

A Gaming Ecosystem Crowdsourcing Deep Semantic Annotations

François Bry, Corina Schemainda, and Clemens Schefels

Institute for Informatics, University of Munich, Germany

Abstract. This article reports on an analysis of the semantic quality of artwork annotations that have been collected by the web platform ARTigo from 2008 to 2014. ARTigo is both, a gaming platform for collecting semantic annotations and a semantic search engine: While playing games featuring artworks, one leaves data describing these artworks that are automatically used for building up, and continuously improving, an artwork search engine. ARTigo is an ecosystem consisting of different types of games aiming at collecting annotations of different semantic depth: Description games aim at collecting mere descriptions, diversification games aim at completing artwork descriptions with more specific characterizations, integration games cluster annotations so as to generate more precise descriptions. The qualitative analysis this article reports about points to both, the effectiveness of ARTigo as a gaming ecosystem and the effectiveness of each of game at collecting annotations with the desired semantic depth.

1 Introduction

ARTigo¹ is an outcome of the five years research project play4science.² Since the project's beginning in 2007, ARTigo collects semantic data on artworks from the footprints left by players of games featuring artwork images. From these footprints and relying on an original higher-order latent semantic analysis conceived for the purpose [12,11], ARTigo automatically builds an artwork search engine.³

In Art History, there is a considerable interest for search engines that would make possible to retrieve artworks from their iconographic content, that is, by referring to what they explicitly depict (like an object, a scene, a person, or a group of people) as well as from their implicit content, that is, the emotions (like love or sorrow) or abstractions (like gallantry, freedom, or power). Some art historians expect such search engines to open up new perspectives in art history by unveiling similarities, or discrepancies, or any other so far unnoticed features of artworks, art schools, or epochs. Even though this expectation has been criticized as too optimistic, there is no doubt that systematic and algorithmic artwork analyzes would, if not revolutionize art history, at least provide

¹ <http://artigo.org/>

² <http://play4science.org/>

³ Accessible at the top right of <http://artigo.org/>

the field with a useful and so far largely missing “cartography” of the artifacts that are, or have been, considered as artworks. Key to such a cartography is the annotation of large number of artworks.

Since art historians rarely have much funds at their disposal, a cheap approach to annotating artworks was necessary. Indeed, annotating a large artwork database can become costly if done for money: low paid workers like students working a few hours a week would for example in Germany cost about €1,500 for a full month work load and would produce in average about three annotations per minute, that is 28,800 annotations per month, yielding an average price for annotation of about €0.05. Annotating 10,000 artworks with an average of 20 annotations per artwork each would therefore cost about €10,000. Requiring further that annotations being proposed by two independent workers and considering, as we have experienced, that only 20% are proposed by two independent workers, rises the cost of annotating 10,000 artworks to €50,000. We were considering an artwork image database of 30,000 to 100,000 artworks from the Prometheus Image Archive⁴ and from museums we were in contact, or collaborating, with, raising the cost for paid annotators to 1.5 to 5 Millions Euro. This was by several orders of magnitude out of our reach. We therefore decided to rely to unpaid lay persons, human computation, and gaming as incentivization.

Whether artworks’ explicit and implicit contents can be conveniently described by annotations so as to make a semantic artwork search possible is the core issue of this article and therefore addressed in the following sections. Whether unpaid lay persons may generate artworks’ descriptions of a quality comparable to that of selected paid workers is worth considering. According to Trant [9], most artwork databases “present the characteristics of objects (such as creator, size, materials, use, provenance) without context, and in isolation from related works”. The jocunde database⁵ of artworks from French museums is such a database with artworks’ annotations that have been generated by paid workers. It contains almost no annotations lay persons would not have produced. For example, Gustave Courbet’s painting “Enterrement à Ornans” (“Burial in Ornans”, a village in Britany) at Musée d’Orsay in Paris is annotated in the jocunde database with (translated from French) “burial”, “priest”, “acolyte”, “peasant”, “cross”, “sorrow”, “blessing”, and “Ornans”. Except the last annotation derived from the painting’s title which, obviously, was known to the annotators, all annotations could have been proposed by lay persons. Further lookups at artwork databases have confirmed that lay persons are capable of annotating artworks as well as selected paid workers.

