

-01_1299013241.jpg	1	195	63.1859	138	48.8354	21	100.7660
-02_1299052989.jpg	2	60	70.3173	164	89.6647	171	75.3399
-02_1299095043.jpg	3	119	79.8049	77	93.1814	31	102.3130
-04_1299210705.jpg	4	230	86.1152	32	39.1059	42	46.2573
-15_1300202721.jpg	5	161	59.8678	150	78.2054	32	110.5277
-20_1300582870.jpg	6	10	80.0254	153	105.1385	8	106
-30_1301499780.jpg	7	89	77.0073	75	85.5518	156	75.8126
-01_1301652317.jpg	8	128	71.8385	80	62.4530	15	105.6218
-01_1301657769.jpg	9	171	62.8986	81	50.2747	18	66.5246
-01_1301672120.jpg	10	218	90.9555	93	73.5556	46	107.9347
-02_1301760330.jpg	11	169	67.7050	150	67.3664	26	29.5686
-03_1301840682.jpg	12	125	82.1643	130	86.6904	106	105.7571
-04_1301911906.jpg	13	65	81.0029	84	98.1270	134	72.4929
-11_1302549887.jpg	14	84	52.9958	147	71.2858	28	110.6424
-12_1302605058.jpg	15	98	68.5011	102	80.5425	230	115.3889
-16_1302951188.jpg	16	11	51.6651	96	71.0357	31	112.6502
-20_1303311589.jpg	17	186	53.7728	58	46.5876	5	24.4844
-21_1303315577.jpg	18	11	53.8055	86	64.8374	22	56.6996
-30_1304020202.jpg	19	11	53.8055	86	64.8374	22	56.6996
-02_1304326775.jpg	20	115	57.4544	38	86.4643	77	82.4598
-05_1304612311.jpg	21	115	57.4544	38	86.4643	77	82.4598
-13_1305295333.jpg	22	115	57.4544	38	86.4643	77	82.4598
-15_1305459987.jpg	23	136	57.8064	30	51.5302	21	91.0131
-17_1305631733.jpg	24	79	100.8371	44	82.0106	21	73.1
-18_1305748725.jpg	25	222	52.9326	154	76.2578	222	117.0254
-23_1306160289.jpg	26	130	83.2750	152	86.9124	13	73.1209
-01_1306943005.jpg	27	165	59.1000	137	60.2206	37	85.8111
-02_1307044273.jpg	28	13	61.2556	100	49.2678	166	41.7378
-05_1307306827.jpg	29	173	55.3228	130	40.511	22	110.4440
-08_1307562209.jpg	30	93	52.8000	127	62.3750	162	75.3981
-08_1307563616.jpg	31	52	92.0000	05	92.0630	34	87.2074
-14_1308053731.jpg	32	108	72.6018	128	72.6018	2	102.0115
-23_1308819860.jpg	33	65	84.9442	159	95.0027	165	74.3665
-25_1309013799.jpg	34	125	80.7868	118	84.5030	164	99.1030
-03_1309689125.jpg	35	112	60.8313	130	70.5844	99	106.9052
-04_1309778814.jpg	36	53	82.4300	131	52.0381	124	39.1552
-04_1309791288.jpg	37	116	62.7000	171	80.5627	239	95.2126
-06_1309940749.jpg	38	102	72.5487	104	61.6707	170	93.0487
-09_1310223035.jpg	39	75	73.2427	238	97.5374	168	70.9089
-10_1310320635.jpg	40	121	85.1909	93	85.0966	53	92.6956
-11_1310372829.jpg	41	224	52.6523	245	85.6991	36	106.6615
-11_1310384253.jpg	42	180	69.5071	102	88.3555	30	96.8556
-11_1310404336.jpg	43	72	52.5638	122	71.4955	223	108.4583
-11_1310416153.jpg	44	201	55.6805	151	60.6620	30	106.5824
-12_1310482743.jpg	45	143	76.0951	68	66.9289	13	99.7664
-19_1311065991.jpg	46	223	93.2323	212	93.5366	247	103.6485
-20_1311151990.jpg	47	17	85.0000	154	93.8263	35	69.8133
-23_1311430736.jpg	48	84	80.1456	79	55.3821	25	110.7296
-24_1311533727.jpg	49	110	83.0042	57	70.0717	30	89.4754
-26_1311678959.jpg	50	50	81.4000	157	80.0000	100	89.9643

# Transforming culture into data

*An exploration of the kind of knowledge that is produced when performing Cultural Analytics on Instagram database material*

ANKE  
HANS

Master thesis New Media & Digital Culture  
Faculty of Humanities, Utrecht University

Dr. Ann-Sophie Lehmann  
31 December 2011

## Index

0. Introduction	p. 3
0.1 Instagram	p. 3
0.2 Cultural Analytics	p. 4
0.3 Research question	p. 5
0.4 Methodological considerations	p. 5
0.5 Argumentative outline	p. 6
1. Instagram as a digital image database	p. 8
1.1 Add database material	p. 8
1.1.1 Users and their motivations for adding database material	p. 12
1.1.2 Storage of database material	p. 13
1.2 Browse database material	p. 14
1.2.1 Browse database material by the Instagram iPhone application	p. 14
1.2.2: Browse database material by third party developed webviewers	p. 15
1.2.3: Browse database material by web feeds	p. 16
1.2.4: Browse database material by external platforms	p. 17
1.3 Change database material	p. 18
1.3.1: Change database material by commenting	p. 18
1.3.2: Change database material by liking	p. 19
1.3.3 Stability vs. sociality of the Instagram database	p. 20
1.4 Destroy database material	p. 21
1.5 De Mul's database ontology refined	p. 24
2. Cultural Analytics as a digital visual research method	p. 25
2.1: An introduction to Cultural Analytics	p. 25
2.2: The making of Manga Style Space	p. 27
2.3 Describing Manga Style Space	p. 28
2.4 Knowledge claim based on Manga Style Space	p. 29
3. Performing Cultural Analytics on Instagram database material	p. 31
3.1: @ankehans Style Space 1	p. 31
3.1.1 The making of @ankehans Style Space 1	p. 32
3.1.2 Describing @ankehans Style Space 1	p. 34
3.1.3 Knowledge claims based on @ankehans Style Space 1	p. 34
3.2: @ankehans Style Space 2	p. 34
3.2.1 The making of @ankehans Style Space 2	p. 34
3.2.2 Describing @ankehans Style Space 2	p. 35
3.2.3 Knowledge claims based on @ankehans Style Space 2	p. 36
3.3: @ankehans Style Space 3	p. 36
3.3.1 The making of @ankehans Style Space 1	p. 36
3.3.2 Describing @ankehans Style Space 1	p. 37
3.3.3 Knowledge claims based on @ankehans Style Space 1	p. 38
3.4: Knowledge claims based on Instagram database material	p. 39
3.5: A future perspective	p. 39
4. Conclusion	p. 41
References	p. 43

## 0. Introduction

Digital databases come in all sorts and sizes. With the introduction of computers, databases were employed in libraries, museums, hospitals and companies to solve problems of information storage and retrieval. Selena Sol, adjunct professor of information and communication technologies at the Singapore Institute of Management, tells us that these digital databases were inflexible in their use. For example, in order to retrieve information from a database, users had to have an understanding of the way the data was structured (Sol 1998). Jos de Mul is a Dutch philosopher in the domain of culture and new media who also has written about digital databases. De Mul has observed that over the last few decades digital databases have become more flexible in their use, with the use of databases requiring less and less understanding of its architecture. This development has caused a rapidly growing range of database applications in all kinds of domains:

Database applications span virtually the entire range of computer software, ranging from mainframe databases for administrative purposes and multimedia encyclopedias on cd-roms to search engines, wiki's and other web 2.0 applications on the Internet (De Mul 2010: 6-7).

Examples of these flexible digital databases are Flickr that stores online images, YouTube that stores online moving image files, and Soundcloud that stores online audio files. These digital database applications are characterized by a high degree of usability. Their graphical interfaces are user-friendly to a degree that any user can become a co-designer of the digital database. Even when they have no designer experience, users can participate in building databases by adding database material without permission. Also, the database material that they add is free of restrictive copyright, which means that it is available to all database users, again without permission (De Mul 2010: 3-9).

Lev Manovich –a new media theorist and professor of Visual Arts at the University of California San Diego, USA who studies the impact of the digital revolution on the domain of visual culture– observed that a change in the flexibility of digital databases has led to a change in the material that they store. Material that any user can create, like photographs, videos, and audio tracks in the case of Flickr, YouTube, and Soundcloud respectively, is called 'user-generated media content' (Manovich 2008a: 204). Over the last few decades digitality has led to an "exponential rise in the numbers of cultural agents worldwide and the media they create" (Manovich 2009: 2).

In this thesis, I will focus on user-generated media content that is stored in these flexible digital databases, and visual content in particular. By 'digital database' is understood "a collection of heterogeneous items, on which the user can perform various operations – view, navigate, search" (Manovich 2001: 194).

### *0.1 Instagram*

As an exemplary digital database that stores user-generated visual media content I have chosen Instagram. On the surface, Instagram is a 6.8 megabyte photo-taking and photo-sharing application for iPhones running iOS 3.1.2 and above. The application was developed by San Francisco-based entrepreneurs Kevin Systrom and Mike Krieger who released it for free download from the iTunes App Store as of October 6, 2010. Instagram allows users to make photographs, apply stylistic filters, and post the result out to Instagram and other digital platforms such as Flickr, Twitter, and Facebook. With

following, liking, and commenting functions, Instagram also is a slimmed-down social network. The application is offered as "a fast, beautiful and fun way to share your life with friends through a series of pictures" (Instagram 2011a). In August 2011 MG Siegler, a columnist for technology weblog TechCrunch, reported that over the first nine months after its release the Instagram application had generated 150 million photo uploads. This is noteworthy, it having taken popular photo-hosting website Flickr nearly two years to reach 100 million photo uploads to their service (Siegler 2011). The Instagram application has grown very fast and contains a large body of digital imagery. In my opinion, this makes Instagram more than an iPhone application and a social network. Instagram also is a digital database that stores user-generated visual media content, and that is exemplary for the nature of current, flexible digital databases.

### *0.2 Cultural Analytics*

The ubiquity of user-generated visual media content as stored in digital databases like the Instagram database has raised some critical questions: do researchers in the field of visual culture have a responsibility to study this content? How can this content be studied? Manovich has thought about these questions that are raised by the vast amounts of online visual content. He was one of the first theorists to observe a gap between traditional visual research methods and the user-generated visual material that is being produced routinely online everyday: "the exponential growth of a number of both non-professional media producers in the 2000s has led to a fundamentally new cultural situation and a challenge to our normal ways of tracking and studying culture" (Manovich 2009: 2). On top of pointing out the discrepancy between user-generated visual media content and visual research methods, Manovich is also designing a method to study this content, notably through his role as the director of the Software Studies Initiative. The Software Studies Initiative was established in 2007 to disseminate the broad vision of 'software studies' through workshops, publications, and lectures (see [lab.softwarestudies.com](http://lab.softwarestudies.com)). Software studies is an emerging field for intellectual inquiry that aims at understanding how software shapes all areas of contemporary societies (Software Studies Initiative 2011). Working with Jeremy Douglass and William Huber, Manovich's colleagues from the Visual Arts department of the University of California San Diego, he is developing a new research paradigm in media theory called Cultural Analytics. By the act of conducting research in the paradigm of Cultural Analytics is understood "to quantitatively analyze the structure of cultural datasets and visualize the results revealing the patterns which lie below the unaided capacities of human perception and cognition" (Manovich 2009: 5). According to Manovich, borrowing digital image analysis and visualization techniques from computing science and applying them to visual cultural datasets allows researchers to formulate claims about the visual language of cultural phenomena (Manovich 2009: 2, Manovich 2010: 8). Cultural Analytics constitutes a new methodology in conducting visual research, being the first to approach cultural phenomena as quantifiable phenomena. Therefore Manovich considers it important for researchers who perform research in the paradigm of Cultural Analytics to be self-reflective:

We need to think about the consequences of thinking of culture as data and of computers as the analytical tools: what is left outside, what types of analysis and questions get privileged, and so on. This self-reflection should be part of any Cultural Analytics study (Manovich 2009: 10).

### 0.3 Research question

As a student in the field of new media and digital culture I have a curiosity for the epistemological consequences of a quantitative analysis of user-generated visual media content. Since my focus will be on the Instagram database, I centered my thesis around the following research question: what kind of knowledge is produced when performing Cultural Analytics on Instagram database material? The hypothesis is that performing Cultural Analytics on Instagram database material will allow researchers to formulate claims about the visual language of Instagram database material.

### 0.4 Methodological considerations

The first step in defining the kind of knowledge that is produced when performing Cultural Analytics on Instagram database material in a self-reflective manner has been to employ the Instagram iPhone application. Since academic resources on Instagram are scarce—the only work thus far is a master's thesis on Instagram user motivations by Zachary McCune, a young digital artist and scholar from New England—employing the Instagram iPhone application has provided me with a sense of the technological structure of the Instagram database. I consider it important to be aware of this structure when formulating knowledge claims on the basis of the computational analysis of its database material. Second, throughout the research process that preceded the writing of this thesis I have performed ethnographic research in the online Instagram community. By uploading images, following other Instagram users, and liking and commenting on their images has produced an understanding of Instagram user practices like 'tagging' and user priorities in using Instagram. Third, I have employed the software that is required to analyze and visualize digital imagery in the paradigm of Cultural Analytics: ImageJ and ImagePlot. Working with this software has enabled me to grasp the epistemological consequences of thinking of culture as data. Throughout this thesis I will refer to my experience in using the Instagram iPhone application, the ImageJ and ImagePlot software and the ethnographic research that I performed in the Instagram community.

My research on the Instagram database and Cultural Analytics is of an exploratory nature yet embedded into the analysis of current theoretical approaches to digital databases, cloud computing, digital image analysis, and datavisualization. Regarding digital image analysis and datavisualization I will frequently refer to the work of Lev Manovich that I discussed in 0.2. Second, the approach that is central to the analysis of Instagram as a digital image database is the work of De Mul on the database in the digital era. De Mul argues that on a fundamental level there are four characteristics that are an integral part of almost all computer software: *Add*, *Browse*, *Change*, and *Delete*. De Mul does not define these elements, yet he does mention that they equal the four basic operations of Structured Query Language, the standardized programming language with which users can manage data in modern-day digital databases. These four operations are *Insert*, *Select*, *Update* and *Delete*: adding new database material, retrieving existing database material, editing existing database material and removing existing database material respectively (Sol 1998). De Mul argues that each of these operations, when made in a database, constitutes a change in the structure of a database. Each operation causes the database to be combined, decombined, or recombined in a unique way. With these four operations in principle all possible combinations of records in a database can be created (De Mul 2009: 103, De Mul 2010: 6). This is why De Mul considers *Add*, *Browse*, *Change*, and *Delete* to be the ABCD of computing. Since software has become omnipresent in today's information society, De Mul argues that this ABCD of computing underlies our society and therefore can

be called a database ontology (De Mul 2009: 94-100). This database ontology is the model that I will employ in my analysis of Instagram as a digital database. This model will be modified when necessary. Adjacent subjects that I consider to be pertinent for the analysis of Instagram as a digital database are: its users, their motivations for creating and sharing database material, the storage of database material, and the social dimension of Instagram. I will treat these subjects throughout my analysis of Instagram as a digital database. Treating these subjects will provide us with information about the nature of Instagram database material. This information will have to be taken into account when formulating knowledge claims on the basis of Cultural Analytics research as has been conducted for this thesis. It will prevent us from formulating knowledge claims that do not directly result from the output of the Cultural Analytics research process. Third, I will refer to the work of Frank Kessler and Mirko Tobias Schäfer, two new media theorists at the Faculty of Humanities of Utrecht University. Kessler is a professor of Media History researching the period of the emergence of cinematography; Mirko Tobias Schäfer is an assistant professor at the Department for Media and Culture Studies and focuses on culture constitutional aspects of computer technology and software. In their article "Navigating YouTube: Constituting a Hybrid Information Management System" (2009), Kessler and Schäfer analyze YouTube's technological structure as a digital database that stores moving image files. They argue that the hybrid nature of the YouTube database determines the retrieval and the presentation of the database material (p. 288). Though YouTube and Instagram manage different types of content (YouTube concerns moving image files while Instagram concerns non-moving image files) and differ in the way their content is based (YouTube is web-based while Instagram is mobile-based), the analysis of YouTube as a hybrid database has been useful for this research. For on a higher level both YouTube and Instagram are concerned with digital collections of user-generated visual media content. Also, their technological structures that are of a hybrid nature are common.

