

# **Abschlussbericht zum Projekt**

## **Entwicklung sozialer Web-Plattformen zur Datengewinnung in den Geisteswissenschaften (KO 1091/4-2)**

### **Verantwortlicher Antragsteller:**

Prof. Dr. Hubertus Kohle / Institut für Kunstgeschichte der LMU München

### **Weitere Antragsteller:**

Prof. Dr. François Bry / Institut für Informatik der LMU München

Prof. Dr. Thomas Krefeld / Institut für Romanische Philologie der LMU München

Dr. Christian Riepl / IT-Gruppe Geisteswissenschaften der LMU München

Prof. Dr. Klaus Schulz / Centrum für Informations- und Sprachverarbeitung der LMU München

### **Berichts- und Förderzeitraum insgesamt:**

1.10.2010 - 30.9.2013

### **Veröffentlichungen**

#### **2010**

François Bry, Fabian Kneißl, Christoph Wieser, Field Research for Humanities with Social Media: Crowdsourcing and Algorithmic Data Analysis. In: Proceedings of 4. Workshop Digitale Soziale Netze im Rahmen der 41. Jahrestagung der Gesellschaft für Informatik (GI), Berlin, Germany (4th - 7th October 2011):

[http://www.pms.ifi.lmu.de/publikationen/PMS-FB/PMS-FB-2011-3/PMS-FB-2011-3-paper.pdf/](http://www.pms.ifi.lmu.de/publikationen/PMS-FB/PMS-FB-2011-3/PMS-FB-2011-3-paper.pdf) <http://digitale-soziale-netze.de/gi-workshop/index.php>

Hubertus Kohle, Artigo: social image tagging pour les oeuvres d'art, in: Béatrice Joyeux-Prunel (Hg.), L'art et la mesure. Histoire de l'art et méthodes quantitatives, Paris 2010, S. 153-164.

#### **2011**

Hubertus Kohle, Kunstgeschichte goes Social Media. Laien optimieren eine Bilddatenbank – mit einem digitalen Spiel, in: Aviso. Zeitschrift für Wissenschaft und Kunst in Bayern, 3/2011, S. 37-43.

Thomas Krefeld, Alter Standard – Neue Medien. Zur Erfassung von Restandardisierungsprozessen im Italienischen in: Dessì Schmid, Sarah (u.a.) (Hrsg.), Koineisierung und Standardisierung in der Romania, Heidelberg 2011, S. 269-281.

Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneißl, and François Bry, Karido: A GWAP for Telling Artworks Apart. In: Proceedings of 16th International Conference on Computer Games (CGAMES2011), Louisville, KY, USA (27th - 30th July 2011), Best Paper Award.

## **2012**

François Bry and Christoph Wieser, Squaring and Scripting the ESP Game: Trimming a GWAP to Deep Semantics. In: Proceedings of the International Conference on Serious Games Development and Applications (SGDA), Bremen, Germany (26-29 September 2012).

Fabian Kneissl and François Bry, Speculation and Bluff for Linguistic Field Research - Social Software and Media for Data Gathering. In: Proceedings of 5. Workshop Digitale Soziale Netze im Rahmen der 42. Jahrestagung der Gesellschaft für Informatik (GI), Braunschweig, Germany (16th - 21st September 2012), Best Paper Award.

Hubertus Kohle, Crowdsourcing in der Wissenschaft. Wie die Massen zum Wissenschaftler werden, in: telepolis, 18.1.2012:

<http://www.heise.de/tp/artikel/36/36239/1.html>

## **2013**

François Bry, Fabian Kneissl, Thomas Krefeld, Stephan Lücke, and Christoph Wieser, Crowdsourcing for a Geographical and Social Mapping of Italian Dialects. In: Proceedings of 2nd International Workshop on Social Media for Crowdsourcing and Human Computation, Paris, France (1st May 2013).

Hubertus Kohle, Digitale Bildwissenschaft, Glückstadt, 2013.

Christoph Wieser, François Bry, Alexandre Bérard, and Richard Lagrange, ARTigo: Building an Artwork Search Engine With Games and Higher-Order Latent Semantic Analysis. In Proceedings of Disco 2013, Workshop on Human Computation and Machine Learning in Games at HComp 2013, Palm Springs, CA, USA (9 November 2013).