In a first phase, from the end of 2007 to the beginning of 2008, a proof-of-concept prototype was realized (by our colleague Dr. G. Schön in collaboration with our art history partner Prof. Dr. H. Kohle) consisting of a slightly modified version of Luis von Ahn’s ESP Game [10]. From February 2008, when this prototype was put on-line, to April 2009, 1,500 distinct players had registered, played, and while playing left 1.5 Million annotations 190,000 of which had been

⁴ <http://www.prometheus-bildarchiv.de/>

⁵ <http://www.culture.gouv.fr/documentation/joconde/fr/pres/htm>

validated, that is, proposed by at least two players. Like other researchers [14,13], we immediately noticed a limitation of the ESP game: its tendency to collect very general annotations. In order to overcome this limitation, we designed, realized, tested, deployed, and publicized a gaming ecosystem extending the (slightly modified) ESP game of ARTigo’s prototype with additional original games [8,1]. From 2009 to 2014, the ARTigo gaming ecosystem has collected about 6,800,000 annotations referring to 220,000 distinct words or phases. 78,000 of these annotations have already been validated (by players). In average, about 150 different players per day have played on ARTigo from 2008 to 2014.

This articles briefly describes ARTigo’s gaming ecosystem and reports on a qualitative analysis pointing to its effectiveness at collecting annotations well describing the explicit as well as implicit contents of artworks. This article is structured as follows. Section 1 is this introduction. Section 2 describes the ARTigo gaming ecosystem consisting of description, diversification, and integration games. Section 3 is devoted to a qualitative analysis of the artworks annotations collected by ARTigo. Section 4 is devoted to possible further work, especially additional games, that have been implied by the qualitative analysis. Section 5 is a conclusion.

The original and so far unpublished contributions of this article are as follows:

- A qualitative analysis of the annotations collected by the ARTigo ecosystem.
- Sketches of new data-collecting games implied by the qualitative analysis.

2 The Gaming Ecosystem

Figure 1 shows how the annotations collected on the ARTigo platform flow from game to game providing the “seed data” necessary for the games to be playable and for collecting “deep semantic” data.

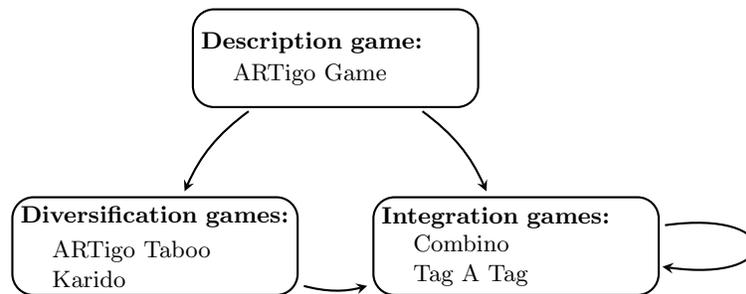


Fig. 1. Seed data flow in the ARTigo ecosystem.

Description Games are games the players of which describe an artwork by proposing annotations related to anything referring to the artwork: objects or characters it depicts, its colors, the materials it is made of, etc. ARTigo’s has a

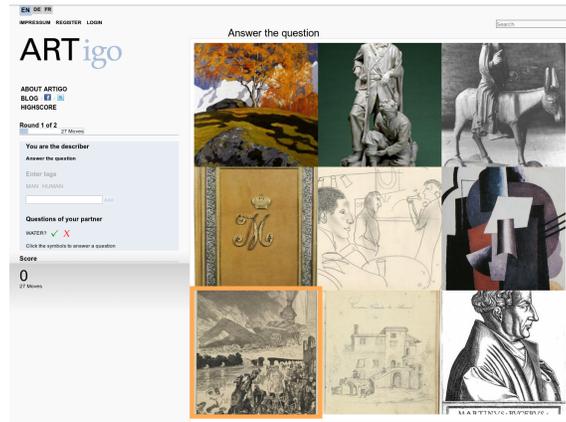


Fig. 2. Describer’s screen during a Karido game (hovering over a picture enlarges it).

single description game,⁶ a variants of the ESP Game [10] called ARTigo game, with following gameplay: When two randomly selected players enter a same annotation, they both score and the annotation is validated. Validation is necessary to ensure annotation correctness (among other exclude malicious annotations). This form of validation has been found sufficient [10]. The annotations collected by description games provide “surface semantics”. They are needed both as basic descriptions and as seed data for games collecting “deep semantics” annotations.

Diversification Games collect annotations that are in general more precise or of a deeper semantics than most annotations collected by description games. Diversification games use annotations collected by description games as seed data. The ARTigo ecosystem has two diversification games: ARTigo Taboo, an ESP Game [10], and Karido [8,6].

