

EMOTO – VISUALISING THE ONLINE RESPONSE TO LONDON 2012

Dr. Drew Hemment, Duncan of Jordanstone College of Art and Design, University of Dundee, Dundee DD1 4HN, UK. d.hemment@dundee.ac.uk

Abstract

In recent years we have moved from data scarcity to data abundance. As a response, a variety of methods have been adopted in art, design, business, science and government to understand and communicate meaning in data through visual form. *emoto* (emoto2012.org) is one such project, it visualised the online audience response to a major global event, the London 2012 Olympic and Paralympic Games. *emoto* set out to both give expression to and augment online social phenomena, that are emergent and only recently made possible by access to huge real-time data streams. This report charts the development and release of the project, and positions it in relation to current debates on data and visualisation, for example, around the bias and accessibility of the data, and how knowledge practices are changing in an era of so-called 'big data.'

Keywords: Data, Social Data, Data Visualisation, Data Art, London 2012, Olympics, Twitter.

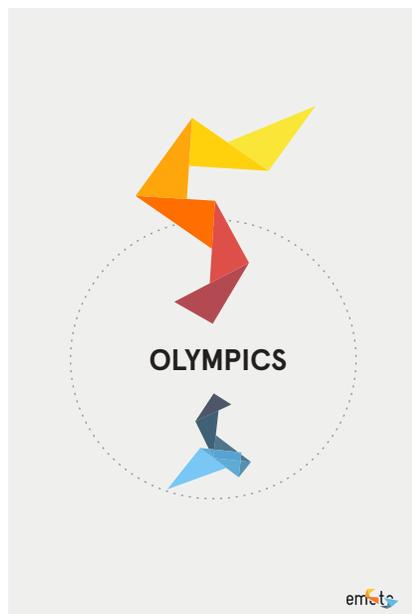


Fig. 1. emoto used origami-like shapes, to display the combination of negative and positive messages for each topic (detail). (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

Introduction

A requirement for data mining, analysis and visualisation has arisen out of new challenges to make sense of an unprecedented upsurge in the amount and complexity of data. This is seen across many domains, from government and scientific research to financial services and smart grids, not forgetting social data, millions of individual interactions that are continuously shared online [1,2].

As a societal tool, visualisation presents the means to interrogate and make intelligible different data dimensions which overlay our everyday — there follow claims that it enhances cognition through novel techniques for data capture and representation [3,4]. Combined with such techniques, this ever expanding domain of data has the potential to unlock new forms of insight, and has been described as “a new type of scientific instrument” for the social sciences [5].

This paper describes an artwork — *emoto* — which visualised the online audience response to the London 2012 Olympic and Paralympic Games. *emoto* set out to give expression to emerging online social phenomena that only becomes possible with access to huge real-time data streams. As such, it was among the first projects to visualise in real-time structured insight on the online response to a major global event [6].

The Data Olympics

The Olympic and Paralympic Games are global sporting events, and in 2012, more than any previous Olympics, London played host to a Games with an online audience [7]. *emoto* was proposed and was commissioned as a data artwork creating a new form of engagement for the Olympics audience, within the London 2012 Festival and Cultural Olympiad programme. One of the principles outlined for the London 2012 online and mobile services was the “comprehensive, unbiased provision of results” [8]. In one sense, *emoto* was a means to achieve this, given that a central tenet of data visualisation is it is *true to the data*.

The *emoto* team were interested in the Olympics as a big data event — huge amounts of data would be generated, from the stream of medals table data, to the interactions of the online audience. The drama of a sporting event is conventionally experienced in the roar of the crowd in the stadia, or through broadcast media, in the home, a bar or public square, alone or with family and friends. However, the 2012 Olympics, as experienced through the conversations and interaction of people online, presented an opportunity for different ways to observe and participate, and to reflect on the mood of individuals or a nation, as events unfold. Data analytics and visualisation offer a new way to access and read the audience response to the sporting event, and

furthermore, the side stories and controversies it engenders — from the staging of the opening ceremony to the number of seats allocated to sponsors. With implications for broadcast and print as well as ‘citizen’ media, such an approach can portray the Games from the perspective of the sporting fans, and also offer audiences new dimensions to the shared experience of the event [9].