## **Arbeits- und Ergebnisbericht**

### **Allgemeines**

Zielsetzung des Projektes war eine neue, über computergestützte Crowdsourcing-Verfahren generierte Datenerhebung in der Kunstgeschichte und der italienischen Linguistik. Dafür wurden Fördermittel zur Entwicklung von sozialer Software beantragt, die den Betrieb von zwei Web-Plattformen ermöglicht, auf denen interessierte Laien Eingaben machen können. Um die Benutzeroberflächen attraktiv zu gestalten und dadurch die Beiträge der Laien sicherzustellen, sollten diese Web-Plattformen als Spielumgebungen (GWAP) realisiert werden, deren kompetitiver Charakter den Reiz zur Dateneingabe erhöht. Während der Förderphase sind, wie geplant, die beiden Web-Plattformen namens Artigo (<http://artigo.org>) und Metropolitalia (<http://metropolitalia.org>) weiter konzipiert bzw. von Null an implementiert, erprobt und letztlich eingesetzt worden. Eine beachtliche Menge an wertvollen Daten wurde bereits gesammelt und auch mit deren Analyse wurde begonnen. Begleitend zu den Spielen (und über den Antrag hinausgehend) wurden zwei weblogs aufgesetzt, die zur Information der Nutzer und zur Sammlung von anregenden Mitteilungen/Kritik etc. verwendet werden. Um das Projekt prägnant und öffentlichkeitswirksam auf einen Namen zu bringen, wurde die Bezeichnung „play4science“ eingeführt. Eine Projekt-Homepage (<http://www.play4science.lmu.de>) gibt Auskunft über die Öffentlichkeitsarbeit, Workshops, Publikationen, etc.

Über die Dauer der Drittmittelförderung hinaus läuft die Web-Plattform mit den beiden Spielen Artigo und Metropolitalia auf den Servern der IT-Gruppe Geisteswissenschaften der LMU weiter und wird gemäß Antrag unter der Leitung von Dr. Christian Riepl von dem dort unbefristet beschäftigten Personal gewartet. Damit ist eine nachhaltige Verfügbarkeit, interdisziplinäre Nachnutzbarkeit und universelle Transformierbarkeit der Daten gewährleistet. Die Datensammlungen können weiter ergänzt werden. Die automatische Nutzung und Auswertung der gesammelten Daten zur Verbesserung der Kunstwerk-Suchmaschinen sowie zur geolinguistischen Erforschung des italienischen Sprachraums können fortgesetzt werden.

Die Software wird als Open Source unter der GNU Affero General Public License (AGPL), der von der Free Software Foundation (FSF) präferierten Lizenz für Software, die auf einem Server läuft, auf der OpenSource-Plattform GitHub kostenlos zugänglich gemacht.

## Artigo

Ziel des kunstgeschichtlichen Teils der Anwendung ([www.artigo.org](http://www.artigo.org)) war es, Annotationen zu Kunstwerken zu generieren, die später zur Suche nach diesen Kunstwerken verwendet werden können. Im Laufe des Berichtszeitraumes konnten auf diesem Weg insgesamt 7 Millionen tags gesammelt werden, die jeweils unmittelbar nach Dateneingabe suchbar werden. Um diese Zahlen zu erreichen, waren eine gefällige und dramaturgisch wirksame Gestaltung der Spieloberfläche sowie eine gute Server-Architektur nötig. Eine Bestenliste zielt darauf ab, die Aktivität der Nutzer zu steigern, da naheliegt, dass diese in der Konkurrenzsituation mit den Mitspielern „bei der Ehre gepackt“ werden. Als Anregung ist auch das täglich wechselnde Teaserbild gedacht, das vor Spielbeginn eingeblendet und mit kurzen und prägnanten Kommentaren versehen wird. Wir gehen dabei davon aus - und das wird von den automatischen Nutzungsberichten bestätigt -, mit einer großen Anzahl von „Seiteneinsteigern“ in das Spiel konfrontiert zu sein, die man auf diese Weise eventuell zu längeren Verweilzeiten animieren kann.

