for engagement with the RECLAIM activities and its scientists, and inform about technical aspects of the collected data (in terms of AI algorithms and robotics components).

Beyond data collected from surveys and focus groups, what content is currently available, its format, IP and licenses must be considered. Producing new content as part of the RDG design is a challenging task that goes beyond what is possible with the time and resources of this task, and therefore the RDG efforts will revolve around **synthesizing** existing content and formulating it as testing challenges (see Section 2). Additional discussions throughout the first reporting period of T6.2 (M3-M9) with RECLAIM partners and other stakeholders in the Greek context have informed this decision and paved the way for collecting appropriate and correct content (see Section 5.2).

5.1 Planned targeted information gathering

The user requirements of the RDG (in D2.1) were collected through a) review of existing literature on citizen science games (CSG), b) focus group discussions with a sample of potential target players and experts in the field of environmental sustainability, waste management, and data annotation, and c) an online survey open to a wider audience (see more details in D2.1). One of the fundamental, underlying elements of the CSG design is the support of intrinsic and extrinsic motivations of the players. To this end, the focus groups and survey provided more grounded and in-depth insights on the needs, requirements, preferences and practices of the target audience.

Regarding the environmental content of the game and based on the user requirements as well as the goals of the RDG (see Section 1) the following main topics were identified. These topics and relevant resources will provide the foundation for the design of the RDG content challenges, quizzes, and puzzles:

- Information about target materials and recyclable objects, e.g., the impact of specific materials on the environment. For instance: design of a challenge where all types of possible packaging waste are listed and explained to the citizens. Examples of such materials are PET containers (e.g., transparent and colored water bottles, beverage bottles, transparent and colored PET jars), HDPE packages (e.g., household detergent bottles, cereal liners, margarine tubs), LDPE Film (e.g., carrier bags, large pieces of LDPE, sacks), PP/PS (e.g., yogurt cases, margarine bowls, crates, big bags including secondary components such as lids, labels, etc./packaging of disposable cups and

plates), ferrous (e.g., beverage and food containers), aluminum cans (e.g. soft drink cans), drinking Cartons (packaging of milk, juice etc.)
- The financial impact of recycling/not recycling. Financial implications of waste landfills.
- Solutions to problems, success stories, real world cases of good recycling practices and/or failures and implications.
- What happens after the recycling bin? The journey of a recyclable item.
- What happens after the material is sorted in the recycling/sorting facilities? The second life of the recycled materials. What happens after the sorting/recycling facility? What happens after the material is recycled in the sorting and recycling units. What products are made from recycled materials?
- Mistakes people make when they sort recyclable objects.
- Showcasing the impact of recycling in real-life (e.g., “if we recycle a plastic bottle, we save energy that can power a 60W light bulb for 3 hours.”)
- Less waste in landfills. The implications of landfills on the environment and on local communities. The financial cost. Recycling means less landfills
- Reduce. As a consumer, be more aware of and conscious about the products you buy. Ways to reduce waste production. Reuse. Ways to reduce waste as a consumer. Barcodes - explaining what they mean, and how to decode them as a consumer.
- Different recycling practices in different countries
- What happens in a sorting facility?
- What are the consequences of not recycling?

5.2 Discussions with stakeholders

To address the educational and information content of the game regarding waste management, recycling, and relevant environmental issues, a team of experts from the partner institutions collaborated closely for the design of the game. Specifically, ISWA (International Solid Waste Association, https://www.iswa.org/) provides information, material and resources relevant to waste management, recycling and implications. HERRCO (Hellenic Recovery Recycling Corporation, https://www.herrco.gr/en/about-us/) provides general information about recycling and waste management, information of local interest (Greece), and insights on the target group needs and requirements. Involvement of experts focused on waste management and recycling ensures validity and authenticity of the game content.

In addition, other valid resources are used for the environmental content of the game such as publicly available information at the Hellenic Recycling Agency (https://grecycle.gr/, https://www.eoan.gr/), and the Greek Ministry of Environment and Energy (https://ypen.gov.gr).

6. Future Work

For the upcoming months, a number of steps are planned for the development of a more robust version of RDG, for the purposes of initial tests in M13 and the broader evaluation of the impact of RDG during T6.3 (Deployment and evaluation of environmental games). Below we discuss the main directions for the coming developments of the RDG.

The developed mini-games presented in section 3 have mainly focused on the technical aspects of collecting annotations from users. We have treated each type of challenge as an isolated mini-game, where a single task is assigned to the user and repeated for a sequence of images. This approach has allowed us to make some initial steps in UI design and solve many technical problems. However, the goal is to combine these modules with other types of content and synthesize more interesting and rich experiences for the player, as per the roadmap described in Section 2. An indicative sequence of such a roadmap would be: (1) informative content, (2) annotation challenges of various types and (3) knowledge challenges of various types. Therefore, these standalone mini-games will need to be integrated into a roadmap structure that also includes other challenge types (see Section 2 for a description of the other challenge types) and their sequence designed and tested in order to keep the interest of the user.