The *emoto* London 2012 Data Visualisation

Launching during the London 2012 Opening Ceremony, *emoto* collected and analysed 12.5 million Twitter messages, consumed in real-time via Twitter’s public API, from 27 July to 12 August 2012. As the Games unfolded, *emoto* processed this data stream and analysed the messages for content and emotional tone. Presentation was in three forms: a real-time sentiment visualisation web application, data journalistic analysis of the Games, and a physical interactive data sculpture.

a) Data mapping

In order to measure the online effect of an event in the real world, the project began by considering the most interesting aspects of the Olympics. The team focused on the response of observers to the Games, and sought data to support rich and nuanced information. One early solution was to consider placing microphones near public screens around the world to record the audience cheering. However, the most accessible and richest source of data that would encompass the response to London 2012 would be social media, and in particular Twitter [10].

b) Visualisation and design process

The visualisation design process — based on the practice of the design lead, Moritz Stefaner, and Studio NAND — begins by examining the available data, and iteratively coding, prototyping and testing. The challenge was to turn the attributes and patterns of social data into a meaningful visualisation that conveys

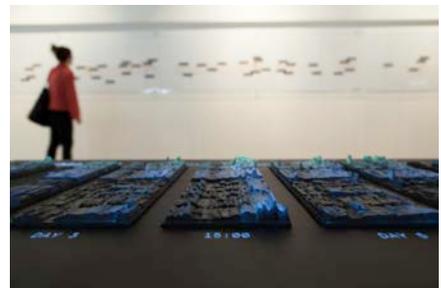


Fig. 2. emoto data sculpture and sentigraph. (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

the drama of the sporting events. The event used real-time data, so in order to test the concepts, model and predict outcomes prior to the Olympics a similar event and dataset was used, in this case the Golf Masters. Through the Golf event, the team observed the ebb and flow of day-by-day events, and questioned anomalies in order to understand what was occurring [11]. On close examination, the most revealing patterns were discovered to be at the level of individual golfers. This led to a focus in *emoto* on individual athletes – and by extension individual sporting disciplines, nations and other topics related to the Games.

c) Technical design

Working with high volumes and data streams posed the biggest technical challenge. In order to address this, a custom infrastructure was developed to ingest the Twitter Streaming API to capture tweets related to the Olympics, and channels were set up for a predefined list of topics. The sentiment analysis software – *Salience Engine* by Lexalytics – annotated each tweet with a sentiment score, which was mapped to equal sized bins on a diverging scale (from -6 ‘very negative,’ to +6 ‘very positive’) [12]. The tweets were then categorised according to topics of interest (disciplines, athletes, countries), and supplied to the viewer in real time, through the online visualisation [13].

d) Visual language and online interface

An aim of the project was to create a meaningful mapping between data and visual form. *emoto* used origami-like shapes (Fig. 1), developed to communicate at a glance the combination of positive or negative attention for each topic in that moment



Fig. 3. Design detail from 'Topics' view – displaying the most discussed topics and a sentiment profile for each. (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

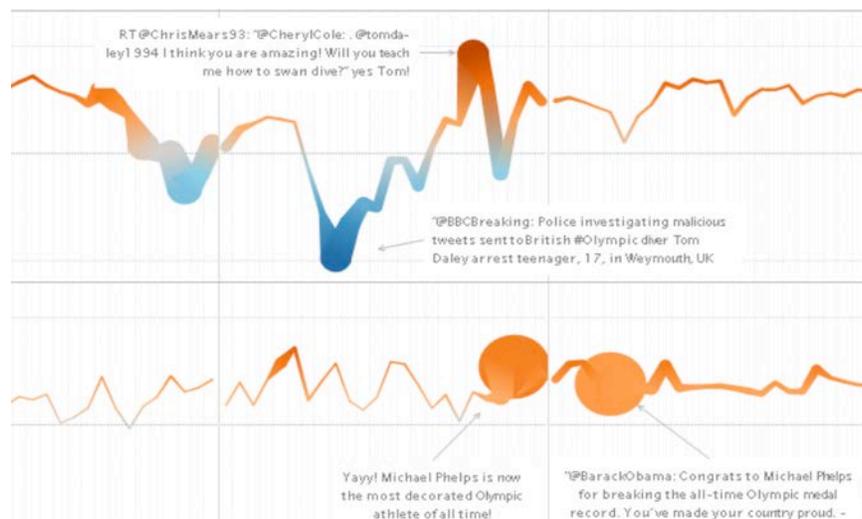


Fig. 4. A custom timeline or ‘sentigraph’ developed through data journalism (detail). (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