ARTigo Taboo is like the ARTigo Game except that annotations already entered by some players can no longer be entered by any player: they are taboo. ARTigo Taboo forces players to enter novel annotations. Therefore, ARTigo Taboo yields more sophisticated descriptions than the ARTigo Game. Tabooing annotations can be seen as a form of “scripting” [1] discussed below since taboos are instructions for players.

The annotations collected by the ARTigo Game on the one hand and by ARTigo Taboo on the other hand are complementary. In contrast to ARTigo Taboo, the ARTigo Game generates not only annotations but also annotation frequencies. These frequencies are useful for “cartographying” the immediateness of what artworks evoke: The higher is an annotation frequency, the more immediate, or important, is the annotation.

⁶ [16] classifies ARTigo Taboo as description game. The analysis below shows that it primarily is a diversification game, hence the change.

Karido [8,6], the second diversification game of the ARTigo ecosystem, has a completely different gameplay than ARTigo Game and ARTigo Taboo. Karido’s gameplay incites players to enter more specific, or “deep semantics”, annotations. It is basically as follows. Nine similar artworks are randomly selected and displayed to randomly paired players in two 3×3 grids such that the artworks are differently ordered in each grid. Image similarity is determined from the annotations, especially from “surface semantics” annotations, so far collected for these artworks [6]. One player is the “describer”, the other the “guesser”. In the next round, they exchange roles. The describer selects one of the nine artworks on her grid and starts annotating it in such a way that the guesser can recognize and select it. The guesser can ask yes/no questions. Since the guesser’s grid and the player’s grid are differently ordered, locational annotations like “South-East” do not help recognizing the selected artwork. The sooner the guesser selects the right artwork, the higher the scores for both players. Figure 2 on the previous page shows a describer’s screen of the Karido game. A guesser’s screen is similar.

Played with artworks that have only few annotations, Karido produces “surface semantics” annotations like ARTigo Game but at a lower speed than that game. In order to collect “deep semantics” annotations, Karido needs sufficiently similar artworks, what in turns requires a sufficient number of “surface semantics” annotations for each artwork. Karido has been designed to collect only valid annotations.

Integration Games cluster annotations yielding more precise descriptions than the unstructured sets of annotations collected by other games. Annotation clusters are often more difficult for players to suggest since they require a deeper understanding of an artwork, in some cases even some specific knowledge. Integration games are therefore sometimes more challenging than other games therefore contributing to a gaming platform’s attractiveness. Two integration games have been specifically designed for the ARTigo ecosystem: Combino and Tag A Tag [1].

While playing games like ARTigo Game, ARTigo Taboo, and Karido players tend to enter single words from descriptive phrases instead of the phrase itself. Indeed, this strategy ensures a faster and better scoring. Combino makes its players bring formerly collected annotations into relation, that is, reconstruct descriptive phrases. Figure 3 shows a Combino session. A same artwork together with a same set annotations formerly collected for this artwork are displayed to randomly paired players. Both players score when they select the same pairs or sets of annotations from the displayed annotations. Combino is, like ARTigo Game and ARTigo Taboo, a variation of the ESP Game [10].

By “squaring” Combino, that is running it with phrases it formerly collected, long phrases like “old + man + sitting” can be stepwise collected from “old”, “man”, and “sitting”. “Squaring” Combino makes sense because as entering single words from descriptive phrases instead of the phrase itself is a good strategy for ARTigo Game and ARTigo Taboo, constructing the shortest phrases is a good strategy for Combino.

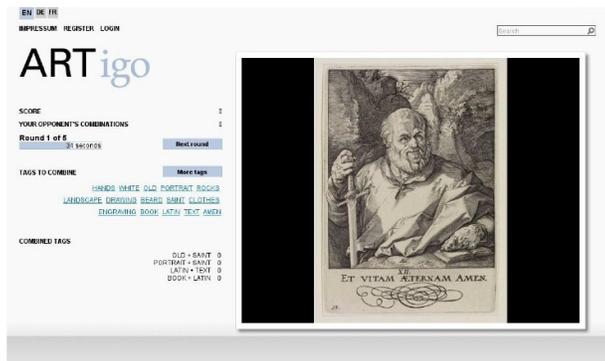


Fig. 3. A Combino screen featuring H. Goltzius: Christus, the twelve apostles with Paulus, 1589, Staatliche Kunsthalle Karlsruhe.