Der Erfolg des Spiels steht und fällt mit der öffentlichen Resonanz: Die intensive, im Antrag vor allem durch Hilfskraftmittel angestrebte Öffentlichkeitsarbeit hatte einen insgesamt überwältigenden Erfolg. Bei Projektende konnten wir ca. 180.000 Mitspieler aufweisen, die entweder als registrierte (ca.18.000) oder als nicht-registrierte Nutzer beteiligt waren und zum Großteil immer noch sind. Täglich spielen im Durchschnitt 150 Personen mit. Dieser Erfolg kam auch durch eine umfängliche Pressereaktion zustande, die wohl in erster Linie auf das zuletzt hochaktuell gewordene Konzept einer *citizen science* (vgl. <http://www.citizen-science-germany.de/>) zurückzuführen ist. Aus der Berichterstattung sei hier nur verwiesen auf:

- die *Deutsche Universitätszeitung* (Die Kraft des Schwarms, 03/2012, S. 32 - auch in: Spiegel-online, 20.3.2012: <http://www.spiegel.de/unispiegel/jobundberuf/lmu-muenchen-laesst-mit-crowdsourcing-bilder-verschlagworten-a-817439.html>),
- die ZEIT (Zocken für die Forschung, 12.9.2013, S. 44),
- die *Süddeutsche Zeitung* (Das Wissen der Massen, 20.2.2012, S. 51) und
- die *Stuttgarter Zeitung* (12.4.2012, S. 29).

Daneben haben sowohl die Radio- als auch die Fernsehmedien über das Spiel berichtet (Interviews in Bayern 2, 26.8.2011, RBB Kulturradio, 18.5.2011, längerer Bericht in ZDF nano, 10.12.2012). Auch durch die intensive Vortragstätigkeit aller Projektbeteiligten konnte eine Öffentlichkeitswirkung erzielt werden (Vorträge in

Berlin, Bonn, Erlangen, Venedig, Köln, Los Angeles, München, Nürnberg, Rotterdam u.a.).

Im Einzelnen konnten folgende beantragte Schritte realisiert werden:

Schon vor Projektbeginn war ein von Dr. Gerhard Schön (IT-Gruppe Geisteswissenschaften) als *proof of concept* erstellter Prototyp des Spieles Artigo vorhanden, mit dem ab 2007 das Konzept überprüft und eine Nutzergemeinschaft von ca. 1500 Mitspielern aufgebaut werden konnte. Der ausschließlich unter Verwendung der Programmiersprachen Javascript und PHP so schlank wie möglich erstellte Prototyp wurde dann im Projekt auf der Grundlage des hochaktuellen Java-basierten Frameworks "JBoss Seam" vollständig neu implementiert. Dabei konnten nicht nur die Verwaltung der Nutzerdaten, Nachrichten und Annotationen neu organisiert, sowie soziale Dienste für die Spieler eingerichtet werden, es wurde auch eine Skalierung der Plattform durch Lastverteilung auf mehrere Webserver realisiert, um die hohen Zugriffsraten, die sich vor allem nach Presseberichten in vielfrequentierte Publikumszeitschriften ergeben haben, abfedern zu können. Schließlich wurde das Grundsystem um konzeptuelle Schnittstellen ergänzt, welche die Erweiterung der Plattform nunmehr ganz erheblich vereinfachen. Der Aufwand für die Implementierung neuer Spiele mit abweichender Spiellogik wird damit um ca. 90% reduziert.

Im Laufe der Projektarbeit zeigte sich, dass nicht alle der im Antrag genannten Spielversionen sinnvoll waren, andere, damals noch nicht angedachte, dagegen schon. Zum einen wurden zwei Basis-Spiele realisiert, die in einem Fall mit, im anderen ohne Tabu-Wörter operieren, also einmal eine völlig freie Dateneingabe ermöglichen, das andere Mal nur neue, noch nicht vergebene Begriffe akzeptieren. Die Unterscheidung hat vor allem Konsequenzen für die Datenanalyse. In der Version, die ohne Tabuwörter operiert, kann die Anzahl der gleich vergebenen Annotationen für eine Bewertung von deren Signifikanz verwendet werden. Die Tabu-Version dagegen steigert die Qualität und Diversität der Annotationen. Das im Rahmen einer informatischen Diplomarbeit entstandene Spiel *karido* steigert die Interaktivität und erreicht damit eine Präzisierung der Dateneingaben. Besonders interessant, wenn auch bislang wenig genutzt, scheint uns *combinio*, das in letzter Konsequenz und Ausbaustufe den Prozess der Zerschlagung von Beschreibungen in Einzelbegriffe, wie er sich in der Spiellogik realisiert, wieder umkehrt und aus den Annotationen Zug um Zug Diskurse generiert. Konkret sollen bei *combinio* zwei zusammengehörige Annotationen, die aber einzeln vergeben wurden, zusammengeführt werden („blau“ und „Pferd“ in einem Bild von Franz Marc zu „blaues Pferd“).