In this work I will not discuss political aspects of online databases such as privacy, control, and ownership. Technical details will be discussed only in as far as they are relevant for the development of a clear understanding of the nature of Instagram database material.

### *0.5 Argumentative outline*

When making knowledge claims on the basis of performing Cultural Analytics on Instagram database material in a self-reflective manner, I consider it important to take into account the technological structure of the Instagram database and the nature of its material. Therefore the goal of the first chapter of this thesis is to understand Instagram as a digital database that stores user-generated visual media content. I will perform an in-depth analysis of Instagram by creating, uploading, retrieving and deleting an Instagram image. I argue that Instagram is a digital database that can largely be described by De Mul's database ontology. However, the user practices that have occurred around Instagram database material require a refinement of De Mul's database ontology.

In the second chapter I will focus on Cultural Analytics. The goal of this chapter will be to understand and apply Cultural Analytics as a research method. I provide a description of the use of the ImageJ and ImagePlot software that is employed when conducting research in this new research paradigm. I will describe the research process by discussing the creation of Manga Style Space, a large-scale visualization created by Lev Manovich. I will assess the knowledge claim that Manovich formulates on the basis of this visualization. I will argue that this claim does not display the self-

reflectivity that Manovich advises researchers to take on when performing research in the paradigm of Cultural Analytics.

Taking into account both the information about the nature of Instagram database material and my criticism about Manovich's knowledge claim regarding Manga Style Space, in the third chapter I will perform research in the paradigm of Cultural Analytics myself. Using ImageJ and ImagePlot, I have created three small-scale visualizations of Instagram imagery. On the basis of these visualizations I will formulate the knowledge claims that can be deduced from these visualizations, while taking into account the need for self-reflectivity. I will argue that, at this moment in time, Cultural Analytics is a digital visual research method that cannot produce relevant cultural insights when applying it to Instagram database material. For this moment it can best be regarded as a transfer of methods from other scientific fields to the field of culture.

## 1. Instagram as a digital image database

Within one week after the release of the Instagram iPhone application into the iTunes App Store it attracted 200,000 users. In the middle of October 2011 Instagram surpassed 10 million users (The New York Times Bits, 19 October 2010). What these 10 million people have in common is that they have an iPhone, they have downloaded the Instagram application, and have subscribed for an Instagram account. These are the three criteria for becoming an 'Instagramer', a user of Instagram. This title confers an entitlement to perform operations on the Instagram database. In this chapter the technological structure of this database is analyzed according to the four elements that constitute De Mul's database ontology: *Add*, *Browse*, *Change*, and *Delete*. I will also discuss the Instagramers, their motivations for using Instagram, the storage of the database material and Instagram's social dimension. Let's move down into Instagram's technological design by creating, uploading, retrieving, and deleting an Instagram image. I will start out with De Mul's first dimension of his database ontology, *Add*. How are Instagram images –also known as 'database material'– created and inserted into the database?

### 1.1 Add database material

Creating an Instagram image is possible only through the Instagram iPhone application. Opening the application returns a screen with five touch-sensitive buttons, which constitute the toolbar that structures the Instagram interface. To create input, I press 'Share', the middle of these five buttons (see Figure 1). This act returns a live-image preview of the iPhone camera. Now I can create input either by taking a picture with the built-in camera of the iPhone or by retrieving an image from my personal Photo Library (see Figure 2).

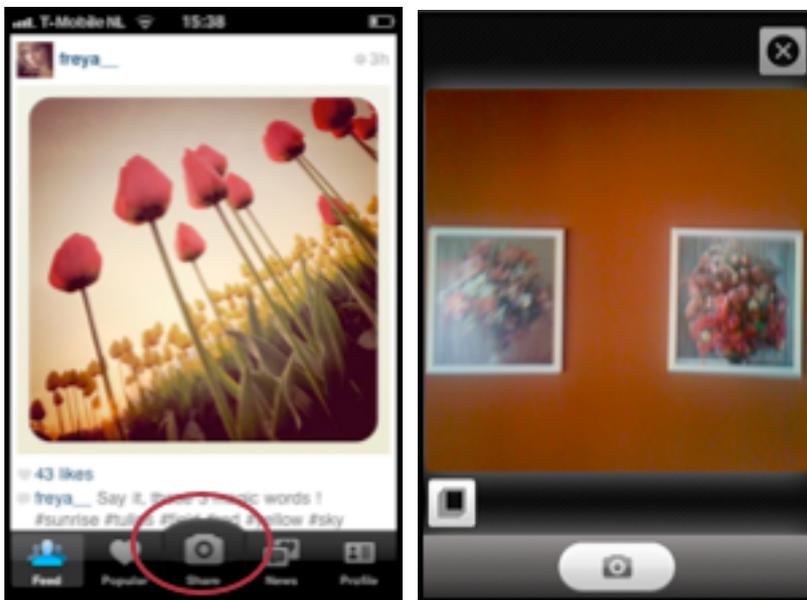


Figure 1 (left): Opening the application returns a screen with five touch-sensitive buttons of which the middle button, 'Share', is used to create database input.

Figure 2 (right): After clicking 'Share', a live-image preview of the iPhone camera appears that is targeted at two small paintings of flowers on my wall. Pressing the bottom left button would return me to my personal Photo Library.

I have chosen to create input by taking a picture of my wall with two small paintings of flowers. After having captured my wall, Instagram's software automatically reformats the image into a square shape that makes one think of the format of photos that are shot with Polaroid cameras. The software presents the result in a new screen and lets me choose one of fifteen filter options to layer over the image (see Figure 3). All of these filters change the mood and tonality of my image in a way that it gets a 'vintage' look: blurred edges, faded edges, added scratches, an enlargement of the contrast, the application of a vignette around the edges, added simulated film grain, added digital photo-paper, or Polaroid borders around the image (see Figure 4 on page 10), but I can also leave the photo unaltered. I have chosen the X-Pro II filter (see Figure 4: first row, second image).

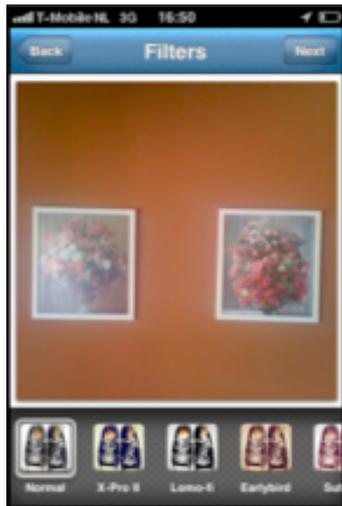


Figure 3: The Instagram software reformats the input I created into a square shape and presents filter options to layer over the image.

I click 'Next' in the upper righthand of the interface. This act returns a publish screen in which the application requests textual details about the image. Here I can add a title to the image, which in Instagram is called a caption. I can type any text that I assume to be an appropriate description for the image of my wall. Because making a caption is an operation performed before the image is inserted into the database, I call this operation an *a priori* option for adding information to Instagram database material. In a database context, this information about the image is called metadata. Professors of computer science at the University of Leuven Duval, Hodgins, Sutton et al. (2002) define metadata as "data providing information about one or more aspects of the data" (p. 1).

In practice the goal of adding a caption to an image is to inform other Instagramers about what is depicted on the picture and/or the circumstances that were involved in capturing it. I choose to type 'my wall'. Frequently authors of images also provide information about the location where the image was shot (i.e. 'pub', 'museum', 'London', 'holidays', 'Asia') and provide details about the photographic process, the device, and the filter that was used to create an image (i.e. 'bugseyeview', 'iphone3gs', 'Nashville').

A caption can consist of text but it also allows for the insertion of hashtags. Hashtags are keywords prefixed with the symbol '#' (identi.ca 2011). For example, in Figure 5 on page 11 (circle 'a') I have complemented the caption 'my wall' with a hashtag '#flowers'. This tag is an interactive textual marker that provides the Instagram

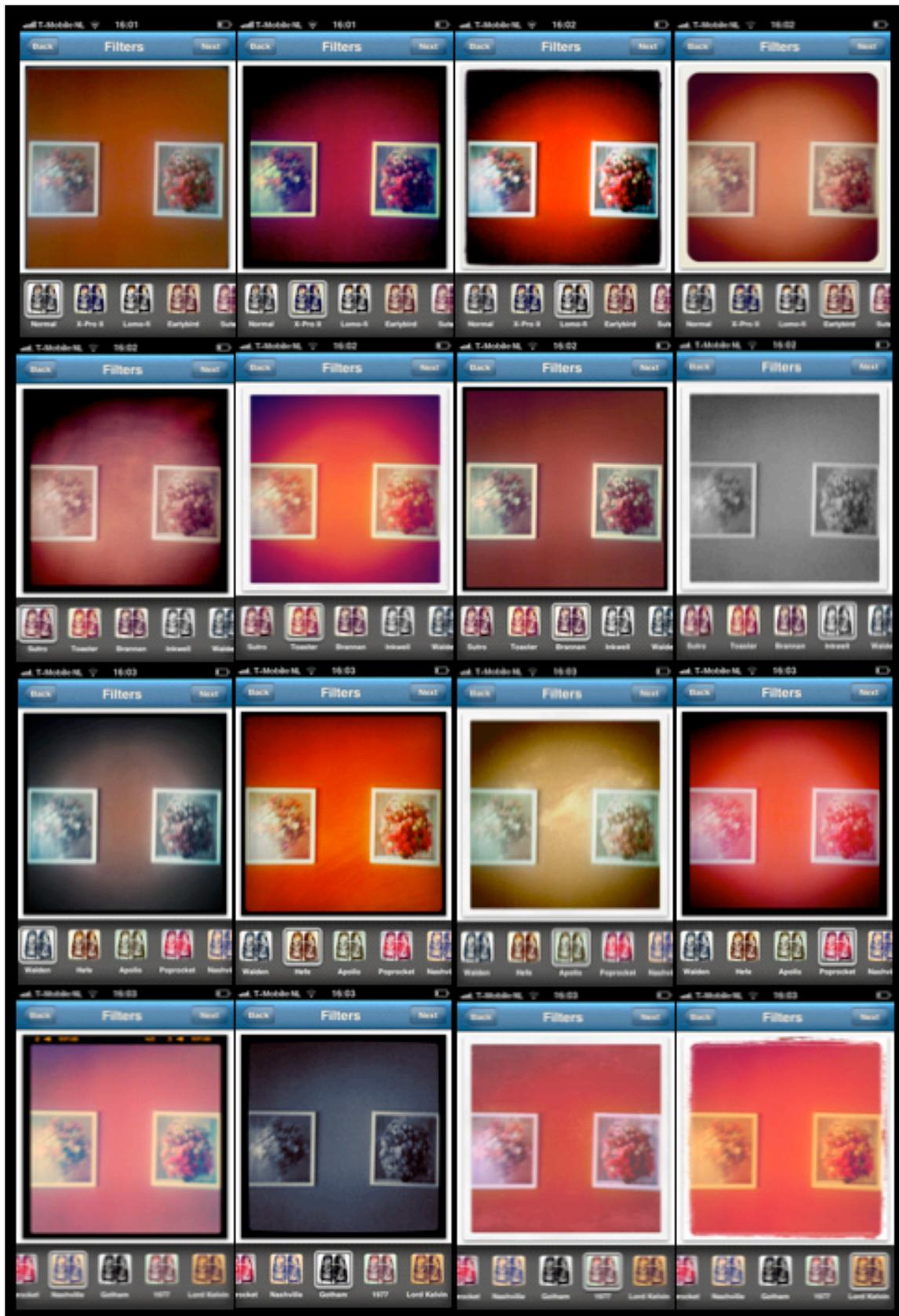


Figure 4: The image of my wall as I captured it with my iPhone camera and the image after modification by the fifteen Instagram filters. First row (from left to right): unaltered, 'X-Pro II', 'Lomo-fi' and 'Earlybird'. Second row: 'Subto', 'Toaster', 'Brannan' and 'Inkwell'. Third row: 'Walden', 'Hefe', 'Apollo' and 'Poprocket'. Fourth row: 'Nashville', 'Gotham', '1977' and 'Lord Kelvin'.

software with semantic information about my image. With this information, the software is allowed to connect my picture to other pictures with the same hashtag and thereby create a pool of images that is related to flowers (in reference to Kessler and Schäfer 2009: 279-280). Since hashtags also provide information about an image, they also are metadata. Although many Instagramers choose to provide their images with a caption and one or multiple hashtags, all possibilities to enter metadata within Instagram are optional.

Under 'Sharing' in the publish screen, I can now choose to publish my image to my profiles on external social platforms like Twitter and Facebook. I want to publish the image to my Twitter account, so I check the toggle box behind 'Twitter' (see Figure 5, circle 'b').



Figure 5: The publish screen. Circle 'a' shows the caption 'my wall', followed by the hashtag '#flowers'. Circle 'b' shows my choice of external publishing platform, Twitter.

By clicking 'Done', the image is being uploaded to my Twitter profile and to my personal Instagram feed. My Instagram feed is a vertical stream of images that the application returns when clicking 'Feed', the first touch-sensitive button of the application's interface (see Figure 6). This stream of images is clustered by the Instagram software based upon whether or not the image is created either by the users that I follow or by myself. 'Following' is a feature in many online social networks that allows users to subscribe to the content of others. For example, when an Instagramer has pressed the 'Follow' button in my user profile (see Figure 7) and I add an image to Instagram, this user will automatically see this image in his personal Instagram feed. Adding an image to the Instagram database thus equals the sharing of this image with my followers.



Figure 6: Viewing my image in my personal Instagram feed.



Figure 7: The follow-button in my user profile, which allows users to subscribe to my Instagram images.

Before moving on to De Mul's second element of the database ontology, I will describe the Instagramers, their motivations for creating and sharing their imagery with others, and the storage of the Instagram database material.

### 1.1.1 Users and their motivations for adding database material

In order to get an impression of the users that create the Instagram database material, I turned to Philippe Gonzalez. Gonzalez is the founder of a global Instagram community called 'Instagramers.com' (see [instagramers.com](http://instagramers.com)). This community is an initiative independent of the Instagram application as developed by Systrom and Krieger. In an e-mail exchange, Gonzalez provided me with the following general profile of an average Instagramer: he or she is aged between 27 and 50 years old, works in communication, advertising, the creative industry, or data systems and is of medium to high socio-economic class. In short they are typical 'early adopters' (Gonzalez 2011). According to Everett Rogers, who described early adopters in his book *Diffusions of Innovations* (1962), these are individuals within a social system who embrace an innovation faster than most other people in their social system do (p. 283).