[14]. These shapes redraw and scale to the number of messages received, to enable easy comparison of a number of topics in a single view.

A ‘Topics’ view (Fig. 3) showed the big picture, which topics were generating the most attention in that moment, and the sentiment profile of each. A ‘Message Stream’ view showed the ephemeral and anecdotal content of individual Tweets making up the big patterns [15]. The viewer could switch between these scales, or view them simultaneously in ‘Hybrid’ mode (Fig. 5). The visualisation also presented structured insight on trends and patterns, such as the development of each topic over time, and an overview for each day.

e) Data journalism

During the Games, the team investigated the data record through ‘data-driven journalism’ [16], giving the opportunity for deeper insight and reflection, as a complement to the fast-moving and ever-changing online visualisation, and shared the findings on the project blog and with interested national teams. Stories were developed and told through charts and graphs (Fig. 4), accompanied by written commentary.

f) Physical data sculpture

Once the Olympics had finished, all of the accumulated data was gathered and transformed into an interactive data sculpture (Fig. 6). Taking the plots from the early experiments as inspiration, the data was milled into 17 objects laid side by side, each representing one day of the Games. Their surface was a relief heatmap, with the height of the sculpture at any point indicating the emotional highs and lows over time. Overlay projections highlighted individual stories and some key themes that emerged

during the Games, with visitors able to scroll through them using a control knob, and investigate which stories were trending at a specific time on a particular day. Another dimension to the gallery installation was a physical sentigraph – a 9-metre long wall piece and timeline depicting a high-level overview of average emotional response to London 2012, the story of the Games as seen by *emoto*.



Fig. 5. Design 'Hybrid' view. (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

Observations and Discussion - Data and Knowledge in Today's Media Landscape

i) Social Data: Properties and Limits

emoto gives expression to emotion as it is performed online, and captured and codified as emotion data. In relation to discussions around epistemology and the ‘truth’ of data visualisation, therefore, it is important to qualify the properties and limits of the dataset. Immediately apparent were the limitations and bias of the dataset being working with – Twitter – and the challenges for a project of this scale, or any scale, in achieving accurate sentiment analysis [17].

Twitter does not represent all people or nations. Access to the Internet is unequal and with social media there is no global space, there is a patchwork of territories, platforms and connectivity. Our visualisation was as ‘global’ –



Fig. 6. *emoto* data sculpture with overlay projection. (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

limited and partial — as Twitter (Fig. 7).

The original intent was to integrate social data from around the world (Twitter, Weibo, Orkut, etc.) and support six widely used languages. But this proved beyond the resources and scope of an art project of this scale, and, following an enforced change in technical partner (detailed in this report), *emoto* focused solely on Twitter and English. The majority of users of the online visualisation — more than 65% — were from four English-speaking countries — UK, US, Canada and Australia. Such a limited perspective does not represent “what the world thinks.”

Other bias is less trivial to quantify. An evaluation of the accuracy of the sentiment analysis, for example, was beyond the scope of the *emoto* project.

ii) Social Data: Vision and Insight

It is necessary to consider how the practicality of working with big data favours some kinds of insight over others, and forces a particular “model of intelligibility” [18]. Pointing to debates in media theory, we see that a visualisation such as *emoto* reveals as much about the medium (Twitter) as about the message (commentary and opinion on London 2012). As such, the findings of this project contribute to discussions around the dimensions of new and traditional media, and how the two may interact. Furthermore, *emoto* highlights commonplace dimensions of social data, and introduces some novel dimensions of its own.