Nachdem zu Beginn des Projektes mit den am Institut für Kunstgeschichte der LMU vorhandenen Beständen der Bilddatenbank gearbeitet wurde, konnten wir diese durch externe Angebote des Rijksmuseums in Amsterdam, der Karlsruher Kunsthalle und des Universitätsmuseums Amherst/Massachusetts erweitern, die inzwischen immer häufiger über offene Schnittstellen zur Verfügung gestellt werden. Strategisch bedeutsam für das Zusammenwachsen der diversen Bestände im bundesdeutschen Kontext ist auch die Tatsache, dass Teile des national und partiell auch international führenden kunsthistorischen Bilddatenbankangebotes (prometheus) ebenfalls über Artigo annotiert werden ([www.prometheus-bildarchiv.de](http://www.prometheus-bildarchiv.de)).

Folgeuntersuchungen sind vor allem in einem Bereich geplant, der von dem ganz auf die Datengewinnung abzielenden Antrag nicht abgedeckt war. Es hat sich im Projektverlauf herausgestellt, dass die Annotationen nicht nur als Suchinstrumente für große Bilddatenbanken zu verwenden sind, sondern dass sie ein Medium eigenen Rechtes sind. Es weist vieles darauf hin, dass sie z.B. Aufschluss über die

geschlechtliche, kulturelle und generationelle Identität des Datenlieferanten geben, vor allem aber auch, dass sie implizites Wissen über die annotierten Kunstwerke enthalten. So kann eine vergleichende Analyse der Annotationen zur Auffindung von „ähnlichen“ Bildern verwendet werden, eine Datenanalyse der Kombination von Beschreibungsdaten könnte aber auch die epochale bzw. stilgeschichtliche Zuordnung der Kunstwerke beinhalten, ohne dass diese selber ausdrücklich genannt wird. Bemerkenswert ist, dass Artigo dank seiner Suchmaschine immer häufiger über die allgemeinen Suchmaschinen von Personen erreicht wird, die mit Suchanfragen wie „Abstrakte Bleistift Zeichnungen“, „Alhambra Südfassade“ oder „Arnold Böcklin der Kampf auf der Brücke“ Interesse an Kunstwerken zeigen.

Wie bei allen digital gestützten Projekten, in denen bei offener Architektur zusätzliche funktionale *layer* über Vorhandenes gelegt werden können, bieten sich andere Erweiterungen an, die eine tagging-Umgebung in eine umfassende, interaktive Kunst-Plattform verwandeln könnten. Naheliegend wäre etwa das Projekt, den Spielern die Möglichkeit anzubieten, mit den von ihnen selbst generierten tags eigene Werkzusammenstellungen vorzunehmen, diese zu diskutieren und zuletzt eigene Kunstausstellungen im Internet zu realisieren.

Eine wirtschaftliche Verwertung von Artigo ist bei massiver Ausweitung der Nutzerzahl insofern denkbar, als hier eine auf ein sehr präzise zu bestimmendes Nutzerpublikum fokussierte Werbeplattform entstehen könnte. Diese aber würde höchstwahrscheinlich die wissenschaftlichen Zielsetzungen des Projektes in zwei Weisen gefährden. Erstens erfolgte damit eine Fokussierung auf andere Inhalte. Zweitens dürfte die Laien-Gemeinschaft der Artigo-Nutzer negativ auf eine wirtschaftliche Nutzung der Plattform reagieren. Diese zweite Gefahr ist bei citizen science-Vorhaben nicht zu unterschätzen, da sie vom Idealismus ihrer Nutzer-Gemeinschaften leben.