For his master's dissertation Zachary McCune researched what drives these early adopters to create and share imagery on Instagram. He carried out ethnographic research, attended an Instagram meetup in London and analyzed user surveys that questioned the motives behind both the creation of images and the sharing of images. From his research McCune found that users are not drawn to the Instagram network by one pre-eminent motivation. Every survey response acknowledged that there is a minimum of three of the following motivations that underlie creating and sharing visual media content on Instagram: documentation, seeing, community, creativity, therapy and sharing. Also, these motivations are often blurred to some extent, which suggests their multi-dimensional nature (McCune 2011a: 72-75). User motivation is a thesis in and of itself, so I will restrict myself to the description of the most common motivation for creating and the most common motivation for sharing visual media content on Instagram. These two motivations, documentation and immediate cultural learning, form the foundation of motivations for Instagram use (ibid., p. 61).

The main motivation for Instagramers to create imagery is documentation. Respondents expressed an urge to capture, record, and preserve memories, experiences,

and moments in their lives. Examples of answers from which McCune drew are: "the world is constantly changing and so I am documenting it" and to "...capture an interesting or emotional moment, before it vanishes". From answers like these, McCune concluded that creating Instagram images is an activity that is employed as "an aid to memory and to posterity". This highlights a perception that photography can "arrest and objectify a world that changes" (ibid., p. 62), a motivation that does not set Instagram photography apart from photography in general.

As explained in 1.1, uploading an image to the Instagram database equals sharing it with ones followers: Instagram imagery is publicly visible by default. It is possible to keep ones images private which forces users who want to follow another user to request their authorization. However, the majority of the Instagramers does not switch this toggle and thus chooses to share their visual content with other Instagramers. The motivation behind sharing photos after creating them does reveal a specificity that sets Instagram aside from photography in general. McCune found that Instagramers share their images because they are interested in immediate cultural learning. There are two ideals that constitute this motivation. The first ideal is of an individualist nature: users want to present a personal worldview to others and find pleasure in allowing photographs to become objects of value to people beyond themselves (ibid., p. 59). The second ideal is the social exchange of vision between different users. Exemplary questionnaire answers that support this ideal are "the ability to see through photographs of people around the world what is going on in their lives" and "people can see through my eyes, and vice versa" (ibid., p. 64). Users share their work with other Instagramers in order to exchange vision and enrich their cultural experience. McCune even argues that the value of an Instagram image lies in the fact that it is shared with other Instagramers: "in fact, this becomes the essential value of the Instagram image, that it is shared, rather than kept and prized" (ibid., p. 60).

### *1.1.2 Storage of database material*

As discussed in 1.1.1, Instagramers create images from a documenting perspective. McCune (2011) noticed that this implies a trust in the Instagram service as an archive in which their images are stored in the long term (p. 62). The question is if their images actually are stored long term and where the storage places are located. In order to answer these questions we need to make a move from the database into the digital cloud. All image files that are being uploaded to Instagram are automatically stored in an online image collection on servers that are supported by Amazon's Elastic Compute Cloud webservice. Renting out the servers that Amazon does not use regularly, they provide Instagram with instantly resizable computing capacity that allows Instagram to meet spikes in traffic or demand (Amazon 2011). This service delivered by Amazon is part of a larger technological development called cloud computing. George Gilder, senior fellow at the Discovery Institute, wrote about cloud computing in the October 2006 issue of *Wired Magazine* as an architectural shift in online network distribution. Whereas a decade ago desktops of separate computers were the main data centers, data is now resident on interconnected servers 'somewhere on the Internet'. Cloud computing constitutes a shift from local computing power to an Internet cloud of shared networks made up of online machines and servers (Gilder 2006). Pelle Snickars, Head of Research at the Swedish National Archive of Recorded Sound and Moving Images, argues that this shift changes our understanding of the binary categories 'computer' and 'Web'. Google Docs for example, allows users to create texts in online documents as opposed to Microsoft Word documents that are stored on a user's desktop. Large-scale examples of cloud computing

companies are YouTube, Flickr, Myspace, and Facebook, who have all built their programs and content *on* the Internet rather than *for* the Internet.

Is the trust that Instagramers have in Instagram as an archive justly attributed? Reading Instagram's Terms of Use, the answer seems to be negative:

... Instagram encourages you to maintain your own backup of your Content. In other words, Instagram is not a backup service. Instagram will not be liable to you for any modification, suspension, or discontinuation of the Instagram Services, or the loss of any Content (Instagram 2011c).

From this we can conclude that there is a possible discrepancy between the archival function that Instagramers ascribe to Instagram and the actual service that Instagram guarantees its users. At the moment, the most reliable option for turning Instagram into an archive is to use a service called 'Instaport'. This service exports all Instagram images of a user's account in a zip format to the user's hard drive, the use of which would imply a provisional return to local computing power (Instaport 2011, McCune 2011b).

There thus is a difference between a traditional archive and a digital database like Instagram. Instagram cannot fulfill a traditional archival function to its users. Due to this uncertainty in relation to the physical storage of Instagram database material, it can therefore be considered an unstable digital database.

## *1.2 Browse database material*

After having added an image to the Instagram database, it can be retrieved in multiple ways: by using the iPhone application, by using third party developed webviewers, and by using web feeds. Also, as mentioned in 1.1, the Instagram application allows users to publish their images to user accounts at external social platforms. These platforms are microblogging service Twitter, social networking service Facebook, blogging platforms Tumblr and Posterous, image hosting website Flickr, and mobile location-based social networking service Foursquare. This means that these platforms also constitute a way to browse the Instagram database. In this section I will address all four options to browse the Instagram database and retrieve the image of my wall.

### *1.2.1 Browse database material by the Instagram iPhone application*

The first option to retrieve my image from the Instagram database is by using the iPhone application. I click 'Feed' that, as described in 1.1, returns a personal stream of database material created by either users that I follow or myself. In this feed I can temporarily find the image of my wall, until it becomes filled with more recent content. A second way of retrieving my image through the application is to formulate a search operation deeper down in the application's menu. Clicking the bottom right button of the toolbar called 'Profile' returns an option for searching the Instagram database by hashtag. Since I tagged my image with the hashtag '#flowers', entering the term 'flowers' in the application's search engine returns an image feed with images that all have been tagged with the hashtag '#flowers' (see Figure 8).

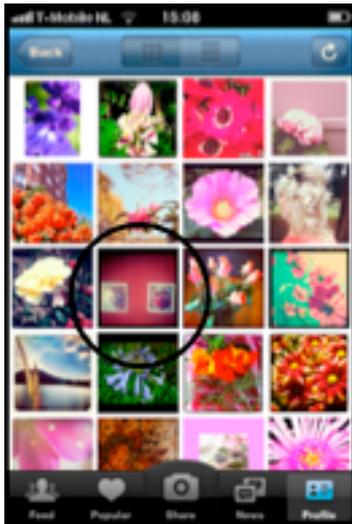


Figure 8: Entering the term 'flowers' into the application's search field returns a screen with a feed of all the images that have been tagged '#flowers'.

### 1.2.2: Browse database material by third party developed webviewers

Originally the Instagram database could be browsed by iPhone application exclusively. Instagram barely had any web presence, until its creators introduced an official API to the public in late February 2011. 'API' is an abbreviation of 'Application Programming Interface' and can best be seen as a direct gate for developers who are interested in Instagram database material. Without the use of a graphical interface –i.e. through programming commands and methods defined by Instagram– an API allows developers to interact with the Instagram database and stream the material that they request into their self-created software applications. This open Instagram database connection has resulted in a wide range of derivative services that all capitalize on the existing popularity of the Instagram iPhone application (McCune 2011a: 30-31). Some developers have created services that print Instagram images. Postagram for example, prints Instagram images on postcards (see [postagramapp.com](http://postagramapp.com)), and StickyGram prints them on magnets (see [stickygram.com](http://stickygram.com)). Another example of a derivative service is Instaport, the service I discussed in 1.1.2 that allows users to export their imagery to their local hard drive. Since these services are developed by parties outside Instagram they are called 'third party developed software': software developed by others which makes use of Instagram database material (in reference to Kessler and Schäfer 2009: 280, McCune 2011a: 30-31). There also have been multiple developers –mostly freelancers involved in web engineering– who have employed the API connection to solve their personal dissatisfaction with the limited image viewing options caused by the small screen of the iPhone. They have created 'webviewers': software that allows users to browse the Instagram database on a computer screen. Examples of these services are Statigram (see [statigr.am](http://statigr.am)), Extragram (see [extragr.am](http://extragr.am)), and Webstagram (see [web.stagram.com](http://web.stagram.com)). The latter service I will use to retrieve the image of my wall. I enter the search term 'flowers' into Webstagram's search field and the software returns an on screen image feed with all the Instagram images that have been equipped with the '#flowers' hashtag (see Figure 9).

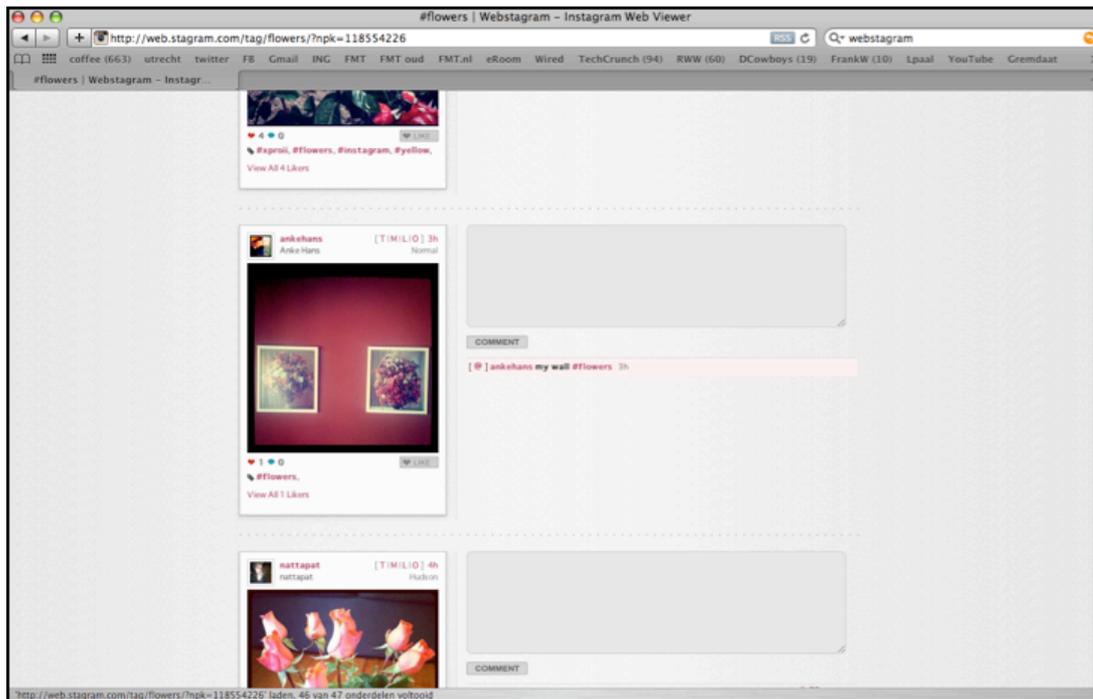


Figure 9: The image feed tagged 'flowers' on Webstagram, a third party developed webviewer that allows users to browse the Instagram database on their computer screens.

Browsing these webviewers is an operation that can be performed independently from using the original iPhone application. Yet for their input the webviewers are dependent upon the API connection that stands between the webviewer and the Instagram database. Therefore McCune considers the third party developed webviewers as distinct parts of Instagram that find themselves on the periphery of what might be called its ecosystem (McCune 2011b).

### 1.2.3: Browse database material by web feeds

A third way to browse the Instagram database is through web feeds: data formats that provide users with news updates from a website in a simple form for a computer. There are different types of web feeds; here I will discuss the most popular web feed called RSS (Really Simple Syndication) feed. An RSS feed is a web format used to publish frequently updated work such as blog entries, news headlines, or Instagram imagery. As with webviewers, RSS feeds enable users to generate a vertical list of images associated with a particular Instagram hashtag. I retrieve the image of my wall by creating a real-time stream of photos that are tagged 'flowers' by entering 'http://instagr.am/tags/flowers/feed/recent.rss' into my URL field (see Figure 10). What distinguishes an Instagram RSS feed from the image feed generated by Webstagram is that through RSS feeds I request a real-time subscription to a particular stream of images. This enables me to see live photos as they are being added to the Instagram database (Instagram 2011b). As the possibility to create RSS feeds has been released by the official Instagram service, I consider RSS feeds to be an integrated part of the Instagram ecosystem.

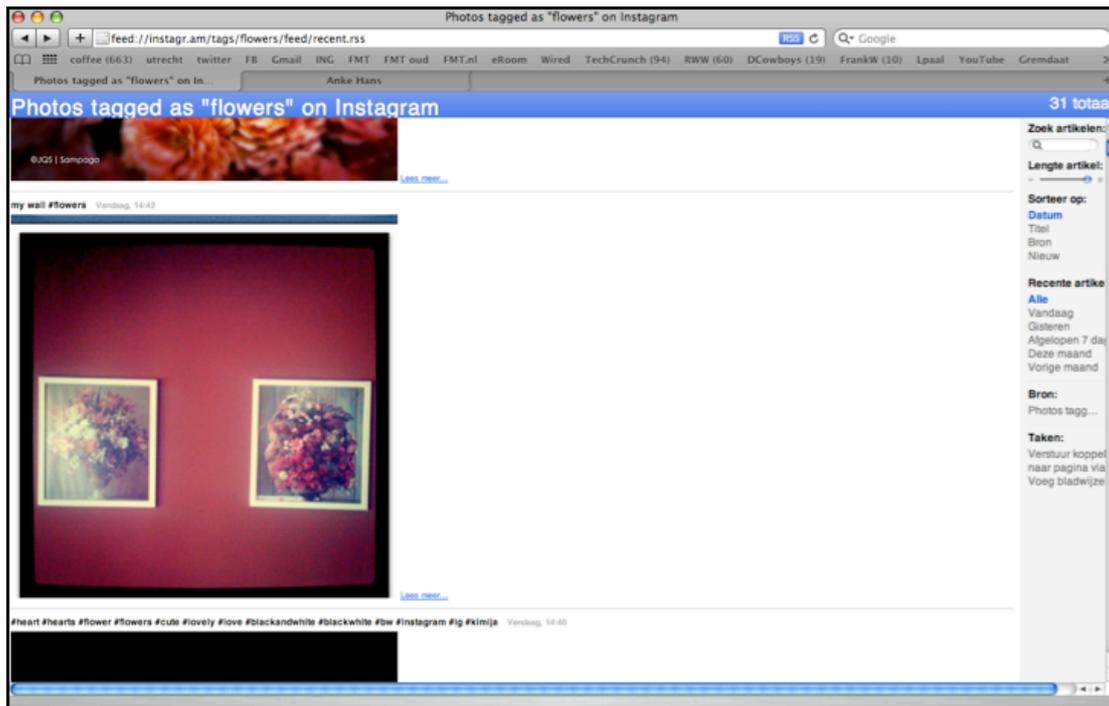


Figure 10: A real-time RSS feed created to browse the Instagram database for images tagged 'flowers'.