The sets of annotations Combino collects are not really phrases but instead unordered sets. This does not impair the quality of semantic artwork search engine, though. For this reason, the ARTigo ecosystem includes neither ordering games nor a refined version of Combino collecting (ordered) phrases.⁷

Tag A Tag is a “squared” [1] and “scripted” [1] ESP Game [10]. “Squaring” has been already introduced in the former paragraph: “Squaring” a game means running it with annotations it had formerly collected. “Scripting” means giving the players a short instruction on the annotations, or pairing, or whatever other input they are expected to enter.⁸ *Tag A Tag* is an ESP Game [10] displaying to its player an artwork AW, an annotations A formerly collected for this artwork AW and asking “Describe A in artwork AW”. Thus, *Tag A Tag* collects annotations on an annotation A in the context of an artwork AW.

Figure 4 on the next page shows a *Tag A Tag* screen asking a player to describe “ring” in the displayed photograph of the French writer Émile Zola. Possible answers are: “finger”, “right hand” or “right” and “hand”, “wedding” (for “wedding ring”), etc.

Scoring at *Tag A Tag* can be difficult. As already observed, this contributes to the attractiveness of the ARTigo gaming platform.

Most games of the ARTigo ecosystem are variations of the ESP game. Is this a good feature? The authors believe it is for the following reasons. First, it makes all games “casual games”, that is, games that can be played in sessions of any duration, especially in short sessions, and requires neither a learning of rules nor much training. Second, keeping the gameplays of all games similar contributes to the “casual game” nature of the gaming platform. Finally, it reduces the

⁷ A Combino-like game collecting (ordered) phrases is easy to specify.

⁸ The denomination is inspired from pedagogical psychology: Scripting is the name of a teaching method.



Fig. 4. A Tag A Tag screen featuring a photograph of the French writer Émile Zola.

effort needed for implementing and maintaining the games thus contributing the cheapness of the platform. Recall that cheapness is a concern in building a semantic artwork search engine.

3 Semantic Quality Analysis of the Collected Annotations

This section reports on an analysis of the annotations collected by the ARTigo gaming platform aiming at determining whether the reflective approach to human computation of the ARTigo gaming ecosystem, which confronts humans to annotations formerly generated by humans, is effective at collecting “deep semantics” data. Since estimating the “semantic depth” of artwork descriptions is, and most likely will remain for a while, out of the reach of algorithmic approaches and since there are no benchmarks on the “semantic depth” of artwork descriptions, we ran a survey. For not having to define “semantic depth”, what might have been suggestive, the survey did not explicitly refer to this notion. The survey was focused at, and structured after, the following questions:

1. Does the time pressure induced by games influences which annotations are collected?
2. Does the diversification game Karido collect annotations of a greater “semantic depth” than the games ARTigo Game and ARTigo Taboo?
3. Does the integration game Tag A Tag collect annotations of a greater “semantic depth” than the game ARTigo Game?
4. Do small sets of annotations collected for an artwork by different games better describe it than small sets of annotations collected by a single game?
5. Do the diversification games ARTigo Taboo and Karido together collect annotations of a greater “semantic depth” than the description game ARTigo Game?

The Survey Questions. The survey had an introduction, five sections numbered 1 to 5, and a final section. The introduction was about one’s experience with the ARTigo gaming platform and one’s interest in art. The final section collected demographical information (age and gender).

Section 1 was devoted to Question 1: Does the time pressure induced by ARTigo’s games influence which annotations are collected? This questions has been motivated by the study [4]. Four artworks⁹ were presented to the participants who were asked to describe them as well as possible, with up to three annotations and without any time limit. The annotations proposed by the survey participants were then compared with those collected for the same artworks by the ARTigo gaming platform.

Section 2 was devoted to Question 2: Does the diversification game Karido collect annotations of a greater “semantic depth” than the games ARTigo Game and ARTigo Taboo? Four groups of four similar artworks were used in this section. Each artwork group was build as follows from a so-called “initial artwork” randomly selected from the ARTigo database. The five most frequent annotations of an initial artwork were used as search query for selecting three further artworks, so-called “test artworks”, from the ARTigo database. Because of their sharing of annotations, the three test artworks can be considered as similar to the initial artwork they have been determined from. Each group of four artworks, an initial artwork with its three associated test artworks, were presented to the one participant group together with a set of five randomly selected annotations collected for the initial artwork by the description game ARTigo Game, to the other participant group together with the set of the five most frequent annotations collected for the initial artwork by the diversification game Karido.¹⁰ The participants of both groups were asked to select the artwork best corresponding to the five annotations.