## **Metropolitalia**

Auf Grund seiner radikal innovativen Konzeption markiert Metropolitalia eine wichtige Etappe bei der Überführung der etablierten Variations- und Varietätenlinguistik in die Digital Humanities; die Plattform erhielt beim *Ars Electronica Award 2013* eine *Honorary Mention*. Im Einzelnen sind technologische, methodologische und einzelsprachliche Aspekte zu unterscheiden, die zwar unbedingt vor dem Hintergrund der aktuellen Forschung zu sehen, aber die bis zu einem gewissen Grad auch experimentell zu werten sind.

In technologischer Hinsicht wurden sehr nützliche und wiederwendbare Tools entwickelt; sie dienen der Erhebung und Etikettierung geolinguistisch relevanter Sprachdaten, ihrer Verwaltung und quantitativen Auswertung sowie ihrer nutzerfreundlichen und laientauglichen Visualisierung auf einer georeferenzierten kartographischen Basis.

In methodologischer Hinsicht schließt das Projekt an die derzeit stark diskutierte perzeptive Neuausrichtung der Varietätenlinguistik an. Im Zentrum dieses Paradigmas steht die Erschließung des epilinguistischen Sprecherwissens, d.h. des Wissens des Laien über seine eigenen Regionalvarietäten (Autoperzeption) aber auch über die von Anderen benutzten Varietäten (Heteroperzeption). Dieses Wissen muss grundsätzlich in die Klassifikation und 'Zonierung' der Sprachräume einfließen, die bislang weitestgehend 'von außen', d.h. durch den Sprachwissenschaftler erfolgt ist. Darüberhinaus berücksichtigt die Konzeption von Metropolitalia die besondere

Rolle der räumlichen Variation: Sie ist in vielen Sprachgemeinschaften quasi allgegenwärtig und daher den anderen Dimensionen, wie zum Beispiel der sozialen Schicht oder dem Alter, vorgeschaltet; darüber hinaus sind räumlich markierte Formen sekundär mit andern Markierungen sozialer, generationeller und geschlechtsspezifischer Art verknüpft. Entsprechende Assoziationen werden daher auf dem Portal ebenfalls erhoben. Die Erschließung des Laienwissens verbindet Artigo und Metropolitalia.

In einzelsprachlicher Hinsicht entspricht das Projekt einem bekannten Desiderat, insofern es weniger auf die lokalen Basisdialekte zielt (die jedoch keineswegs ausgeschlossen sind), sondern in erster Linie die weiter verbreiteten und gelegentlich sogar standardhaften Formen erfassen will (das so genannte Regionalitalienisch). Während die lokalen Dialekte Italiens Gegenstand einer außerordentlich reichhaltigen Forschungstradition sind, ist über die Existenz komplementärer Regionalstandards praktisch nichts bekannt; hierfür ist gerade die von Metropolitalia fokussierte Divergenz von Selbstwahrnehmung und Fremdwahrnehmung ein entscheidendes Kriterium, denn die einschlägigen Merkmale sind in der Regel nur für diejenigen salient, die sie nicht selbst gebrauchen; für ihre Benutzer sind sie qua Standard eben unauffällig. Metropolitalia ist mit anderen an der LMU angesiedelten, teils DFG-geförderten geolinguistischen Projekten von Stephan Lücke und Thomas Krefeld verknüpft und trägt zum Aufbau eines umfassenden digitalen Observatoriums des italienischsprachigen Raums bei (vgl. *Atlante linguistico digitale dell'Italia e della Svizzera meridionale*, <http://www.adis.gwi.uni-muenchen.de/AIS.php>; *Atlante sintattico della Calabria*, <http://www.asica.gwi.uni-muenchen.de/>, DFG-Kennziffern KR 1730/3-1 und KR 1730/2-2). Metropolitalia unterscheidet sich von den genannten Projekten jedoch durch seinen absolut experimentellen Charakter, der in seiner vollständigen und absoluten Fundierung in den Sozialen Medien zu sehen ist: Metropolitalia generiert – im Unterschied zu Artigo – sowohl Primär- als auch Metadaten, ohne auf plattformexterne Datenbestände zurückzugreifen; die Plattform ist in gewisser Hinsicht auf mediale Selbstgenügsamkeit hin angelegt. Dazu gehört auch der Versuch, durch internetspezifische Kommunikationsformen Nutzer zu gewinnen; eine besondere Strategie bestand darin, einzelne sprachliche Formen oder Ausdrucksweisen des Datenbestands zum Aufhänger von Blogs zu machen, die sich über das Sprachliche hinaus, mit der italienischen Aktualität befassen. Diese Blogs können über Twitter und Facebook kommentiert werden. Die Web-Plattform Metropolitalia ist erst vor einem Jahr veröffentlicht worden. Die Zahl ihrer Nutzer und folglich auch der mit der Plattform gesammelten Daten ist dementsprechend noch gering. Die Erfahrung mit den meisten Crowdsourcing-Projekten lehrt aber, dass sich die Anfänge lang hinziehen. Erst nach mindestens fünfjährigem Betrieb wird zu ermitteln sein, ob die Plattform erfolgreich ist oder nicht. Aus diesem Grund ist die sprachwissenschaftliche Auswertung der bislang verfügbaren Datenbestände noch nicht angebracht.