#### 1.2.4: Browse database material by external platforms

The fourth option for browsing the Instagram database is through the external social platforms that are integrated with but not related to the Instagram iPhone application. After users have indicated that they wish to publish their image to their accounts at either Twitter, Facebook, Tumblr, Posterous, Flickr and/or Foursquare, a URL to the image concerned will appear on their profile. In 1.1 I mentioned that I have opted for my image to be posted to my Twitter account. Checking my tweets indeed returns in a tweet with the caption 'my wall #flowers' followed by a URL (see Figure 11). Clicking the link returns my image in a new screen in an Instagram template (see Figure 12).



Figure 11 (left): The link to my image as it appears in my Twitter timeline.



Figure 12 (right): Opening the image from the link in my Twitter timeline.

These external platforms can be used for publishing images outside the Instagram service but they do not offer the possibility to view other images that also are tagged 'flowers'.

Publishing Instagram images on an external social platform thus means that there are fewer options for the retrieval of images that are related to the image of my wall. Therefore the external platforms can be positioned as integrated parts of the Instagram ecosystem, which are more distinct than the webviewers (McCune 2011b).

Having discussed the options for browsing the Instagram database, we can conclude its scope goes beyond the original iPhone application. There are different entrances into the Instagram ecosystem that allow for different levels of browsing activities.

### 1.3: Change database material

After having retrieved an image by either of the four browsing options, users have two options at hand to edit its position in the database. They can either place a comment by pressing the 'comment' button or 'like' an image by pressing the 'like' button. Both buttons are placed underneath every Instagram image (see Figure 13, circle 'a'). In practice, Instagramers perform commenting and liking operations to provide the author of an image with feedback about his/her work. At the same time however, these operations modify the partition of the images in the Instagram database. For example, comments on an image often lead to the incorporation of this image into multiple image feeds within the database.

#### 1.3.1: Change database material by commenting

A comment is a written reaction that users can place underneath an image. Contentwise these reactions can be diverse. In general they are compliments or technical questions to the author of an image. These comments may also include hashtags that the user considers to be appropriate labels for the image. Placing a comment under an image is an act that can be performed by all Instagramers including the author of an image. In Figure 13, circle 'b' for example, I have placed a comment in the form of a second hashtag, 'paintings', underneath the image of my wall.



Figure 13: The image of my wall. Circle 'a' shows the 'like' and 'comment' button under my Instagram image. Circle 'b' shows a self-placed comment consisting of a 'paintings' hashtag under the caption of my image.

In 1.1 I have explained how hashtags in a caption serve as metadata that is created before an image is uploaded to the Instagram database. Therefore I called the creation of a caption an *a priori* option for adding metadata to Instagram database material. Hashtags in comments also serve as metadata, which allows the Instagram software to connect my image to other images with the same hashtag. The result is an incorporation of my image into two image feeds: one image feed called 'flowers' and one image feed called 'paintings'. With hashtags in comments the Instagram software is thus allowed to 'duplicate' an image into another image feed. Adding a hashtag in a comment thus causes a change in the distribution of the image through the database. In reference to both Kessler and Schäfer (2009) and De Mul (2010) I argue that creating a comment including one or multiple hashtags is a second option for Instagramers to provide the software with metadata that recombines the Instagram database material in a unique way (p. 284 and p. 6 respectively). Since metadata in comments is added to an image after it has been uploaded I call it an *a posteriori* option of adding metadata to Instagram database material.

When it comes to creating hashtags Instagram does not put any limits to the creativity of its users. Instagramers can literally create any hashtag they desire, both in captions and in comments. I have already discussed the traditional function of tagging in 1.2: the tags 'flowers' and 'paintings' that I attached to the image of my wall describe what is depicted on the image. However, the liberty of tagging Instagram imagery has resulted in additional functions of this database operation. First of all, users tag images to notify one or more Instagramers. For example, tagging an image with '#longram', an Instagram-centered community in London, does not ensure that the image has anything to do with London. Instead a user wants the Longram community to see the image (McCune 2011a: 50). During my integration in the Instagram community I found that another function of tagging is to set up meetings among Instagramers, witness the following caption by user @r3kahsttub: "I'll be headed to #Bangkok this weekend for #holiday; any Igers over there want to meet up for a #coffee? I'll be at a #Starbucks this Sunday morning". Third, tagging has incited the organisation of virtual competitions. Tagging an image with '#jjchallenge' for example, implies ones participation in a challenge among Instagramers for the best photo using the Lomo-filter organized by user @joshjohnson (Instagram 2011b). Last, hashtags can be employed in order to create a personal subdatabase of ones favorite images, i.e. '#ankesfaves'.

### 1.3.2: Change database material by liking

Besides commenting, the second operation that Instagramers have at hand to edit the position of Instagram images is to 'like' an image. This operation is performed by pressing the 'like' button beneath an Instagram image. As with commenting, liking simultaneously provides the author of the image with feedback about his/her image. Figure 14 for example shows that the image of my wall has been liked by users @jmgoebelt and @spems.



Figure 14: The image of my wall has been liked by users @jngoebelt and @spems.

The number of likes that my image collects are not only noticed by me as the author of the image, they also are read and stored by Instagram's 'Popular Algorithm'. This algorithm is a part of the Instagram software that was written to trace the images that receive the most likes within five minutes after their upload. On the basis of this ranking the software structures its 'Popular' feed. This feed is a 4x8 photo grid that is returned after pressing the second touch-sensitive button in the application's toolbar. The grid displays the 32 most popular images of the moment. Because the act of liking results in information that is processed by Instagram software to structure (a part of) the database, I consider liking the second *a posteriori* option for adding metadata to Instagram database material.

### 1.3.3 Stability vs. sociality of the Instagram database

I have discussed three operations that Instagramers can employ to add metadata to Instagram database material: writing a caption with hashtags, writing a comment with hashtags, and liking. However, as mentioned in 1.1, none of these three operations are obligatory. This means that if none of these options are employed, the image rests 'bare' and the modes of retrieval are restricted. The Instagram software will not be able to recognize the image and place it into streams of images that are tagged with identical terms. The image will only temporarily show up in the personal image feeds of its author and the author's followers, but is bound to largely disappear after the feeds have filled up with more recent content.

The assignment of metadata to Instagram database material is a practice that comes across as categorization in a rather random and messy way. First, not all the Instagram images necessarily are equipped with metadata in the first place. Second, as discussed in 1.3.1, metadata can be assigned by different Instagramers and can serve multiple functions, which implies that the metadata itself also is unstructured. Kessler and Schäfer (2009) observed similar categorization practices in their analysis of YouTube. They found that YouTube is characterized by a "... rather incoherent and unorganized way in which videos are indexed, labeled, and filed" (p. 286). Would a catalog be made, "it would not contain similar information for each video" (p. 282). Due to this characteristic, Kessler and Schäfer consider YouTube to be a database that is relatively unstable. Its only

stability is generated through the random addition of metadata over time (Kessler and Schäfer 2009: 286). We can conclude that besides the instability in relation to the physical storage of Instagram database material as discussed in 1.1.3 there also is an instability in relation to the assignment of metadata to the Instagram database material.

One might wonder what attracts so many users to participate in building a database that can neither ensure that their material is stored nor that their personal material can be retrieved. As described in 1.1.2 Instagramers create images because it serves as life documentation. Does this not require a guaranteed archival function? In practice this does not appear to be the case. Whereas the usefulness of a library is proportional to its data storage and retrieval efficiency (Sol 1998), on Instagram the assignment of traditional metadata to database material is not what drives Instagram to a success. During my integration in the Instagram community I have found that Instagramers do not regard their personal image collection to be a collection that should archive their past in an objective manner. Rather, they consider their collection to be a living collection that first of all is about a shared present that is open to inquiry and feedback by other Instagramers. McCune (2011a) confirms that the use of metadata exceeds the objective categorization of Instagram imagery: "tagging operations extend and expand the social potential of the image, which becomes increasingly social with more comments and likes" (p. 56): "the comments and likes that are often collected rapidly, can be considered instant feedback loops, which amplify the already accelerated digital photographic process" (ibid., p. 66-67). This social function is emphasized by the 'News' button, the fourth touch-sensitive button of the application's toolbar. 'News' informs me who has liked my images (see Figure 15, circle 'a') and who has commented on my images (see Figure 15, circle 'b'). It is not the traditional archival function but Instagram's social features that tie users to other users to form a worldwide Instagram community (McCune 2011b). Rather than facilitating Instagram as an archive, the assignment of metadata facilitates Instagram as a community.

Having discussed how users can add, browse and change the position of Instagram database material, we can confirm that Instagram is a digital database in which users are co-designers: they generate its material (1.1), they can browse the database to retrieve material (1.2), and they determine its position in the database by providing it with metadata, before uploading by adding a caption (1.1) and/or after uploading by commenting and liking (1.3). Let's take a look at the extent to which users have a voice in the fourth element of De Mul's database ontology: the destruction of database material.

#### *1.4: Destroy database material*

There are three ways to delete an image from the Instagram database. All three options delete the image concerned as well as the metadata that may have been attached to it. First, I can delete my own image by pressing the button that is placed underneath my image (see Figure 16) and consequently press the 'delete' button, the first of the four photo options that is offered (Figure 17).

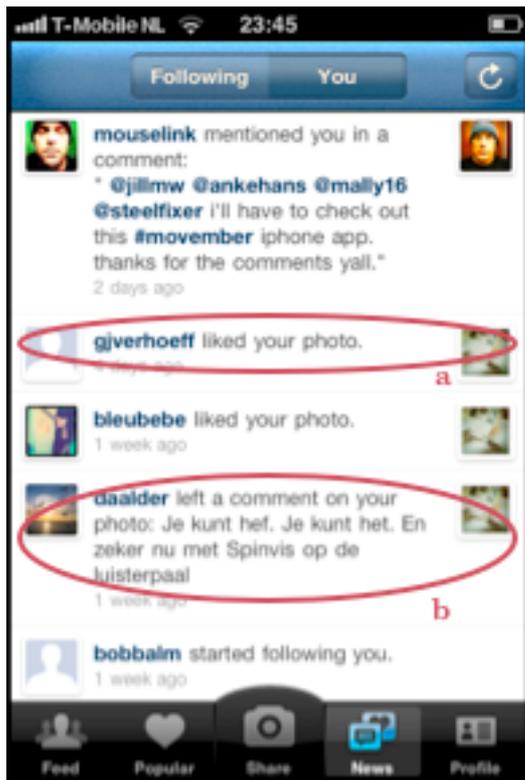


Figure 15: The News section in the Instagram application. Circle 'a' shows that user @giverhoeff has liked one of my images. Circle 'b' shows that user @daalder has commented on one of my images.



Figure 16 (left): The button that returns a menu with Instagram's option for the deletion of database material.

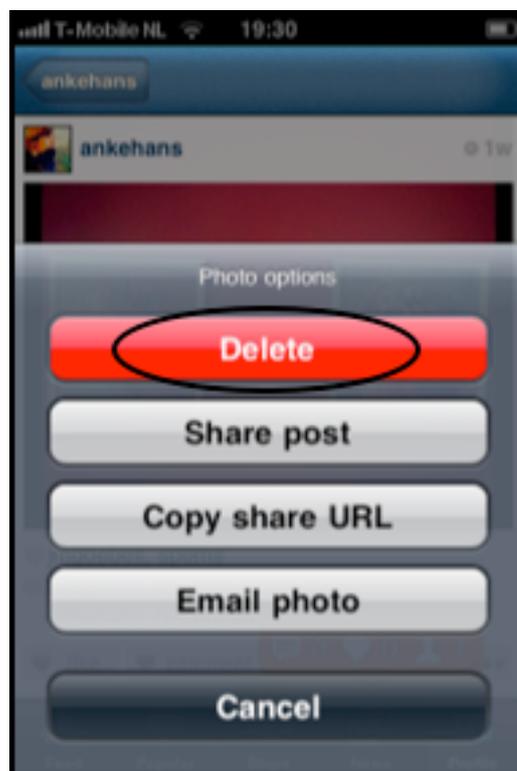


Figure 17 (right): The 'delete' button with which I can delete my own image.

Second, I can participate in the deletion of another user's image. For example, if I would find user @playboy's image of their Miss June 1986 objectionable because it depicts a nude woman, I can press the button placed underneath this image (again circled in Figure 18) that returns one photo option –the number of photo options are reduced because I am not the author of this image– called 'Flag for review' (see Figure 19). Flagging means reporting an image for objectionable content such as nudity, copyright infringement, abuse, or other violations of the lines of conduct as formulated in Instagram's Terms of Use. I press the 'Flag for review' button and am asked to select the proper reason for reporting this image: nudity, copyright or terms of service (Figure 20). I choose 'Nudity' and get a confirmation that the image is flagged for review (Figure 21).

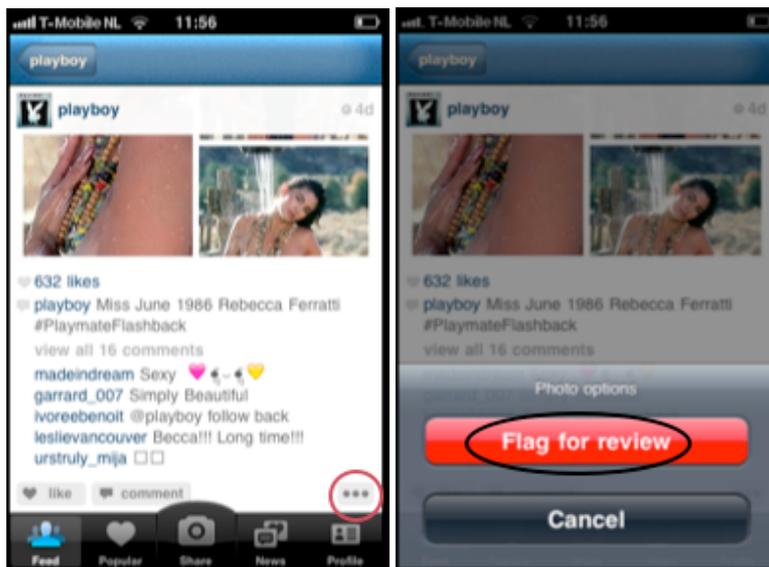


Figure 18 (left): The button that returns a menu with Instagram's option for flagging other users' images.  
 Figure 19 (right): The 'flag' button with which I can take part in deleting user @playboy's database material.

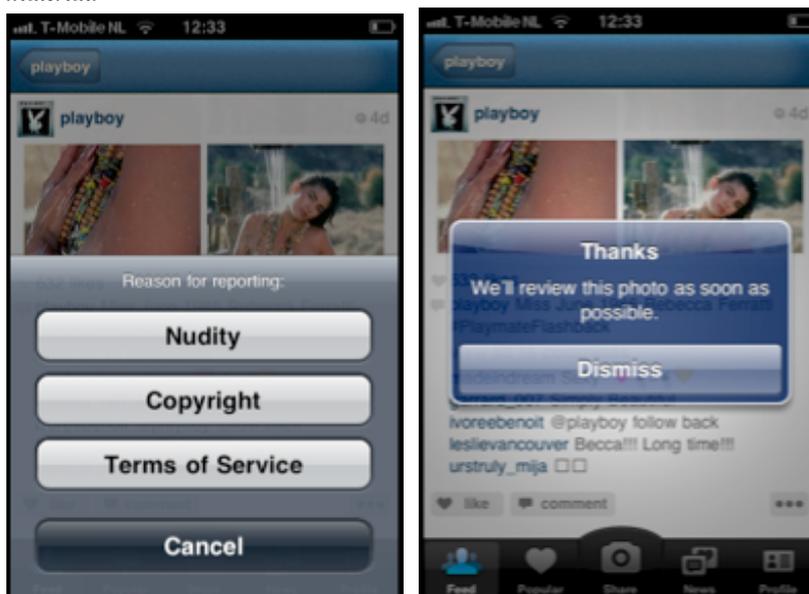


Figure 20 (left): The menu from which I have to select the reason that I am reporting this image.  
 Figure 21 (right): The message that is returned after flagging user @playboy's image for nudity, which confirms that this image is now flagged for review.