In some cases, *emoto* revealed similar stories to those seen in broadcast media, but with greater immediacy, and greater scope to see patterns emerge. Viewers

could observe the mounting anxiety and then joy in Britain around the wait for Team GB Golds. Other events observed were native to social media, such as the ‘trolling’ of athletes [19]. Both old and new media at times became the story, such as in controversies surrounding NBC and Twitter. Medium and message were sometimes confused, such as when the BBC received a negative score when it covered a negative story. When the *emoto* team observed anomalies, they were investigated to see if they were actual events or artifacts of the visualisation [20].

The ‘views from above’ are instances of a statistical viewpoint that comes from representing social phenomena in a quantified way, through data analysis and computational tools [21]. Similarly the ‘ephemeral and anecdotal’ is a common frame of reference in social media, wherever there is a stream of social updates pushed out immediately to subscribers.

When captured in digital memory the past is available in new ways. The data from *emoto* is archived in its entirety, making possible data journalism during the Games, and ongoing ad hoc analysis and research of the data after the Games concluded. And yet, there was no option of replay. Within the constraints of time and resource faced, *emoto* was not able to preserve context and hold the data in memory so that a prior moment in the live stream could be simulated and an event be visualised a second time. Here the digital tool places an emphasis on the present moment — the ‘perpetual now’ — due to the difficulty in accessing and representing historic data.

Novel dimensions in *emoto* also

included the option to interrogate and query structured insight, and the use of data journalism to bring new meaning to the data record.

iii) Social Data: Openness and Accessibility

emoto tested the ability of a small artistic team to create its own application for visualising live data, combining bespoke and off the shelf software, something that in the near future might be within reach of everyone. Inseparable from the artistic development was the ‘institutional critique’ involved in negotiating the many layers of permissions involved, firstly to build relationships with London 2012 stakeholders and secure the commission, and secondly around access to the social data, both with a data reseller and Twitter directly. Through this, questions around the openness of social data emerged.

Twitter has become a defacto newswire for everyday events, by making it very easy to share and observe commentary and opinion. Tweets are publicly visible by default, unregistered users can read posts. Crucially, Twitter provides licensing and APIs so third party developers can build on top of it and Tweets can be accessed through many platforms. All of these factors made it an ideal choice for *emoto*, enabling the team to develop its own interface to visualise Twitter data.

emoto had been offered access to augmentations of the ‘Firehose’ — the real-time stream of all (100%) the public tweets — by one of two data resellers with full access. Then Twitter tightened their Terms of Service, and limited third party access, in order to better monitor

and monetise the flow of data [22]. Shortly before launch the reseller informed the team that no data or derived insight obtained from them could be publicly displayed in any form [23]. The project faced the real prospect of cancellation. To work around this, it was necessary to build new infrastructure in just two weeks, and also to use the ‘public API’ – giving access to 1% of the 400 million tweets per day it was theoretically able to access. This highlights a fragility of artworks and, indeed, industries built on the APIs of social media owned by commercial monopolies. And a more general point, that today only the big data companies themselves have complete access to all the data and the means to analyse it [24]. Of note here is Twitter’s willingness to undermine its own carefully cultivated USP and reputation in order to enhance profitability.

Conclusions - Learning and Reflections

This report reflects on the development and release of *emoto*, and highlights multiple questions it raises in relation to current debates on data, visualisation, epistemology, media theory and economics – that are beyond the scope of the report to fully address.

emoto provided an instrument to observe and participate in the online emotional response to London 2012. It was an experimental deployment of online social infrastructure combining human and machine agency, a ‘social machine’ [25]. This was data art [26,27] as a mode of enquiry. It was an instance of ‘visualisation exploration’ [28] – self-directed, and idealistic, seeking to investigate the limits of the possible, and to provoke reflection on emerging trends. Reflection on this practice gave us cause to question how accessible and replicable these tools can be, the way knowledge practices are changing, and what it means to perceive and experience events at the scale – in size and connectivity – of big data.