Nachfolgeprojekte sollten mindestens in zwei unterschiedliche Richtungen gehen:

1. Im Sinne des digitalen Observatoriums der italienischen Gegenwartssprache ist es sinnvoll, den Datenbestand durch systematische Extraktion bereits im Internet verfügbarer regionalsprachlicher Formen zu ergänzen; dazu sollten alle über die Spielplattform gesammelten Formen systematisch zum Gegenstand von Suchmaschinenabfragen werden; da markierte Sprachmerkmale im Diskurs in aller Regel nicht isoliert vorkommen, ist in den jeweiligen Vorkommenskontexten mit kookkurierenden anderen Merkmalen identischer Markierung zu rechnen. In regionaler Hinsicht wäre eine Erweiterung durch gezielte Adressierung der italienischen

Migranten vielversprechend (Deutschland, Schweiz, Belgien, Großbritannien, Brasilien, Australien, Kanada USA u.a.), denn Migrationsvarietäten sind auf Grund ihrer konservativen Natur einerseits und andererseits auf Grund ihrer jeweiligen kontaktsprachlichen Einflüsse von besonderem sprachwissenschaftlichen Interesse. Dasselbe gilt analog für italienische Immigrationshintergründe der 2. Generation, also etwa das Italienische von Abkömmlingen der rumänischen, albanischen, westafrikanischen und chinesischen Migranten in Italien.

2. Wegen der bestens erprobten Leistungsfähigkeit sollten die robusten Tools auch auf andere Einzelsprachen angewandt werden; aus romanistischer Sicht von besonderem Interesse sind die so genannten plurizentrischen Sprachen, das heisst die romanischen Sprachen mit mehreren nationalen Normen, also die unterschiedlichen Ausprägungen des Spanischen in den südamerikanischen Staaten, das Portugiesische in Brasilien und Portugal, das Französisch in Québec, Westafrika usw.

## **Zusammenfassung**

Mit Artigo und Metropolitalia sind zwei Spiele realisiert worden, mit der eine Laienöffentlichkeit in den wissenschaftlichen Prozess eingebunden werden soll. Im ersten Fall wird diese Öffentlichkeit adressiert, um deren Kompetenzen in der Beschreibung und Deutung von Kunstwerken fruchtbar zu machen. Im zweiten Fall geht es um Einschätzungen italienischer Muttersprachler bei der regionalen und lokalen Zuordnung sprachlicher Formen. Das ganz praktische Ziel von Artigo ist es, mit diesem crowdsourcing genannten Verfahren in möglichst kurzer Zeit eine möglichst große Menge von Begriffen zu erhalten, mit denen in großen Bilddatenbanken nach bestimmten Kunstwerken gesucht werden kann. Metropolitalia dient Erkenntnissen der Varietätenlinguistik in einem dialektal hoch ausdifferenzierten Land wie Italien.