However, this image will not be reviewed by an Instagram employee. Once multiple users will also have flagged this image, it will automatically be removed from the database by the Instagram software –a much debated part of Instagram. Instagram is aware of the responsibility it ascribes to its users when it comes to judging the content of its database material: "On removed photos and accounts, we (Instagram Team) rely heavily on the Instagram community to help keep the photos on Instagram within our App Store guidelines and Terms of Use" (Instagram 2011c). However, at the same time Instagram makes it clear that they retain final authority over the images in its database:

We may, but have no obligation to, remove Content and accounts containing Content that we determine in our sole discretion are unlawful, offensive, threatening, libelous, defamatory, obscene or otherwise objectionable or violates any party's intellectual property or these Terms of Use. [...] Although Instagram will normally only delete Content that violates this Agreement, Instagram reserves the right to delete any Content for any reason, without prior notice. Deleted content may be stored by Instagram in order to comply with certain legal obligations and is not retrievable without a valid court order. [...] We reserve the right to modify or terminate the Instagram service for any reason, without notice at any time (Instagram 2011c).

Although users manage the Instagram database to a large extent, when investigating the destruction of Instagram database material, it becomes clear that the final rights to all the database material lie with Instagram. In reference to Kessler and Schäfer (2009), I argue that this uncertainty as to whether or not the database material will remain available on Instagram increases the instability regarding the physical storage of Instagram database material as discussed in 1.1.3 (p. 276).

### *1.5 De Mul's database ontology refined*

In this chapter I have analyzed Instagram as a digital image database according to De Mul's database ontology. This model –*Add, Browse, Change, and Delete*– is adequate to describe the Instagram database. We have found that users, who generally can be described as 'early adopters', add material to the Instagram database that is simultaneously shared with their followers. Before uploading material, the author of an image can add *a priori* metadata in the form of a caption and one or more hashtags. These hashtags, among other things, make an image retrievable when browsing the database by the iPhone application itself or in third party developed webviewers, web feeds, and external social platforms. Instagramers can change the position of images by 'commenting' or 'liking' them, two options for adding *a posteriori* metadata to Instagram imagery. The main motive behind the creation of the imagery is life documentation and the main motive behind the sharing of the imagery is an interest in cultural exchange through images. Due to an uncertainty regarding the physical storage of the material, the database can be called unstable. A second type of instability is caused by the inconsistent assignment of metadata that users attach to Instagram database material: metadata is not necessarily attached and hashtags do not necessarily represent what is depicted on an image. This is explained by the fact that tagging on Instagram often serves a social rather than an archival function.

However, these user practices of tagging database material require two refinements of De Mul's database ontology. The first refinement concerns a division between 'imagery' and 'metadata' as two elements that can be added to the Instagram database. This calls for the second refinement that is caused by the different moments in time on

which Instagramers can add metadata to an image: tagging in captions as discussed in 1.1 is assigned before uploading database material (*a priori*) and tagging in comments and liking, as discussed in 1.3, is assigned after uploading database material (*a posteriori*). Whereas adding imagery falls within the scope of *Add*, the addition of metadata can fall under either the scope of *Add (a priori)* or *Change (a posteriori)*.

## 2. Cultural Analytics as a digital visual research method

Having analyzed Instagram as a digital database that stores user-generated visual media content brings us to the question how to study this content. In the introduction of this thesis I discussed the work of Lev Manovich who co-develops Cultural Analytics, a new approach for the study of visual cultural content. In this chapter I describe Cultural Analytics as a digital visual research method. After a short introduction to this method I will make explicit how to conduct research in the Cultural Analytics paradigm by discussing the creation of Manga Style Space, a large-scale visualization of Japanese manga imagery created by Lev Manovich. Then I will assess the knowledge claim that Manovich formulated on the basis of this visualization, whilst being aware of the need for self-reflectivity in any Cultural Analytics study.

### *2.1: An introduction to Cultural Analytics*

At the Software Studies Initiative Manovich co-develops Cultural Analytics. The breeding grounds of Cultural Analytics are large-scale quantitative data analysis and information visualization methods that have already been employed in businesses, governments, and life sciences for five decades. For example, engineering scientists Vijayakumar and Gharpure (2011) discuss the way that medical imaging software is employed in the biomedical community to automatically segment and measure brain tumors in MR images (p. 2). Also, rheumatologists Dacre and Huskisson (1989) describe how edge detection algorithms in digital image analysis software can be employed to measure joint space in the knee size using digital image analysis (p. 506). Cultural Analytics aims to incorporate these existing quantitative image-processing methods and apply them to the domain of culture. The software and the approach to data remain the same, yet the datasets that are analyzed are different: they are contemporary cultural datasets (Manovich 2009: 2, Manovich 2010: 8):

We feel that the ground has been set to start thinking of culture as data (including media content and people's creative and social activities around this content) that can be mined and visualized. In other words, if data analysis, data mining, and visualization have been adopted by scientists, businesses, and government agencies as a new way to generate knowledge, let us apply the same approach to understanding culture (Manovich 2007: 3).

How to treat culture as data? The answer lies in the way Manovich defines the act of conducting research in the Cultural Analytics paradigm: "to quantitatively analyze the structure of cultural datasets and visualize the results revealing the patterns which lie below the unaided capacities of human perception and cognition" (Manovich 2009: 5). Manovich has already performed large-scale research in this new research paradigm in order to explore the visual language of different types of visual media content. His projects include the analysis and visualization of 4535 covers of Time Magazine dating from 1923 to 2009, 780 paintings by Vincent van Gogh, and one million manga pages (Manovich 2011: 8). However, the datasets that are the starting points of a Cultural Analytics research process do not necessarily have to be large. As Manovich observed: "... even with small cultural data sets consisting from hundreds, dozens or even only a few objects it is already viable to do Cultural Analytics" (Manovich 2009: 5). To make clear how to perform research in the paradigm of Cultural Analytics, I will now describe a visualization project called Manga Style Space, as executed by Lev Manovich and his colleagues Jeremy Douglass and William Huber. Out of practical considerations I will refer to these three researchers with 'Manovich'.

## 2.2: *The making of Manga Style Space*

Manga Style Space has been a large-scale visualization project centered around scanlated editions of Japanese manga. Scanlation –a portmanteau of 'scan' and 'translation'– denotes the practice exercised by Japanese manga fan communities who translate manga pages from Japanese to English, scan them and distribute them among an online public (Douglass, Huber and Manovich 2011: 12-14, 32). The first step in this project has been to collect input: a collection of 1,074,790 unique pages of scanlated Japanese manga that originate from 883 different manga series. The pages were downloaded from a website called One Manga (see onemanga.com). The second step in the project was to insert this collection of images into an image-processing program called ImageJ. Researchers in the field of biomedical imaging Abramoff, Magalhaes and Ram (2004) describe ImageJ as an open-source software tool developed at the National Institute of Mental Health in Maryland, USA to analyze scientific imagery (p. 2). By performing pixel value statistics on the images that are inserted, the software defines them on eight visual features: median brightness, standard deviation, entropy, sobel, contrast, correlation, energy and homogeneity (Manovich 2010, Douglass, Huber and Manovich 2011: 14). Third, the ImageJ software processed the images on the aforementioned visual features. While processing, the ImageJ software generated numerical descriptions that described the score of each individual manga image for each of the eight visual features that ImageJ can measure (Manovich and Douglass 2011: 326). The fourth step in the research process was to visualize this collection of images. For that Manovich extended ImageJ with ImagePlot, a self-developed tool that can visualize images on the basis of the numerical descriptions for visual features as defined by ImageJ (Manovich 2010, Manovich and Douglass 2011: 335). In order to visualize his dataset, Manovich ran the ImagePlot application in ImageJ and selected two variables of interest. He opted for two visual features: contrast (the range of gray tones in an image) and mean of entropy (the average amount of texture and fine details in an image). Fifth, in order to create a visualization that is manageable in that it can be observed by the eye of a researcher, the ImagePlot software reduced the size of the manga pages to about 2% of their original size. This is a manipulation of the dataset that Manovich calls 'sampling'. Last, the ImagePlot software rendered all the reduced images into a graph according to the numerical values that ImageJ generated for contrast and mean (see Figure 22).

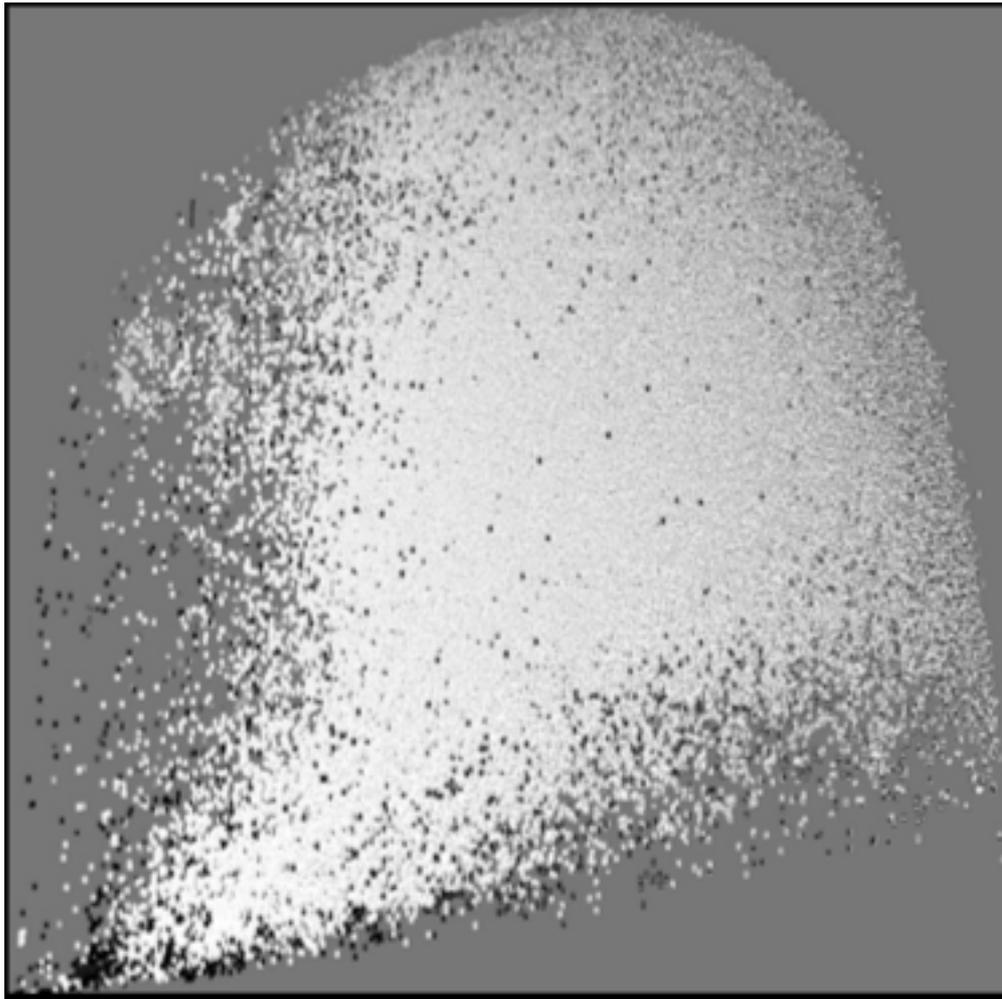


Figure 22: *Manga Style Space*; a visualization of 1,074,790 scanlated manga images (*x*-axis: contrast; *y*-axis: mean of entropy) (Manovich and Douglass 2011).

### 2.3 Describing Manga Style Space

Since this graph in Figure 22 depicts the manga images directly instead of points that represent the images, Manovich calls this output an 'image graph' or an 'image plot'. This way of creating a graph is relatively new. With mainstream information visualization techniques, information is mapped onto a graph by translating data into graphical primitives like lines and points. This new method does not involve the translation of the images into graphical primitives. Instead, the images are reduced by the ImagePlot software and visualized immediately. Manovich calls this visualization technique 'direct visualization' (Manovich 2010: 8, 10, 11, 17).

The direct visualization of scanlated manga imagery has resulted in an image graph where the bottom of the image represents the *x*-axis and the left borderline of the image represents the *y*-axis. For the *x*-axis Manovich selected a variable called contrast. Manovich explains that the images that only contain black and white appear on the extreme right and the images where all grey tones fit within a smaller range appear on the extreme left side of the visualization. For the *y*-axis Manovich selected the variable mean of entropy. The pages that are graphic and contain the least amount of detail appear in the bottom part of the visualization and pages with fine textures and lots of details appear in the upper right part. In between the four extremes there are many possible stylistic variations. This is why Manovich also describes this image graph as 'the space of visual possibilities explored in Japanese

manga culture' or 'Manga Style Space' (Douglas, Huber and Manovich 2011: 33, Manovich and Douglass 2010).

#### *2.4 Knowledge claim based on Manga Style Space*

Besides describing the image graph in Figure 22, Manovich also interprets Manga Style Space. By looking at the way the images are distributed, the first observation that he made was that in some areas the concentration of images is denser than in other areas. From this he concluded that within some parts of Manga Style Space, cultural evolution explores all its possibilities (Manovich 2010: 30). Second, from the sparse areas in the bottom and the left part of the image graph, Manovich concluded that some stylistic possibilities that are possible in theory are never observed in practice and that some stylistic choices are used more often than others, which implies that within cultural evolution some directions are more important than others (ibid., p. 40). This led him to conclude that cultural evolution does not actualize all its possibilities and that therefore cultural variability is finite (ibid., p. 38). Third, looking at the central, most dense part of Manga Style Space, Manovich observed that visual style varies continuously (ibid., p. 33). Together, these three observations have led Manovich to formulate the following knowledge claim: the visual language of Japanese manga cannot be divided into discrete categories. This endangers the concept of visual style because this concept assumes that a set of works can be partitioned into a small number of discrete categories. Instead, Manovich suggests understanding manga visual language as a continuous variable (ibid., p. 33).