The Internet made the store of human knowledge and culture in principle available to all. One year after London 2012 were revelations about abuses of big data by national security agencies. In light of this, a clear understanding is needed of the legibility social data affords, and wider access to the tools to make it meaningful.

emoto illustrates a tension and contradiction in the emerging landscape of data, information and knowledge. On

the one hand, it embodies the fascination that the “global village” has become reality, and it is possible to visualise millions of interactions around the world in real-time. And, on the other, it makes clear the limits of this perspective when we consider issues of ownership and control, or look beyond specific territories (geographic, political or economic) to the world as a whole [29].

The project occurred on the cusp of Twitter’s move to a more commercial and closed business model. It brought that into sharp relief, and with it an urgent debate around the openness and accessibility of social data. The ‘new scientific instrument’ of the social web is arguably one of the greatest resources of our times. *emoto* provides a timely warning that this instrument is not publicly owned, or publicly accountable, it can be switched off on a whim, and its full power is available to only a few.

To conclude, artworks such as *emoto* investigate the potential for data visualisation to enhance insight and experience around emerging social phenomena. Data visualisation combines visual aesthetics and interface design with understanding of data science, and is a practice that strives to be ‘true’ to its materials of data and code. This report suggests this ‘truth’ needs to be qualified by consideration of multiple factors, including: i) the properties and limits of the dataset; ii) the model of intelligibility involved; and iii) the openness and accessibility of the data. There is now understanding within the academic visualisation community of the necessity to communicate confidence levels along with a visualisation [30]. boyd & Crawford argue that any big dataset has limitations and flaws, and the outcome of data sampling and cleaning depends

on the attributes the analyst has access to and makes use of. Further, there is always interpretation, in any attempt to understand what data means, and hence big data analytics can never present a simple ‘objective’ truth [31].

Areas For Further Enquiry

This report suggests a need for further research around effective ways to communicate levels of uncertainty, and the nature of uncertainty, in data analysis and visualisation, in view of challenges in making sense of an unprecedented upsurge in the amount and complexity of data, and also the *a priori* understanding that such data is always partial and limited.

It also highlights a requirement for ongoing debate around the openness and accessibility of social data in light of the economic model of today’s social networking sites.

In *emoto*, evaluation was through participant observation and reflection by the project team during development and deployment. Only a very early stage user experience study of the online visualisation was possible within the scope of the project. Future iterations would benefit from more extensive study of user experience, evaluation of the accuracy of the visualisation, and further documentation and evaluation of the design decisions and process [32].

Ad hoc analysis and research of the data has continued after the Games concluded. The collected data was supplied to the Data Science Laboratory of the ISI Foundation. This has been used in the development of the *EMOTO Topic Explorer*, an exploratory tool designed to expose the topical and temporal dimensions of Twitter conversations [33].

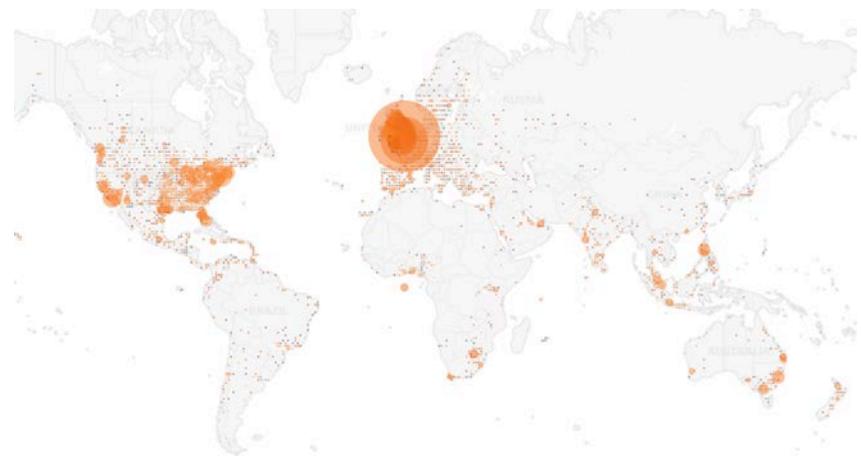


Fig. 7. The location of geo-coded tweets captured by *emoto*, highlighting a bias towards the UK and North America. (© Moritz Stefaner, Drew Hemment, Studio NAND. Some Rights Reserved.)