Section 3 was devoted to Question 3: Does the integration game Tag A Tag collect annotations of a greater “semantic depth” than the game ARTigo Game? Three artworks, each artwork together with six of the annotations collected for it on the ARTigo platform, were presented to the participants who were asked to rank these annotations after their relevance for the artworks. The six annotations associated to each artwork included both, single word annotations collected by ARTigo Game and combined annotations collected by Tag A Tag.

Section 4 was devoted to Question 4: Do small sets of annotations collected for an artwork by different games better describe it than small sets of annotations collected by a single game? Three groups of one artwork together with four sets of six annotations collected for each artworks were presented to the participants. For each artwork, the six sets of annotations were as follows: The first set consisted of six of the ten annotations for this artwork most frequently collected by ARTigo Game, the second set, of randomly selected annotations collected by

⁹ Here and in the following, artworks of different styles and epochs were selected. However, not all styles and epochs could be represented in the survey.

¹⁰ Here and in the following, annotations collected by diversification games were only selected if they had not been also collected by the game ARTigo Game.

ARTigo Game, the third set, of three annotations collected by ARTigo Taboo and of three annotations collected by Karido, the fourth set, of annotations most frequently collected by ARTigo Game, ARTigo Taboo, and Karido.

Section 5 was devoted to Question 5: Do the diversification games ARTigo Taboo and Karido together collect annotations of a greater “semantic depth” than the description game ARTigo Game? The participants were asked to assess the relevancy, the specificity, and the level of detail of annotations collected for six artworks by the games ARTigo Game, ARTigo Taboo, and Karido. Three of the six artworks were presented to the participants with eight annotations: four of the ten most frequently collected by ARTigo Game and four of the ten most frequently collected by ARTigo Taboo. The participants had to rank each annotation as “important”, “visible at first sight”, “broad”, and “generic” with respect to their associated artworks on five-level Likert scales.

To avoid so-called primacy-recency [7], or serial position, effects, the artworks respectively annotations were presented to each participant in random orders.

The Survey Participants. The survey was run on-line, in German,¹¹ and open to everyone over four months. It had been advertised in the ARTigo blog, in the blog of the research project play4science, and in social media. No experience with the ARTigo gaming platform was required for participating in the survey. The survey participants were not paid. As an incentive to participate, twelve Amazon vouchers à €25 were drawn as prizes. 159 people fully answered the survey of whom 77 were randomly assigned to to a first group, the remaining 82 to a second group. 60% of the participants were female, 2.5% younger than 20 years, 47% 20 to 29 years old, 20% 30 to 39 years old, 22% 40 to 59 years old, and 5.6% 60 years old or older.

Among the survey participants, there were almost as many ARTigo players as non-players: about 55% participants had already played on the ARTigo platform 87.5% of whom had played the game ARTigo Game, about 30% each of the diversification games ARTigo Taboo and Karido and 15% an integration game, Combino or Tag A Tag. 92% of the players had expressed an interest in art, 63% even a high interest in art. 30% of the non-players had expressed no or only some interest in art, 28% a high interest in art. 45% of the participants told that art, art history, or art education were related to their profession.

Results. In Section 1, the participants had to describe four artwork as well as possible with up to three annotations. Table 1 summarizes the answers. Most participants entered the requested three annotations. Half of the given annotations, but only a third for Artwork 3, had unique words, the other had words

¹¹ Thanks to a good media coverage over several years, the German version of ARTigo has a sufficient number of regular players for the data it has collected to be sufficient for this analysis. This still is not the case of the English and French versions of ARTigo because, so far, we could not invest the effort necessary for ARTigo to get sufficient a coverage in the English and French speaking media.

Table 1. Results of Section 1: Influence of time pressure.

Artwork	Annotations	Unique Words	Combined Annotations
1	227	93	10
2	213	105	15
3	214	70	3
4	227	103	13

that had been proposed by more than one participant. Although this was not requested, most participants proposed single word tags.¹²

In Section 2, the interesting question was whether the participants would identify among groups of four artworks the so-called initial artworks (see above) from the given annotations. Table 2 shows that this was the case of most participants for the two first artwork groups as well as for the second two groups of artworks. Interestingly, the best accuracy was achieved for the first group through the annotations collected by ARTigo Game, for the second group by Karido.

Table 2. Results of Section 2: “Semantic effectiveness” of the game Karido.