In the introduction I cited Manovich in his call for self-reflection when thinking of culture as data: "We need to think about the consequences of thinking of culture as data and of computers as the analytical tools: what is left outside, what types of analysis and questions get privileged, and so on" (Manovich 2009: 10). Let's take a closer look at the central claim Manovich makes on the basis of Manga Style Space. He suggests understanding the visual language of Japanese manga as a continuous variable because it cannot be divided into discrete categories. This claim assumes that Manga Style Space represents the visual language of Japanese manga. I have three objections to this assumption. First, the dataset that Manovich chose to process in ImageJ is merely a sample of the population of Japanese manga images over the last 37 years. This dataset cannot be considered as representative for Japanese manga as a whole. It consisted of 1,074,790 scanned Japanese manga pages that were available on One Manga. These images originate from 883 manga series that date from 1976 to present. However, Japanese manga is older than 35 years. Paul Gravett, a British journalist and an author on the subject of comics, informs us that manga has a long and complex pre-history in early Japanese art. The term manga was first coined in 1814 by a Japanese artist called Hokusai. In its modern form manga dates from 1945, the first publication being instigated by the bombings of Hiroshima and Nagasaki that lead to Japan's surrender in World War II and the American occupation (Gravett 2004: 6). Recalling Manovich's demand for self-reflectivity we can say that in choosing this particular dataset a lot of other Japanese manga pages were left outside of the analysis. If there is a knowledge claim to be made on the basis of this dataset, it has to include that the material is merely a sample of the imagery that constitutes Japanese manga culture, namely 1,074,790 scanned Japanese manga images that date from 1976 to present. My second objection to the assumption that Manga Style Space represents the visual language of Japanese manga concerns the truthfulness of the dataset as visualized by the ImagePlot software. During the visualization process the quality of the full-color manga pages, like the cover pages, had to be reduced to black

and white images: "To be able to fit all images into a single large image (the original is 44,000 x 44,000 pixels - scaled to 10,000 x 10,000 for posting to Flickr), we rendered everything in grey scale" (Douglass, Huber and Manovich 2011: 34). However, since contrast is defined as "the difference in color and brightness" (Wolfram Mathematica Documentation Center 2011), reducing a dataset to black and white is likely to make a difference in the way the dataset is plotted on an image graph. When asking ourselves what questions get privileged after reducing full-color manga pages to black and white manga pages, we can conclude that these are not questions about the original dataset but questions about the modified dataset. If there is a knowledge claim to be made on the basis of this dataset, it has to include that the reliability of the visualization is questionable since the ImagePlot software rendered a part of the dataset in non-original coloring. Third, I question the fact that Manga Style Space can be an indicator for a visual language. Manga Style Space is a visualization in which the positions of images are mapped according to their numerical values on two visual features as measured by the ImageJ software. However, the variables contrast and mean of entropy cannot be regarded as representative for the concept of visual language as a whole. In order to get an impression of the visual language of this dataset, the dataset would also have to be analyzed and visualized on the basis of the other visual variables and combinations of these variables. The analysis that was privileged in this study has not been the analysis of a visual language, but the analysis of the variables contrast and mean of entropy. If there is a claim to be made on the basis of Manga Style Space, it has to include that it presents the visual dataset when letting ImagePlot create a visualization on the basis of the variables contrast and mean of entropy.

With these three objections I do not want to suggest that there is no knowledge that can be produced by conducting research in the paradigm of Cultural Analytics. Manga Style Space does allow making knowledge claims about visual characteristics of the manga imagery, yet these claims should incorporate more information about the nature of the dataset and the visualization process in order to guard self-reflectivity.

Taking my three objections into account, regarding the formulation of knowledge claims I propose to first provide a description of the dataset and a description of the software processing and then formulate one or multiple knowledge claims. In the case of Manga Style Space I suggest the following: The dataset consists of 1,074,790 pages of scanlated Japanese manga that originate from 883 manga series from 1976 to present, that were downloaded from One Manga. In order to be able to render the imagery into a manageable visualization not all the material could be rendered in its original coloring. This possibly affected the way the imagery has been visualized. When letting the ImageJ software measure this dataset on contrast and mean of entropy and letting the ImagePlot software visualize this dataset, the visualization shows that in some areas all possible combinations of contrast and range of fine textures have been realized in the dataset. In other areas there are combinations of these two visual features that are possible in theory yet have not been observed in the dataset.

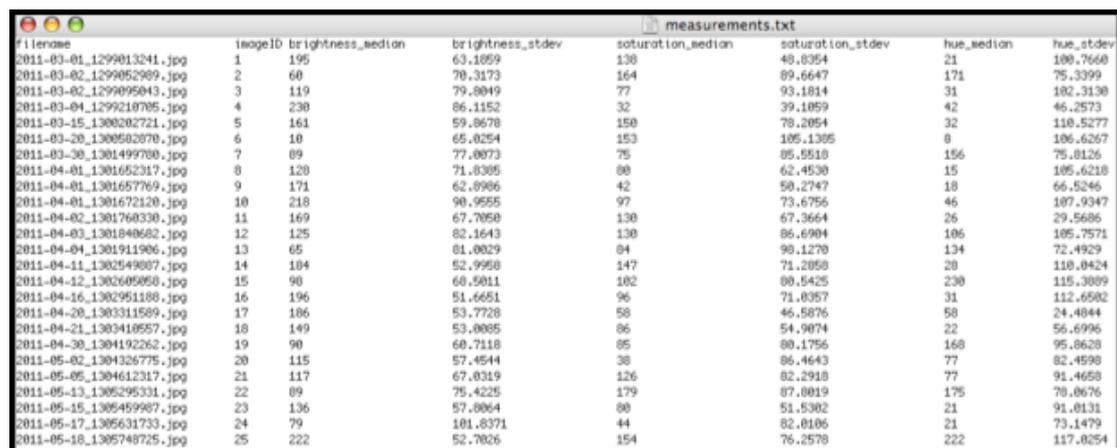
### 3. Performing Cultural Analytics on Instagram database material

Having analyzed Instagram as a digital image database and Cultural Analytics as a digital visual research method, in this chapter I will perform research in the paradigm of Cultural Analytics myself. Instead of manga imagery I used my personal Instagram database material to create three small-scale visualizations. Taking into account both my objections regarding the knowledge claim made by Lev Manovich on the basis of Manga Style Space and the information about the nature of Instagram database material, I will discuss these visualizations and formulate the knowledge claims that directly result from them.

#### 3.1: @ankehans Style Space 1

##### 3.1.1. The making of @ankehans Style Space 1

In creating visualizations I have followed the same procedure as Manovich did in creating Manga Style Space. First, I collected the input for the visualizations. I downloaded my personal Instagram images (accountname: @ankehans) to my hard drive. In order to do this I used Instaport, the third party developed Instagram service that I have mentioned in the first chapter of this thesis. The directory contained 129 filtered images that date from March 1, 2011 to December 15, 2011. These images represent my first nine and a half months as an Instagrammer. While downloading these images to my hard drive the Instaport software generated 129 filenames that represent the dates that I added the images to the Instagram database (the first file name is '2011-03-01\_1299013241.jpg' and the last one is '2011-12-15\_1323987336.jpg'). Second, I inserted this dataset into ImageJ for digital image analysis. Third, the ImageJ software defined all 129 images on three visual features that were available to me: brightness, saturation, and hue. For each of these three features, the software calculated both their median value and their standard deviation value. While processing, the ImageJ software generated a text document with the file names, ImageID (a sequential number of the image in the folder selected) and six columns of numerical descriptions of visual features (see Figure 23).



filename	imageID	brightness_median	brightness_stddev	saturation_median	saturation_stddev	hue_median	hue_stddev
2011-03-01_1299013241.jpg	1	195	63.1859	138	48.8354	21	100.7668
2011-03-02_1299052989.jpg	2	60	78.3173	164	89.6647	171	75.3399
2011-03-02_1299059043.jpg	3	119	79.8049	77	93.1814	31	182.3138
2011-03-04_1299218785.jpg	4	238	86.1152	32	39.1059	42	46.2573
2011-03-15_1300282721.jpg	5	161	59.8678	158	78.2054	32	118.5277
2011-03-20_1300582878.jpg	6	10	65.8254	153	105.1305	8	106.6267
2011-03-30_1301499788.jpg	7	89	77.0873	75	85.5518	156	75.8126
2011-04-01_1301652317.jpg	8	128	71.8385	88	62.4538	15	105.6218
2011-04-01_1301657769.jpg	9	171	62.8986	42	58.2747	18	66.5246
2011-04-01_1301672128.jpg	10	218	90.9555	97	73.6756	46	187.9347
2011-04-02_1301768338.jpg	11	169	67.7058	138	67.3664	26	29.5686
2011-04-03_1301848682.jpg	12	125	82.1643	138	86.6984	186	105.7571
2011-04-04_1301911986.jpg	13	65	81.8829	84	98.1278	134	72.4929
2011-04-11_1302549887.jpg	14	184	52.9958	147	71.2858	28	118.8424
2011-04-12_1302609888.jpg	15	98	68.5811	182	88.5425	238	115.3889
2011-04-16_1302951188.jpg	16	196	51.6651	96	71.8357	31	112.6582
2011-04-20_1303311589.jpg	17	186	53.7728	58	46.5876	58	24.4844
2011-04-21_1303418557.jpg	18	149	53.8885	86	54.9874	22	56.6996
2011-04-30_1304192262.jpg	19	98	68.7118	85	88.1756	168	95.8628
2011-05-02_1304326775.jpg	20	115	57.4544	38	86.4643	77	82.4598
2011-05-05_1304612317.jpg	21	117	67.8319	126	82.2918	77	91.4658
2011-05-13_1305295331.jpg	22	89	75.4225	179	87.8819	175	78.8676
2011-05-15_1305459987.jpg	23	136	57.8864	88	51.5382	21	91.8131
2011-05-17_1305631733.jpg	24	79	181.8371	44	82.8186	21	73.1479
2011-05-18_1305748725.jpg	25	222	52.7826	154	76.2578	222	117.8254

Figure 23: The first part of the text document with numerical descriptions generated by ImageJ when defining my Instagram images on median brightness, standard deviation of the brightness, median saturation, standard deviation of the saturation, median hue and standard deviation of the hue.

The fourth step was to run the ImagePlot application in ImageJ and select two variables of interest for the image graph. As the file names equal the dates my images were added to Instagram, I was enabled to analyze my images on a time dimension. Therefore I chose ImageID as a variable for the x-axis. This means that the software positioned the imagery in its original sequence. Without any pre-conceived expectations regarding brightness, saturation and hue values of my imagery, I randomly picked median brightness as a variable for the y-axis. Fifth, the ImagePlot software sampled the images to about 2% of the original size. After resizing, the ImagePlot software spread the values for publication date over the x-axis and the calculated values for median brightness over the y-axis. It plotted all 129 images on a graph according to their publication date and their numerical descriptions for median brightness. Since the data set is relatively small ImagePlot was able to render all the images in their original color (see Figure 24). This process is what Manovich and Douglass call "turning 'culture' into 'data' " (2011):

We need a set of x and y coordinates to draw the points. This means that we have to map a cultural process into a set made from pairs of numbers. In each pair, one number would represent a moment in time (x-axis) and the other number would represent some characteristic of the process at that time (y-axis.) In short, we have to turn "culture" into "data" (p. 324).

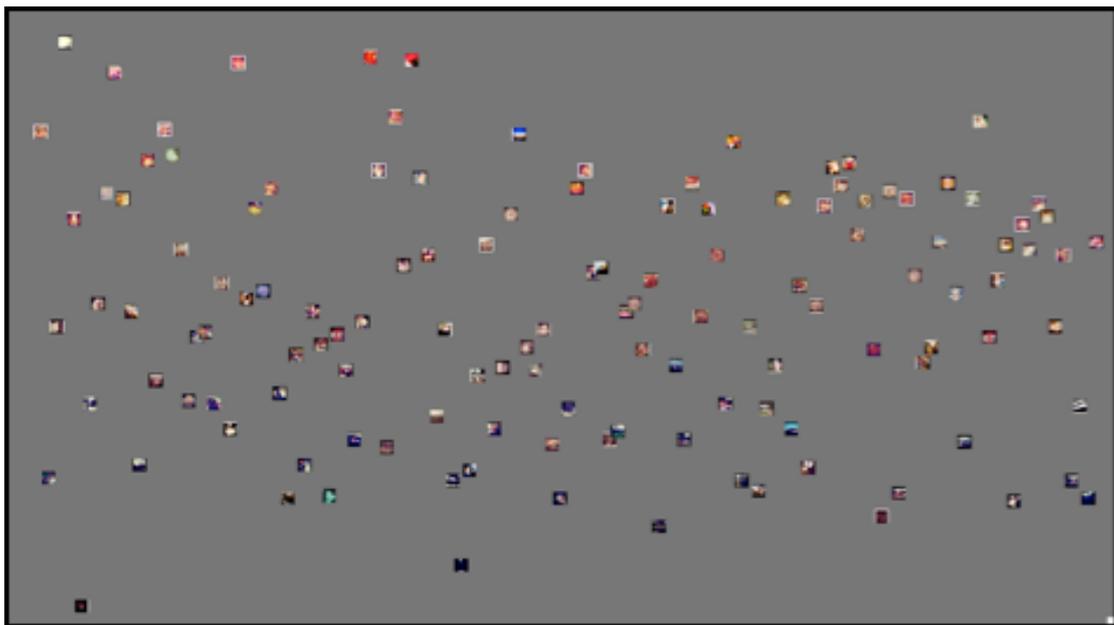


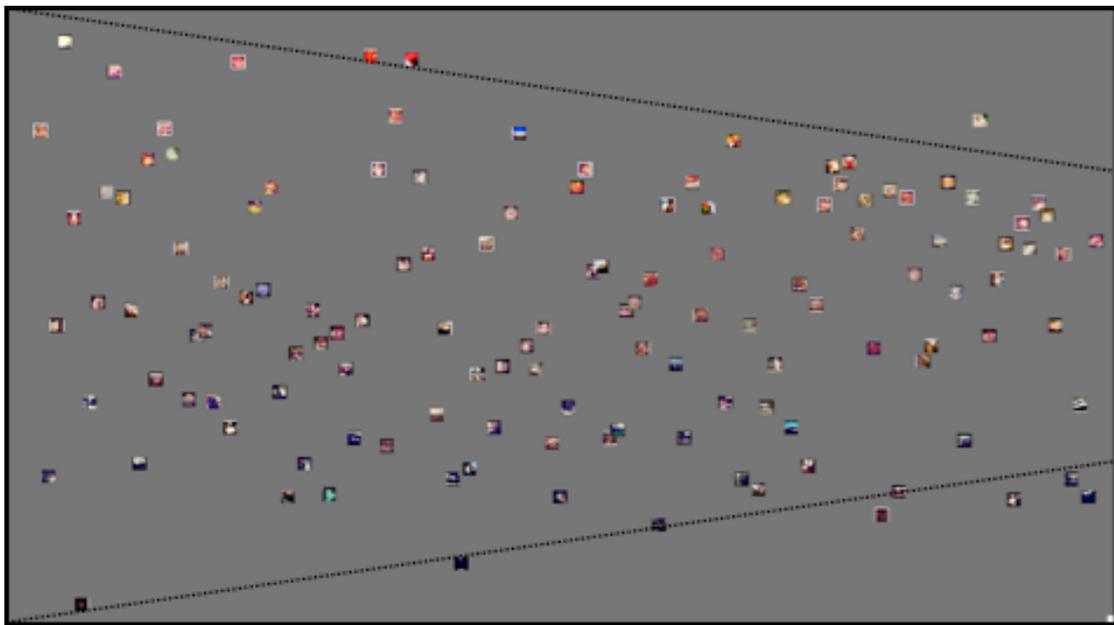
Figure 24: @ankehans Style Space 1; a visualization of 129 Instagram images over a nine and a half month period (x-axis: ImageID –as an approximation of date of image publication; y-axis: median brightness).

### 3.1.2 Describing @ankehans Style Space 1

Looking at Figure 24, we see an image graph in which the 129 Instagram images that I added to Instagram through the application are distributed in a heavily reduced format, yet in their original coloring. However, the colors of these images have been transformed by different filters of the Instagram application. These colors thus are no objective report of colors as they appear in reality. The images are clustered by the ImagePlot software on the basis of their values on two variables as measured by the ImageJ software. For the x-axis I selected the variable 'ImageID' that equals the date the images were added to the Instagram database. The images that were published in

the beginning of March appear on the left and the images that were published in the first half of December appear on the right. For the y-axis I selected median brightness, where brightness defines the average luminance of an image (Wolfram Mathematica Documentation Center 2011). The images that contain a low luminance appear in the lower part of the visualization and the images that contain a high luminance appear in the upper part. The visualization thus shows the temporal development of luminance of my Instagram images.