As already stated in previous sections, a core goal of the RDG is to collect user-generated data for re-use in the AI algorithms of the project. Therefore, the RDG is heavily dependent on an underlying, central database which supports various aspects of its operation. A rough outline of this database is that it stores a collection of data-items that can be used to synthesize content for the user, including the annotation challenges, informative content and knowledge challenges. Second, a collection of data that describe a specific player's activity history, including: (1) their current Resources, Achievements, and other player profile information, (2) their previously submitted annotations, (3) their responses to specific questions posed during knowledge challenges, (4) the informative content that they have been presented with and, perhaps, more aspects of their activity. This database will be developed in coming months.

As discussed in Section 5, there is currently a need for educational and informational content that satisfies other needs of the RDG (see Section 1) and, additionally, provides different types of challenges to lower the fatigue of the user when annotating difficult waste images. As discussed in Section 5, there is currently a methodology and a plan for collecting and synthesizing such data from existing sources, and the coming months will also focus on collecting such data and preparing them in a way that is playful and challenging for players of the RDG (e.g. as quizzes and informational tidbits in-between annotation challenges. There is already a pipeline for collecting this processed information, and the coming months will also see the development of mini-games and interfaces for this type of challenge.

An additional direction for the upcoming developments of the RDG is the visualization of a player profile which allows users to view their achievements so far, their resources, their performance in different annotation tasks or quizzes, and potentially their position in a global leaderboard etc. This aspect of the RDG, while documented in the Game Design Document (see Section 2) will need to be carefully designed in order to ensure that it is appealing to a user and promotes their reflection regarding the goals of the RDG (see Section 1). Therefore, multiple iterations of this player profile will be refined and tested during T6.3 (M13-M36).

Finally, an important area for improvement is the user experience. The initial mini-games described in Section 3 have been tested on mobile devices and are both responsive and intuitive. However, further refinements in tandem with other developments described above (mission structure, educational content, database integration) will require additional iteration on the user interface and visualizations. This refinement to the user experience is expected to continue during the testing phase undertaken by T6.3 (M13-M36).

7. Conclusion

This report presented the current state of the game design and development for the Recycling Data Game. The goals of the RDG are many and often orthogonal, as is the case in many citizen science applications and games, while user requirements highlighted the need for casual and low-effort interactions of a few minutes as the ideal user experience. In light of these specifications, we designed a game consisting of sequences of small bite-sized challenges that we have themed after the goals of the RDG as specified in the DOA. Initial prototypes of one type of challenge, that of annotation, which is the hardest and least straightforward, have been implemented and presented in this report. Other types of challenges, related to improving social awareness as per the *PIL-4* of RECLAIM, are currently in preparation with content collection processes as discussed in Section 5. Future iterations of this report will implement the online functionalities of the RDG, including validation of user data and reward systems as discussed in Section 4, and build up additional types of challenges into a consistent roadmap.

8. References

[Arnab2019] Arnab, S., Lewis, M., Bogliolo, A., Klopfenstein, L. C., Delpriori, S., & Clarke, S. (2019). Player Interaction with Procedurally Generated Game Play from Crowd-Sourced data. *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 333–339. https://doi.org/10.1145/3341215.3356257

[Botev2007] Botev, Z. (2007). Nonparametric Density Estimation via Diffusion Mixing (Technical report). University of Queensland.

[Foody2014] Foody, G. M. (2014). Rating crowdsourced annotations: Evaluating contributions of variable quality and completeness. International Journal of Digital Earth, 7(8), 650–670. https://doi.org/10.1080/17538947.2013.839008

[Savvani2018] Savvani, S. (2018, September). State-of-the-art duolingo features and applications. In *Proceedings of the International Conference on Interactive Collaborative Learning* (pp. 139-148). Springer, Cham.

[Sicart2008] Sicart, M. (2008). Defining game mechanics. *Journal of Game Studies*, volume 8, issue 2.

[Sumner2020] Sumner, J. L., Farris, E. M., & Holman, M. R. (2020). Crowdsourcing Reliable Local Data. *Political Analysis*, 28(2), 244–262. https://doi.org/10.1017/pan.2019.32

[Strobl2019] Strobl, B., Etter, S., Meerveld, I. van, & Seibert, J. (2019). The CrowdWater game: A playful way to improve the accuracy of crowdsourced water level class data. *PLOS ONE*, 14(9), e0222579. https://doi.org/10.1371/journal.pone.0222579