Artw. Set	Ann. Set	Choice				
		Init. Artw.	Artw. 2	Artw. 3	Artw. 4	no choice
1	Karido	88.31%	0%	7.79%	1.30%	2.60%
	ARTigo Game	93.90%	2.44%	0%	0%	3.66%
2	Karido	45.12%	0%	12.20%	23.17%	19.51%
	ARTigo Game	51.95%	1.30%	28.57%	6.49%	11.69%
3	Karido	86.59%	4.88%	2.44%	0%	6.10%
	ARTigo Game	15.58%	72.73%	3.90%	3.90%	3.90%
4	Karido	28.57%	22.08%	24.68%	14.29%	10.39%
	ARTigo Game	3.66%	50.00%	4.88%	28.05%	13.41%

In Section 3, the participants had to tell whether single word annotations collected by ARTigo Game or combined annotations collected by Tag A Tag better describe artworks. As Table 3 shows, most participants gave a slight preference to combined annotation.

In Section 4, the participants had to select those sets of annotations collected from different games, or different groups of games, that best describes three artworks. Table 4 on the next page shows that the best annotations *for all three* artworks had been collected from all games. However, Table 4 also shows that the best annotations for one artwork had been collected by ARTigo game, for another by the two diversification games, ARTigo Taboo and Karido. Since the annotations collected for an artworks by ARTigo Game only, for the other by the

¹² Possibly as a consequence of a widespread tagging practice on social media.

Table 3. Results of Section 3: “Semantic effectiveness” of the game Tag A Tag.

Artw.	Ann. Type	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
1	Combin.	64.78%	49.67%	51.41%	42.06%	47.86%	46.02%
	Single	35.22%	50.33%	48.59%	57.94%	52.14%	53.98%
2	Combin.	50.94%	47.74%	47.10%	53.79%	56.67%	41.23%
	Single	49.06%	52.26%	52.90%	46.21%	43.33%	58.77%
3	Combin.	66.04%	47.22%	54.69%	57.14%	38.61%	36.46%
	Single	33.96%	52.78%	45.31%	42.86%	61.39%	63.54%

diversification games only, performed extremely poorly each on an artwork, the gaming ecosystem as a whole can be seen as better performing than its parts.

Table 4. Results of Section 4: “Semantic effectiveness” of the ecosystem.

Artwork	Annotation Set			
	ARTigo Game (most frequent)	ARTigo Game (random)	Karido and ARTigo Taboo	all Games
1	4.40%	11.95%	58.49%	20.13%
2	30.82%	14.47%	6.92%	39.62%
3	46.54%	15.72%	22.64%	12.58%

In Section 5, the participants had evaluate the relevancy, semantic depth, specificity, and the level of detail of annotations collected for six artworks by the various games of the ARTigo gaming ecosystem. Table 5 shows that, for five of the six artworks, the annotations collected by the description game ARTigo Game were considered more important, better visible at first sight, broader, and more generic than the annotations collected by the diversification games ARTigo Taboo and Karido. For one of the six artworks, interestingly an abstract painting, the annotations collected by the diversification game Karido were rated as more important, better visible at first sight, broader, and more generic.

Interpretation. The results of Section 1 suggest that the time pressure induced by games does not have a significant influence on which tags are collected from humans by a human computation system. However, common sense suggests that, the deeper the semantics of annotations a game strives to collect from humans, the more counter-productive time pressure might be.

The results of Section 2 suggest that the diversification game Karido is “semantically effective” in the sense that it provides annotations useful for distinguishing from each other similar artworks. This confirms the intuition which led to the design of Karido. Note, however, that so far, for some artworks not enough annotations have been collected for Karido to perform as a diversification game

Table 5. Results of Section 5: Compared “Semantic effectiveness” of description and diversification games.

Dimension	Game	Artwork					
		1	2	3	4	5	6
Relevancy	ARTigo Game	76.62%	71.43%	67.99%	71.64%	49.35%	75.92%
	ARTigo Taboo	29.55%	31.17%	55.18%	-	-	-
	Karido	-	-	-	32.32%	59.09%	17.99%
Sem. depth	ARTigo Game	88.31%	77.60%	78.66%	73.48%	41.56%	85.67%
	ARTigo Taboo	19.48%	38.63%	67.68%	-	-	-
	Karido	-	-	-	29.57%	67.53%	24.09%
Detail	ARTigo Game	59.41%	55.19%	44.21%	43.60%	30.20%	58.23%
	ARTigo Taboo	21.10%	16.88%	44.82%	-	-	-
	Karido	-	-	-	20.73%	56.17%	11.59%
Specificity	ARTigo Game	69.80%	58.12%	57.71%	63.11%	39.61%	66.16%
	ARTigo Taboo	28.57%	37.34%	55.79%	-	-	-
	Karido	-	-	-	49.39%	53.90%	53.05%

on these artworks. On artworks with limited sets of annotations Karido performs like a description game [11].