Looking at the distribution of the images in @ankehans Style Space 1 I observe a large variety of luminance in my images. However, considering the facts that on the x-axis the images have been plotted on the basis of their publication date, and that I have rarely published more than one or two images per day, it would be almost impossible to have a core of images. When looking from the perspective of the y-axis, I can say that my images roughly cover the range of median brightness values. Even though in theory there are options regarding the luminance of images that I have not added to the Instagram database in my first nine and a half months as an Instagrammer, these options would contain minor deviations from the median brightness values in my existing images. Taking into account the x-axis I observe that the temporal development of median brightness values is continuous: over time I have added images with a range of median brightness values. Third, I observe that the corner areas of the right half of @ankehans Style Space 1 are relatively sparse (see Figure 25 for @ankehans Style Space 1 with help lines that show these corner areas; these lines have not been determined statistically but have been drawn by myself for illustration). This means that, when compared to the first half of the publication period, over the latter half of this period the options that I explored regarding luminance are slightly drawn towards average brightness values. In the Instagram imagery that I published in the first half of the time period I thus have explored more possibilities regarding brightness than in the latter half.



*Figure 25: @ankehans Style Space 1 with help lines that show a slight concentration towards median brightness values over nine and a half months.*

### 3.1.3 Knowledge claims based on @ankehans Style Space 1

Based on the description of @ankehans Style Space 1 in 3.1.2 I can now formulate the following knowledge claims. The dataset consists of 129 Instagram images that were added to the Instagram database between March 1, 2011 to December 15, 2011. The images are in their original coloring, yet the colors have been transformed by Instagram filters. This means that the colors do not represent the colors as they appeared in reality during the creation of the image. When letting the ImageJ software measure this dataset on ImageID and median brightness and letting the ImagePlot software visualize this dataset on an image graph, the visualization shows that there are possibilities concerning the luminance of my images that in theory are possible yet I have not added in my first nine and a half months as an Instagrammer. Second, the temporal development of luminance in my imagery is continuous. Third, over time my images show a slight tendency towards images with average luminance, for at the end of the publication period the extreme luminant and the extreme non-luminant images have disappeared.

### 3.2: @ankehans Style Space 2

#### 3.2.1 The making of @ankehans Style Space 2

For the second visualization that I have created the first three steps of the research process have been identical to the process as described for @ankehans Style Space 1. I have used the same numerical values for visual features that were generated by the ImageJ software on the same dataset. However, in the fourth step of the process I changed my choice for median brightness to another visual variable for the y-axis: median hue. Hue is a visual feature that represents the dominant color as seen by an observer (Wolfram Mathematica Documentation Center 2011). Choosing median hue for the y-axis implies that my images will be positioned according to the shading of their dominant color as seen by an observer: on the lowest part of the visualization will be the dominantly red colored images. As we go upwards, the dominant color of the images becomes orange, yellow, green, blue, and finally purple. See Figure 26 for @ankehans Style Space 2.

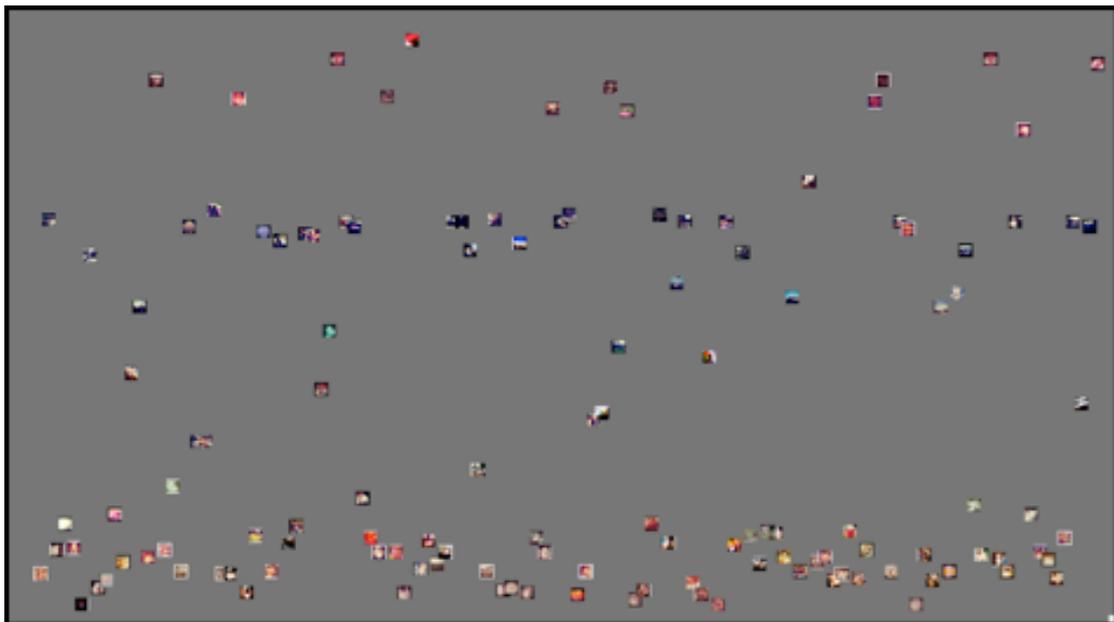


Figure 26: @ankehans Style Space 2; a visualization of 129 Instagram images over a nine and a half month period (x-axis: ImageID; y-axis: median hue).

### 3.2.2 Describing @ankehans Style Space 2

When looking at @ankehans Style Space 2, we can see an image graph on which are distributed 129 heavily reduced Instagram images. These images have all been added to the Instagram database by myself using the Instagram application over a period from March 1, 2011 to December 15, 2011. These images are in their original uploaded colors. However, these colors have been transformed by filters of the Instagram application. The colors thus are no objective report of colors as they appeared when creating the image. The images have been clustered by the ImageJ and ImagePlot software. For the x-axis the software clustered according to the dates the images were added to the Instagram database; for the y-axis it clustered on the basis of the numerical values of median hue. This variable represents the dominant color of an image as seen by an observer. This means that on the y-axis my images are positioned according to the shading of their dominant color. In the bottom part will be the mainly red and orange images, then will follow the yellow, green, blue and purple images as we go upwards. @ankehans Style Space 2 represents the development of hue values in my Instagram imagery over a nine and a half month period.

Looking at the distribution of the images in @ankehans Style Space 2 I first observe that there is no core of images to be located. This can be explained by the fact that for the x-axis the images have been plotted on a time dimension. Second, observing the visualization from the left to the right I observe a consistent temporal pattern. Over a nine and a half month period, there have been no notable trends or changes in my images regarding median hue values. This temporal pattern can be defined as follows: there is a relatively high density in the lowest quarter of the visualization, which implies a high amount of red and orange images. The second layer shows a low density, implying few yellow and green images. The third layer of the visualization displays an average density, implying an average amount of blue images. The upper part of the visualization shows a low density, implying few purple images (see Figure 27 for @ankehans Style Space 2 with help lines that show this temporal pattern; again, these lines have not been determined statistically but have been drawn by myself for illustration).

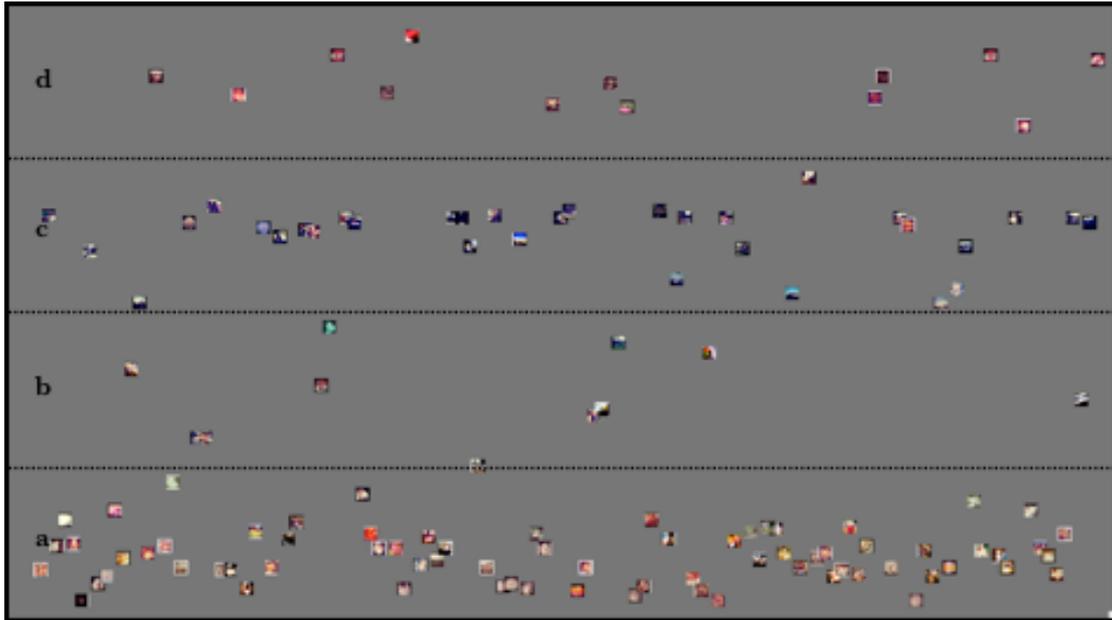


Figure 27: *@ankehans Style Space 2 with help lines that show that the visualized dataset can be divided into four areas. Part 'a' shows a high amount of red and orange images. Part 'b' shows a low amount of yellow and green images. Part 'c' shows an average amount of blue images. Part 'd' shows a low amount of purple images.*

### 3.2.3 Knowledge claims based on @ankehans Style Space 2

From the description of @ankehans Style Space 2 I can now deduce the following knowledge claims. The dataset consists of 129 Instagram images that were added on Instagram from March 1, 2011 to December 15, 2011. The images are in their original coloring, yet these colors have been transformed by Instagram filters. When letting the ImageJ software measure this dataset on ImageID and median hue and letting the ImagePlot software visualize this dataset, the visualization shows that over this time period I have consistently added an above-average amount of images that are predominantly perceived as red and orange, an average amount of images that are predominantly perceived as blue and a below-average amount of images that are predominantly perceived as yellow, green and purple.

## 3.3: @ankehans Style Space 3

### 3.3.1 The making of @ankehans Style Space 3

For the third visualization of my Instagram imagery the first three steps of the visualization process again have been identical to the process in creating @ankehans Style Space 1. This time I did not choose for a visualization with a temporal dimension, but a visualization based on two visual variables. In the fourth step of the research process I chose median hue as a variable for the x-axis, implying that on the x-axis my images will be positioned according to the shading of their dominant color as seen by an observer. On the left side of the visualization will be the dominantly red colored images. As we go rightwards, the dominant color becomes orange, yellow, green, blue, and finally purple. In @ankehans Style Space 2 this variable was positioned on the y-axis. For the y-axis I chose the visual variable median saturation, where saturation is defined as "the amount of dilution of the color with white light" (Wolfram Mathematica Documentation Center 2011). This means that the images that contain the least amount of color will be positioned in the bottom of the visualization

and the most colorful images will be positioned in the upper part of the visualization. See Figure 28 for @ankehans Style Space 3.

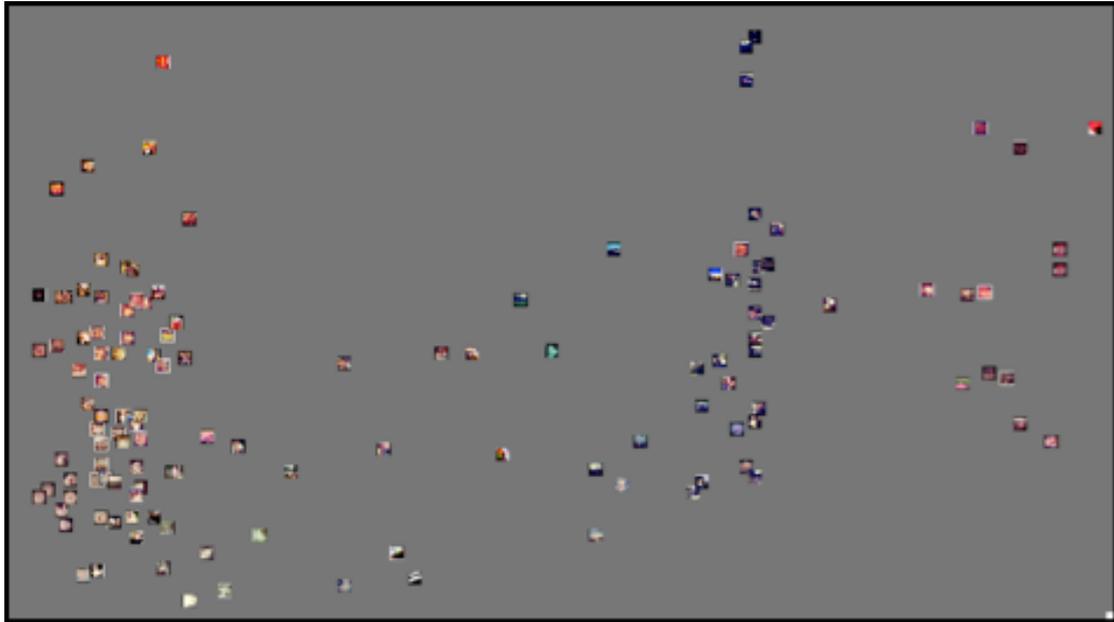


Figure 28: @ankehans Style Space 3; a visualization of my 129 Instagram images over a nine and a half month period (x-axis: median hue; y-axis: median saturation).

### 3.3.2 Describing @ankehans Style Space 3

@ankehans Style Space 3 is a visualization on which is distributed my Instagram imagery. These are 129 images that have been heavily reduced in order to be rendered into this image graph. All these images I have added to the Instagram database over a period from March 1, 2011 to December 15, 2011. The images are rendered in their original colors. However, these colors have been modified by filters of the Instagram iPhone application and thus cannot be regarded as a truthful report of the colors in reality. The images have been processed by the ImageJ software and visualized by the ImagePlot software. For the x-axis the ImagePlot software clustered the images according to their values for the visual feature median hue; for the y-axis it clustered on the basis of the values for median saturation. @ankehans Style Space 3 thus represents both the colors of my images (x-axis) and the intensity of these colors in my images (y-axis).

Looking at the way the images are distributed in @ankehans Style Space 3, I first observe that there are two relatively dense areas of images: one on the left side of the visualization and one in the center-right area of the visualization (see Figure 29 for @ankehans Style Space 3 with non-statistically determined, self-drawn circles to illustrate these two concentrations of images). The concentration on the left side of the image (see Figure 29, circle 'a') is positioned in an area with relatively low median saturation values. This means that a considerable part of the images that I have added on Instagram are dominantly red and orange colored, with a below-average intensity. In between the two concentrations of images we find a large yet thin populated area of images that are yellow and green, with a below-average intensity. Moving rightwards, we see a second concentration of images (see Figure 29, circle 'b'). These images are predominantly perceived as blue. The intensity of the images in this area varies: there are three images that have a high intensity of blue, while the rest of the blue images has an average intensity. On the right side of the visualization are positioned the

images that are perceived as purple. This area is thin populated and contains approximately the same amount of images as the area with yellow and green images. The intensity of the purple images can be called average.