Acknowledgements

emoto (emoto2012.org) was by Moritz Stefaner, Drew Hemment and Studio NAND. A FutureEverything project for the London 2012 Festival and the Cultural Olympiad. Funded by Arts Council England and WE PLAY/Legacy Trust UK. Infrastructure design and development by Gerrit Kaiser. Citizen journalism by Andy Miah. Evaluation and blogging by Ege Sezen. Project management by Leon Seth and Nick Lawrenson. Communications by Jo Williams, Anita Morris Associates. Sentiment Analysis Partner: Lexalytics. Manufacturing Partner: Tischlerei Bächer. Thanks to Debbi Lander, Creative Programmer for the London 2012 Cultural Olympiad (Northwest). *emoto* was produced by FutureEverything and Studio NAND.

The *emoto* online visualisation was exhibited within London 2012 Festival, 27 July - 9 September 2012. The *emoto* data sculpture was exhibited at *WePlay*, London 2012 Festival/Cultural Olympiad, 7 - 9 September 2012, and *Information in Style*, CAFA Art Museum/Beijing Design Week, Beijing, China, 20 September - 15 October 2013.

Moritz Stefaner was lead designer on *emoto*. He works as a "truth and beauty operator" on the crossroads of data visualization, information aesthetics and user interface design. Stefaner is especially interested in the visualization of large-scale human activity.

Dr. Drew Hemment initiated the *emoto* project and was a member of the artistic team. He is Founder and CEO of FutureEverything, and a Dundee Fellow, Reader at Duncan of Jordanstone College of Art and Design, University of Dundee.

Studio NAND, consisting of Stephan Thiel, Steffen Fiedler and Jonas Loh, were involved in all aspects of creative development, and had a lead role on design of the physical data sculpture. Studio NAND is a multi-disciplinary design practice based in Berlin, practising and researching at the intersection of design, science and technology.

FutureEverything is an art and innovation organisation founded in 1995 which investigates various dimensions of today's digital culture by commissioning and participating in art and design experiments.

Thanks to Moritz Stefaner, Stephan Thiel and Mel Woods for contributions to this report.