The results of Section 3 suggest that Tag A Tag collects useful additional information on artworks. Depending on the artwork and its already collected annotations, Tag A Tag performs very differently, though. In some cases, Tag A Tag players hardly propose much, or new, annotations. In some cases, Tag A Tag players propose highly valuable annotation. Figure 5 is such an example. The survey participants found the annotation “woman” very expressive for this artwork, the combined annotation “woman + dress” not.¹³ The reason probably is that “dress” does not add much information to “woman” in the context of a 17th century European artwork. In contrast, the participants found the combined annotation “stirring + child” collected by the integration game Tag A Tag highly expressive. Interestingly, they did not found expressive for this artwork the single word annotation “stirring” collected by the description game ARTigo Game. The reason seems to be that, in this artwork, the “stirring” is not visible at first glance but only noticed after looking at the child. The combined annotation “stirring + child” collected by the integration game Tag A Tag brings the attention to the child achieving what is needed.

The results of Section 4 reinforce those of Section 3. They suggest that the gaming ecosystem ARTigo as a whole collects better artwork annotations than its parts. Even though games simpler to play like ARTigo Game, ARTigo Taboo, and Karido might have the lion share of annotation collecting, games more difficult to play like Tag A Tag and Combino do collect useful annotations, too.

Section 5 points to the usefulness of having each of ARTigo Game, Karido, and ARTigo Taboo in the gaming ecosystem. The annotations collected by the description game ARTigo Game are assessed as more important, more generic,

¹³ Translated from German. The original annotations are: Frau, Kleid, rühren, Kind.



Fig. 5. Godfried Schalcken: Der Rommelpotspieler, 1665-1670, Staatliche Kunsthalle Karlsruhe.

less detailed, and highly visible at first sight. The tags of the diversification games Karido and ARTigo Taboo are assessed as less relevant and less obvious but as more specific and more detailed.

4 Perspectives for Future Work: New Games With A Purpose

Inspired by the survey results presented in the former section, we suggest the following new description/integration games, that, probably, would enhanced the “semantic depth” of the artwork annotations collected by the ARTigo ecosystem.

A first game with codename Tag An Aspect 1 is a “scripted” [1] version of the ESP Game [10] (and therefore has the same gameplay as ARTigo Game, ARTigo Taboo, and Tag A Tag). The scripting consists in instructing the players to describe an aspect of the artwork. In contrast to data-collecting games formerly proposed like, among others, Peekaoom [17], the selected aspect is not necessarily a location but instead anything, from a location (like the boxes of Figure 5) to a concrete or abstract concept (like, referring again to Figure 5, “stirring + child”, “pair”, “communication”). Note that Tag An Aspect 1 would be both, a diversification and an integration game.

A second game with codename Tag An Aspect 2 would be a variation of Tag An Aspect 1 in which one player, a describer, selects an aspect unknown to the other player and describes it so as to make the other player, a guesser, capable of soon recognize, and name, that aspect. A more sophisticated gameplay, possibly giving the guesser the possibilities to ask queries, would most likely be necessary for “deep semantic” aspect being playable, that is, recognizable by a guesser. Like Tag An Aspect 1, Tag An Aspect 2 would be both, a diversification and an integration game.

Third and a fourth games with codename Unveil 1 and Unveil 2 would be “scripted” [1] version of Tag An Aspect 1 and Tag An Aspect 2 respectively explicitly requesting the players to play with hidden, or hardly recognizable aspects. Working out gameplays for Unveil 1 and Unveil 2 that would be “playable” and would result in a good collecting of semantic data would be not trivial. Like both Tag An Aspect games, the two Unveil games would be both, diversification and an integration games.

Note that the four afore mentioned games would not only enhance a gaming data-collecting ecosystem, but also only be possible within a gaming ecosystem like ARTigo. Indeed, these four new games like current diversification and integration games of ARTigo require seed data.

We are also considering an analysis of the *linguistic* quality of the annotations collected by the ARTigo ecosystem and an analysis of the ecosystem’s efficiency defined as number of annotation per hour of human activity, possibly weighted by a measure of the annotations’ semantic values.