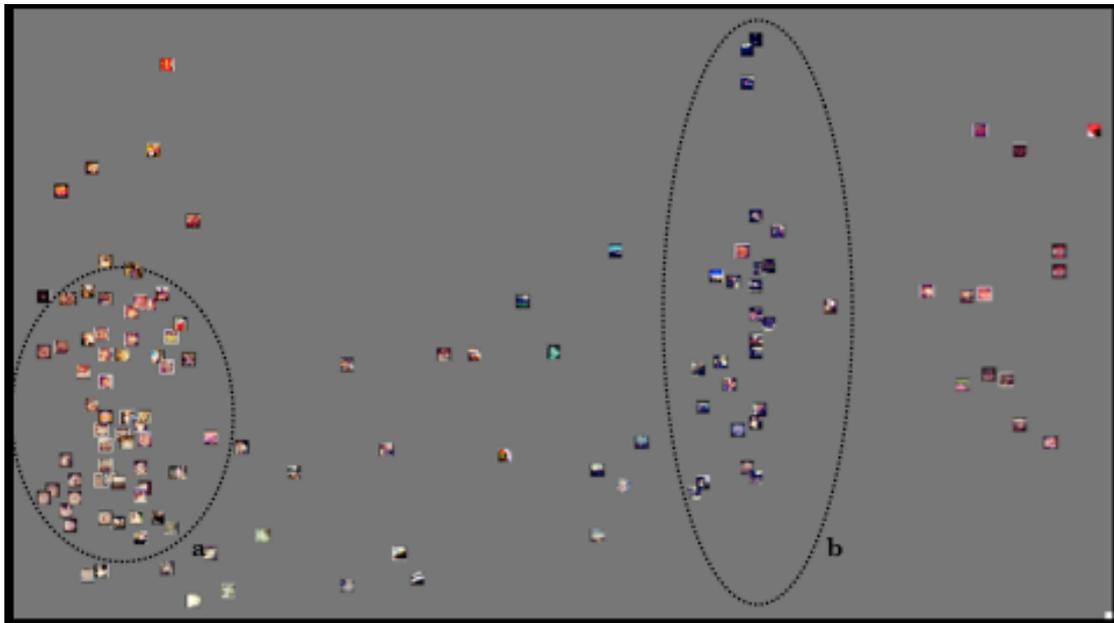


Figure 29: *@ankehans Style Space 3*. Circle 'a' shows a concentration of red and orange images. Circle 'b' shows a small concentration of blue images.

### 3.3.3 Knowledge claims based on *@ankehans Style Space 3*

From the description of *@ankehans Style Space 3* I can formulate the following knowledge claims. The dataset consists of 129 Instagram images that were added to the Instagram database from March 1, 2011 to December 15, 2011. The images are in their original coloring, yet the colors have been transformed by Instagram filters. When letting the ImageJ software measure this dataset on median hue and median saturation and letting the ImagePlot software visualize this dataset, the visualization shows that most of the images are predominantly perceived as red and orange. These images have a relatively low intensity. After red and orange images, blue images appear second most frequently, with an average intensity. There are few images in which yellow, green and purple constitute the predominant colors. Their intensity can be called below-average and average respectively.

Cultural Analytics is a digital visual research method that has arisen in a response to vast amounts of user-generated visual media content. It borrows image-processing software –ImageJ– from scientific fields that employ it to analyze individual images. The extension of this software –ImagePlot– bridges the gaps between scientific versus cultural analysis and the analysis of individual images versus large amounts of images. I have employed both ImageJ and ImagePlot to perform a cultural analysis of a small part of user-generated visual media content. In this analysis I took into account both the nature of the visual dataset and the questions that are privileged when subjecting a visual dataset to a computational analysis. We can conclude that at this moment in time, with current image-processing and visualization techniques, the cultural analysis of my dataset has not led to relevant cultural insights. The knowledge claims that I have formulated on the basis of the three visualizations are of a rather abstract nature. They either concern statements about the temporal development of a visual

characteristic –brightness, hue or saturation– in the dataset, or statements about the arrangement of the material based on its analysis on two of these three visual characteristics. Therefore, at this moment in time, Cultural Analytics can best be seen as a transfer of methods from scientific fields to the field of culture.

#### *3.4: Knowledge claims based on Instagram database material*

Future knowledge claims based on datasets from the Instagram database should incorporate the following information about the nature of this material. First, this material is created by Instagramers: people that can be typified as early adopters who own an iPhone running iOS 3.1.2. and above, who have downloaded the Instagram iPhone application and have signed up for an Instagram account. Their motives behind creating and sharing their imagery are documentation of their lives and cultural enrichment respectively. Second, the imagery originates from an unstable, flexible digital database. The database is organized by metadata that is assigned by either the authors of the images or by other Instagramers. Since tagging on Instagram serves more functions than mere categorization –notably a social function–, tags do not always reflect what is depicted on an image. This should be taken into account when formulating knowledge claims on the basis of visualizations of datasets that have been extracted from the Instagram database on the basis of hashtags. Third, tagging within Instagram is not obligatory. This implies that not all the database material is tagged. Since the Instagram software cannot recognize untagged material, a researcher will not be able to retrieve all the material on which is depicted a subject of interest. The knowledge claims based on the visualizations of Instagram datasets should include that it only concerns Instagram imagery that has been tagged with the hashtag of interest.

#### *3.5 A future perspective*

The current development of recognition software potentially compensates the inconsistent attachment of metadata to Instagram database material. Dr Bennie Coetzer, managing director of Thales Advanced Engineering (the Johannesburg-based manufacturer of high level electronic surveillance security systems), informs us about digital image processing in the security industry. In the security industry digital image analysis is employed to proactively detect and prevent crime. Digital CCTV (Closed Circuit TeleVision) security systems record videos that are subjected to a computational analysis of the recorded images. Their software is programmed to match facial images and vehicle number plates in order to determine whether a person or object has been captured at previous incidents (Coetzer 2004: 1-2). According to Coetzer, the potential of CCTV analysis is large:

Image analysis has the potential to very rapidly become a pervasive technology – it can be used in a host of applications such as manufacturing and production for quality control; to automatically grade fruit by size, colour and other parameters; in military applications to track targets and guide missiles; for traffic pattern analysis monitoring and adjustment of traffic lights to ease congestion, the list is endless (Coetzer 2004: 1).

Recognition software integrated into the Cultural Analytics software would result in software that could 'read' what is depicted on images that it processes. This means that the computational analysis of Instagram imagery would no longer be dependent on the inconsistent assignment of user-generated metadata. When recognition software would be incorporated, the certainty of formulating knowledge claims concerning a

precise cultural phenomenon as depicted on Instagram imagery would increase. For example, the integration of recognition software would enable to take analyze aesthetic aspects of Instagram imagery, which at this moment in time cannot be taken into account.

## 4. Conclusion

In this thesis I have investigated the kind of knowledge that is produced when performing Cultural Analytics on Instagram database material. In the first chapter I analyzed Instagram as a flexible, digital database that stores user-generated visual media content. I employed De Mul's four elements that constitute a database ontology: *Add*, *Browse*, *Change* and *Delete*. This model suffices to describe the technological structure of the Instagram database, yet the user practices of tagging database material required two refinements of the database ontology. First, a division has to be made between 'imagery' and 'metadata' as two elements that can be added to the Instagram database. Second, concerning metadata, users can add *a priori* metadata to imagery by tagging in captions, and *a posteriori* metadata by liking and tagging in comments. The analysis of Instagram has provided us with information about the nature of the Instagram database material.

In the second chapter I have described Cultural Analytics as a digital visual research method. The aim of this method is to transfer qualitative research methods that are already being employed in multiple scientific fields, and apply them to visual cultural datasets. By processing datasets on visual variables and visualizing the datasets on the basis of the numerical descriptions for these visual variables, culture is transformed into data. I have discussed Manga Style Space, a large-scale visualization of scanlated Japanese manga imagery created by Lev Manovich. On the basis of this visualization, Manovich formulated a knowledge claim that assumes that Manga Style Space can represent the visual language of Japanese manga. Recalling the need for self-reflectivity in any Cultural Analytics study, I have formulated three objections to this assumption. First, the dataset that Manovich employed did not represent Japanese manga culture as a whole. Second, the knowledge claim did not incorporate the fact that not all the images were visualized in their original coloring which likely affects the knowledge claims that can be made on the basis of Manga Style Space. Third, Manga Style Space is based on a mapping of the dataset on the variables contrast and mean of entropy that do not represent the concept of visual language as a whole. Considering these three objections, I concluded that the analysis of Manga Style Space does not privilege questions about the visual language of a cultural phenomenon. Instead, it allows to formulate knowledge claims about the arrangement of the dataset based on its analysis on two visual variables. At this point of the research it already was clear that the hypothesis of this research –performing Cultural Analytics on Instagram database material will allow a researcher to make claims about the visual language of Instagram database material– would not be supported.

In the third chapter I performed research in the paradigm of Cultural Analytics myself. I employed the ImageJ and ImagePlot software to create three visualizations of my own Instagram imagery. On the basis of the knowledge claims that I formulated I can answer the research question that has been central to this thesis –What kind of knowledge is produced when performing Cultural Analytics on the Instagram image database? The knowledge claims that are produced when performing Cultural Analytics on Instagram database material concern the temporal development of a visual characteristic –brightness, hue or saturation– in the dataset concerned, or the arrangement of the dataset based on its analysis on two of these three visual characteristics.

At this moment in time the results of a cultural analysis of Instagram database material are less pertinent than expected. However, the current development of

recognition software as applied in CCTV analysis for example, brighten the future perspective.

## References

- Abramoff, M.D., Magalhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International* 11: 36-42.
- Amazon (2011). Amazon EC2 FAQs. <<http://aws.amazon.com/ec2/>> [last checked 30 December 2011].
- Coetzer, B. (2004). Image analysis takes hold as the evolutionary CCTV trend. *Hi-Tech Security Solutions* June 2004. <<http://www.securitysa.com/>> [last checked 30 December 2011].
- Dacree, J.E. and Huskisson, E.C. (1989). The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. *Rheumatology* 28: 506-510.
- Douglass, J., Huber, W. and Manovich, L. (2011). Understanding scanlation: how to read one million fan-translated manga pages. <<http://softwarestudies.com/>> [last checked 30 December 2011].
- Duval, E., Hodgins, W., Sutton, S. et al. (2002). Metadata Principles and Practicalities. *D-Lib Magazine* 8: 1-10.
- Gilder, G. (2006). The Information Factories. *Wired*, 14 October 2006. <<http://www.wired.com/wired/archive/14.10>> [last checked 30 December 2011].
- Gonzalez, P. (2011). Personal e-mail conversation on characteristics of Instagramers, 16 August 2011.
- Gravett, P. (2004). *Manga: sixty years of Japanese comics*. London: Laurence King Publishing Ltd.
- Identi.ca (2011). What is a hashtag? <<http://identi.ca/>> [last checked 30 December 2011].
- Instagram (2011a). [instagr.am](http://instagram.com/). <<http://instagram.com/>> [last checked 30 December 2011].
- , (2011b). Blog hashtags. <<http://instagr.am/blog/38/hashtags>> [last checked 30 December 2011].
- , (2011c). Terms of use. <<http://instagram.com/legal/terms/>> [last checked 30 December 2011].
- Instaport (2011). Export & Download your Instagram photos with Instaport. <<http://instaport.me/>> [last checked 30 December 2011].
- Kessler, F. and Schäfer, M.T. (2009). Navigating YouTube: Constituting a Hybrid Information Management System. In: Snickars, P., Vonderau, P. (eds.) (2009). *The YouTube Reader*. Stockholm: National Library of Sweden, 275-291.

- Manovich, L. (2001). *The Language of New Media*. Cambridge, MA: The MIT Press.
- , (2007). Cultural Analytics: analysis and visualization of large cultural data sets. <[http://www.manovich.net/cultural\\_analytics.pdf](http://www.manovich.net/cultural_analytics.pdf)> [last checked 30 December 2011].
- , (2008a). *Software Takes Command*. <<http://lab.softwarestudies.com/2008/11/softbook.html>> [last checked 30 December 2011].
- , (2008b). Cultural Analytics: Visualizing Cultural Patterns in the Era of "More Media". <[softwarestudies.com/cultural\\_analytics/Manovich\\_DOMUS.doc](http://softwarestudies.com/cultural_analytics/Manovich_DOMUS.doc)> [last checked 30 December 2011].
- , (2009). How to Follow Global Digital Cultures, or Cultural Analytics for Beginners. <[http://softwarestudies.com/cultural\\_analytics/](http://softwarestudies.com/cultural_analytics/)> [last checked 30 December 2011].
- , (2010) Manga genre report. <<http://lab.softwarestudies.com/2010/11/one-million-manga-pages.html>> [last checked 30 December 2011].
- , (2011). Trending: The Promises and the Challenges of Big Social Data. <<http://lab.softwarestudies.com/2011/04/>> [last checked 30 December 2011].
- Manovich, L. and Douglass, J. (2011). Visualizing Change: Computer Graphics as a Research Method. In: Grau, O., Veigl, T. (eds.) (2011). *Imagery in the 21<sup>st</sup> century*. Cambridge, MA: MIT Press, 315-338.
- Manovich, L., Douglass, J. and Huber, W. (2011). Digital Image Analysis and Interactive Visualization of 1000000 Manga Pages. *Digital Humanities 2011*.
- McCune, Z. (2011a). Consumer Production in Social Media Networks: A Case Study of the "Instagram" iPhone App. Master's dissertation submitted 9 June 2011, University of Cambridge.
- McCune, Z. (2011b). Personal e-mail conversation on Instagram as a database, 12 July 2011.
- Mul, J. de (2009). The work of art in the age of digital recombination. In: Boomen, M. van den, Lammes, S., Lehmann, A.S. et al. *Digital Material. Tracing New Media in Everyday Life*. Amsterdam: Amsterdam University Press, 95-106.
- Mul, J. de (2010). Lecture 'Redesigning (Open) Design'. First addressed at PICNIC '10, 24 September, Amsterdam. <<http://opendesignnow.org/index.php/>> [last checked 30 December 2011].
- Rogers, E.M. (1962). *Diffusions of innovations*. New York: Free Press.
- Snickars, P. (2009) The Archival Cloud. In: Snickars, P., Vonderau, P. (eds.) (2009). *The YouTube Reader*. Stockholm: National Library of Sweden, 292-313.

Software Studies Initiative (2011). Software studies. <<http://lab.softwarestudies.com>> [last checked 30 December 2011].

Sol, S. (1998). Introduction to Databases for the Web: Pt. 1. <<http://iat.ubalt.edu/>> [last checked 30 December 2011].

Siegler, MG. (2011). The Latest Crazy Instagram Stats: 150 Million Photos, 15 Per Second, 80% Filtered. *TechCrunch*, 3 August 2011. <<http://techcrunch.com/2011/>> [last checked 30 December 2011].

The New York Times Bits (2011). A Photo-Sharing App With Bigger Aspirations. 19 October 2011. <<http://bits.blogs.nytimes.com/2010/>> [last checked 30 December 2011].

Vijayakumar, C., Gharpure, D. C. (2011). Development of image-processing software for automatic segmentation of brain tumors in MR images. *Journal of Medical Physics* 36: 147-158.

Wolfram Mathematica Documentation Center (2011). Image Processing – Mathematica 8 Documentation. <<http://reference.wolfram.com/mathematica/>> [last checked 30 December 2011].