Images licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

References and Notes

1. D. Bollier, *The Promise and Peril of Big Data* (Washington, DC: The Aspen Institute, 2010).
2. D. Fisher, R. DeLine, M. Czerwinski, S. Drucker, "Interactions with Big Data Analytics." *Interactions*, Volume 19 Issue 3 (2012).
3. S. K. Card, J. D. MacKinlay, B. Shneiderman, eds., *Readings in Information Visualization: Using Vision to Think* (Morgan Kaufmann Publishers, 1999).
4. C. Ware, *Information Visualization: Perception for Design* (Morgan Kaufman, 2004).
5. D. Hubbard, *Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities* (John Wiley & Sons, 2011).
6. Related works include: Jonathan Harris and Sep Kamvar, *We Feel Fine* (2006) and Stamen, *MTV: Live Twitter Visualizations* (MTV Video Music Awards, 2009).
7. London 2012 was dubbed 'the social media Games' in the popular press. See for example <<http://www.bbc.co.uk/news/technology-18889406>> (accessed 30 June 2013).
8. A. Balfour, *London 2012 Olympics and Paralympic Games-time Digital Report*. (LOCOG, 2013), <<http://www.slideshare.net/balf/london-2012com-olympic-games-digital-round-up-13-august-2012>>.
9. Indeed, during London 2012 social media became integral to the television coverage of the Games.
10. In the early stages, the *emoto* team made the case for making Olympics data freely available using open standards to the New Media team at the London Organising Committee of the Olympic and Paralympic Games (LOCOG). LOCOG did release data through the London Datastore, but in the end *emoto* used data from a single proprietary service, Twitter.
11. In this context, the team developed a chart type it called "sentigraph," which encodes sentiment in color and vertical position of a line, and at the same time the number of messages in line strength. See early process heats maps documented at <<http://www.nand.io/visualisation/emoto>> (accessed 30 June 2013).
12. *emoto* used the Lexalytics *Salience Engine*, which uses various text processing, natural language processing and text analytics technologies for sentiment analysis <<http://www.lexalytics.com/technical-info/salience-engine-for-text-analysis>> (accessed 30 June 2013).
13. The project was developed by building on open source code libraries such as Processing.org and D3.js. The infrastructure consisted in a Node.js java based server, sentiment analysis tools developed by a partner, Lexalytics, Redis for fast and flexible storage, and HTML5 for the frontend.
14. One early visual concept consisted in particles representing individual tweets that would coalesce and form humanoid figures representing each discipline. But this approach would work only in a limited set of circumstances – for example, it was not effective with too high or too low numbers of tweets.
15. In *emoto* this featured an "inverse parallax" perspective - where tweets that are "more important" (cited or referred to more often) are larger and float above other tweets, while moving more slowly across the screen.
16. M. Lorenz, "Data driven journalism: What is there to learn?" *IJ-7 Innovation Journalism Conference*, Stanford, CA (2010).
17. A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *7th International Conference on Language Resources and Evaluation LREC10*, Valletta, Malta (2010).
18. D. boyd and K. Crawford, "Six Provocations for Big Data" *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society* (2011).
19. In one case, a British diver, Tom Daley, was sent an abusive tweet from a 17 year old, who was later arrested. This threw his sentiment profile from positive to negative, then back to positive, as fans expressed first anger then support.
20. One anomalous event was the Irish triangle. It was observed that Ireland only had one single dominant emotion, "quite positive," shown as a one large triangle. The explanation lay in the huge impact of teen celebrities on the web – the response captured on Twitter was dominated by one message retweeted over 25,000 times, by Niall Horan, member of the boyband "One direction."
21. L. Manovich, "Trending: The Promises and the Challenges of Big Social Data" in M. K. Gold, ed., *Debates in the Digital Humanities* (The University of Minnesota Press, 2012).
22. Twitter tightened their Terms of Service, and limited third party access <<https://dev.twitter.com/blog/changes-coming-to-twitter-api>> (accessed 30 June 2013).
23. Upon notification, the *emoto* team asked Twitter for an exception for a non-commercial, one-off artwork, pointing out that the project aimed for a high level of abstraction, and offering to build in a Twitter layer. The team explored the legal definition of data, and made a case that the installations proposed (one a soundscape, one milled onto solid material) are far removed from the data stream and do not "distribute data" in a way that would enable the original information to be reconstructed. The *emoto* team spoke to Twitter in the USA, in Germany, to their Olympics lead in the UK, but to no avail.
24. Manovich [21].
25. N. Shadbolt, D. Smith, E. Simperl, M. Van Kleek, Y. Yang, W. Hall, "Towards a Classification Framework for Social Machines", *WWW 2013 Conference*, Rio, Brazil Vinilson (2013).
26. G. Judelman, "Aesthetics and Inspiration for Visualization Design: Bridging the Gap between Art and Science," *International Conference on Information Visualisation*, IV 04. IEEE (2004) pp. 245-250.
27. F. B. Viegas and M. Wattenberg, "Artistic Data Visualization: Beyond Visual Analytic," *Lecture Notes in Computer Science*, 4564 (2007).
28. A. V. Moere, and H. Purchase, "On the role of design in information visualization," *Information Visualization* 10: 356 (2011).
29. This observation was made by Moritz Stefaner.
30. K. Brodrie, R. A. Osorio, A. Lopes, "A Review of Uncertainty in Data Visualization," in J. Dill, R. Earnshaw, D. Kasik, J. Vince and P. C. Wong eds. *Expanding the Frontiers of Visual Analytics and Visualization* (London: Springer Verlag, 2012).
31. boyd and Crawford [18].
32. Evaluation Study: *Emoto Project*, a report by Ege Sezen (2012) for the HighWire Doctoral Training Centre at Lancaster University. Conducted from July 27-Aug 12 2012, this focused on questions around augmentations to the experience of the sporting events, based on analysis of interviews conducted with users following the Games, and by monitoring online comments, web articles, shares and messages.
33. Twitter Topic Explorer <<http://www.datainterfaces.org/2013/06/twitter-topic-explorer>> (accessed 30 June 2013).