Der grundsätzliche Ansatz des Projektes schreibt sich ein in die hochaktuelle sogenannte citizen science, in der das Internet genutzt wird, um Menschen zur Übernahme von Aufgaben zu bringen, die vom Computer allein nicht zu lösen sind. Die Ergebnisse sind beeindruckend und in ihrer innovativen Entstehungsweise auch von der Presse intensiv kommentiert worden. So konnten wir etwa im Fall von Artigo im Projektzeitraum an die 7 Millionen Annotationen sammeln, die unmittelbar nach Eingabe für die Suche zur Verfügung standen. Da die Spiele auch über den Projektzeitraum hinaus weiterlaufen, rechnen wir mit einer erheblich größeren Anzahl im Lauf der nächsten Jahre. In einem zweiten Schritt sollen zudem die Potentiale der gesammelten Annotationen datenanalytisch und auf implizites Wissen hin untersucht werden. So wäre etwa zu fragen, ob eine bestimmte Annotationsmenge in ihrer Spezifik Hinweise auf die Entstehungszeit eines Kunstwerkes liefern kann. Grundsätzlich glauben wir, mit der Plattform einen Ansatz realisiert zu haben, der auch in anderen empirisch fundierten Geistes- und Sozialwissenschaften zur Anwendung kommen kann.

## **Förderung des wissenschaftlichen Nachwuchses**

Im Rahmen des Projekts dokumentieren eine Reihe von Studien-, Magister- und Diplomarbeiten sowie Dissertationen die Förderung des wissenschaftlichen Nachwuchses. Zum Teil sind Schriften auch unter Beteiligung von Diplomanden und Doktoranden entstanden, eine davon erhielt die Auszeichnung „Best Paper Award“.

Laura Commare, Social tagging in der Kunstgeschichte. Magisterarbeit am Institut für Kunstgeschichte der LMU München, 2011.

Silvia Cramerotti, Zur subjektiven Beurteilung und Erkennbarkeit italienischer Stadtakzente. Magisterarbeit am Institut für Romanistik der LMU München, 2011:  
<http://epub.ub.uni-muenchen.de/12717/>

Elena Levushkina, Computerlinguistische Methoden in Community-basierten Anwendungen, Dissertation 2013.

Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneißl, and François Bry, Karido: A GWAP for Telling Artworks Apart. In: Proceedings of 16th International Conference on Computer Games (CGAMES2011), Louisville, KY, USA (27th - 30th July 2011). Best Paper Award.

Dissertation von Fabian Kneißl, in Vorbereitung, Disputation in 2014.

Dissertation von Christoph Wieser, in Vorbereitung, Disputation in 2014.

Bartholomäus Steinmayr, Designing Image Labeling Games For More Informative Tags, Diplomarbeit, 2011:  
[http://www.pms.ifi.lmu.de/publikationen/diplomarbeiten/Bartholomaeus.Steinmayr/DA\\_Bartholomaeus.Steinmayr.pdf](http://www.pms.ifi.lmu.de/publikationen/diplomarbeiten/Bartholomaeus.Steinmayr/DA_Bartholomaeus.Steinmayr.pdf)

Mislav Boras, Integration des e-learning GWAPs Termina in das digitale Backchannel Backstage, Institute for Informatics, LMU, Munich, Bachelorarbeit, 2013:  
[http://www.pms.ifi.lmu.de/publikationen/#PA\\_Mislav.Boras](http://www.pms.ifi.lmu.de/publikationen/#PA_Mislav.Boras)

Florian Stoerkle, Combino - A GWAP for Generating Combined Tags, Institute for Informatics, LMU, Munich, Bachelorarbeit, 2012:  
[http://www.pms.ifi.lmu.de/publikationen/#PA\\_Florian.Stoerkle](http://www.pms.ifi.lmu.de/publikationen/#PA_Florian.Stoerkle)

Diego Havenstein, Towards MapReduce Algorithms for the Higher Order-Singular Value Decomposition, Institute for Informatics, LMU, Munich, Bachelorarbeit, 2012:  
[http://www.pms.ifi.lmu.de/publikationen/#PA\\_Diego.Havenstein](http://www.pms.ifi.lmu.de/publikationen/#PA_Diego.Havenstein)

Philipp Langhans, JSMapReduce - Distributed Computing with Web Clients and LAMP, Institute for Informatics, LMU, Munich, Bachelorarbeit, 2012:  
[http://www.pms.ifi.lmu.de/publikationen/#PA\\_Philipp.Langhans](http://www.pms.ifi.lmu.de/publikationen/#PA_Philipp.Langhans)