5 Conclusion

This article has first briefly described the ARTigo gaming ecosystem aiming at collecting “deep semantic” descriptions of artworks. To the best of the knowledge of the authors, ARTigo is the first ever proposed, conceived, and deployed ecosystem of data-collecting games, or games with a purpose [10]. Ecosystem means here that the games and the gaming system are conceived in such a manner that some of the data-collecting game collect from the footprints left by players the seed data necessary for other data-collecting games to be playable.

This article has then reported on an analysis pointing to the effectiveness of the ARTigo gaming ecosystem at collecting “deep semantic” data. Not only do each of the games of ARTigo collect useful semantic artwork descriptions, but also (1) the ecosystem as a whole performs better than its parts and (2) the seed data needed by some games are provided by other games of the ecosystem.

Finally, this article has suggested four novel data-collecting games, or games with a purpose [10], that, most likely, would enhance an ecosystem like ARTigo by collecting complementary “deep semantic” data.

Acknowledgements

The research reported about in this article as been founded in part by the German Foundation for Research (DFG, Deutsche Forschungsgemeinschaft) within the project play4science.

The authors are thankful to Dr. Gerhard Schön for the implementation of a first prototype, to Dr. Christoph Wieser for his essential contributions to the conception and implementation of ARTigo’s gaming platform and search engine, and to Prof. Dr. Hubertus Kohle, art historian at the University of Munich, for useful suggestions.

References

1. Bry, F., Wieser, C.: Squaring and scripting the ESP game. In: Proc. of the 4th Human Computation Workshop (HCOMP). AAAI Press, Palo Alto, CA, USA, 2012
2. Bry, F., Wieser, C.: Squaring and scripting the ESP game: Trimming a GWAP to deep semantics. In: Proc. of the International Conference on Serious Games Development and Applications (SGDA). pp. 183–192. Springer, Berlin, Germany, 2012
3. Golbeck, J., Koepfler, J., Emmerling, B.: An experimental study of social tagging behavior and image content. *Journal of the American Society for Information Science and Technology* 62(9), pp. 1750–1760, 2011
4. Lawley, L.: Social consequences of social tagging, 2005 http://many.corante.com/archives/2005/01/20/social_consequences_of_social_tagging.php, accessed: January 12, 2015
5. Schemainda, C.: Qualitative Analysis of the ARTigo Gaming Ecosystem. bachelor thesis, Institute of Computer Science, Ludwig-Maximilian University Munich, 2014
6. Steinmayr, B.: Designing Image Labeling Games For More Informative Tags. diploma thesis, Institute of Computer Science, Ludwig-Maximilian University Munich, 2010
7. Tourangeau, R., Couper, M.P., Conrad, F.: Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly* 68(3), 368–393 (2004)
8. Steinmayr, B., Wieser, C., Kneißl, F., Bry, F.: Karido: A GWAP for telling artworks apart. In: *Computer Games (CGAMES)*, 2011 16th International Conference on. pp. 193–200. IEEE, 2011
9. Trant, J., with the Participants in the steve.museum Project: Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia* 12(1), pp. 83–105, 2006
10. Von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI). pp. 319–326. ACM, 2004
11. Wieser, C., Bry, F., Bérard, A., Lagrange, R.: Artigo: Building an artwork search engine with games and higher-order latent semantic analysis. In: Proc. of Disco 2013, Workshop on Human Computation and Machine Learning in Games (HComp). AAAI Press, 2013
12. Shah, P., Wieser, C. Bry, F.: Parallel Higher-Order SVD for Tag-Recommendations. In: Proc. the International Conference WWW/Internet, 2012
13. Shaili J., Parkes, D. C.: A Game-theoretic Analysis of Games with a Purpose, *ACM Transactions on Economics and Computation*, Vol. 1, Nb. 1, Article 3, 2013
14. Robertson, S., Vojnovic, M., Weber, I.: Rethinking the ESP Game, Proc. of the Conference on Human Factors in Computing Systems (CHI), 2009
15. Von Ahn, L., Dabbish, L.: Designing Games With a Purpose, *Communications of the ACM*, Vol. 51, Issue 8, pp. 58–67, 2008
16. Wieser, C.: Building a Semantic Search Engine with Games and Crowdsourcing, Doctoral Thesis, Institute for Informatics, University of Munich, 2014
17. Von Ahn, L., Ruoran Liu, and Blum, M.: Peekaboom: A Game for Locating Objects in Images, Proc. of the Conference on Human Factors in Computing Systems (CHI), pp. 55–64, 2006