Observing Social Machines Part 2: How to Observe?

David De Roure
Oxford e-Research Centre
University of Oxford
Oxford, United Kingdom
david.deroure@oerc.ox.ac.uk

Clare Hooper
IT Innovation Centre
University of Southampton
Southampton, United Kingdom
cjh@it-innovation.soton.ac.uk

Kevin Page
Oxford e-Research Centre
University of Oxford
Oxford, United Kingdom
kevin.page@oerc.ox.ac.uk

Ségolène Tarte
Oxford e-Research Centre
University of Oxford
Oxford, United Kingdom
segolene.tarte@oerc.ox.ac.uk

Pip Willcox
Bodleian Libraries
University of Oxford
Oxford, United Kingdom
pip.willcox@bodleian.ox.ac.uk

ABSTRACT
Social machines are increasingly attracting study. In our paper "*Observing Social Machines Part 1: what to observe*?" we scoped the task of observing them. Several exercises that have followed have further informed our thinking and methodologies. Here, in Part 2, we reflect on *how to observe*? We promote a variety of methodologies that transcend the study of individual social machines, recognizing social machines as co-constituted processes within the evolving Web, and the intersection of social machines with the physical world through the Internet of Things. Our approaches emphasize the importance of sociality and human-centric perspectives.

Categories and Subject Descriptors
H.4.m. **[Information systems applications]**: Miscellaneous.

General Terms
Theory; Human Factors; Design.

Keywords
Social Machines; Web Observatories; Internet of Things.

1. INTRODUCTION
The concept of social machines, as initially defined by Berners-Lee [1], provides us with a perspective that focuses on the purposeful human-machine collectives of the social web. Often exemplified by Wikipedia and citizen science, social machines are attracting increasing study, offering a useful new approach to the analysis and design of socio-technical systems.

WebSci '15, June 28 - July 01, 2015, Oxford, United Kingdom

ACM 978-1-4503-3672-7/15/06…$15.00
DOI: http://dx.doi.org/10.1145/2786451.2786475

Many examples of individual social machines have received attention in the Web Science community, including Wikipedia [2], the Mechanical Turk [3], and Tumblr [4]. The methodologies for studying social machines are also discussed, for example in the development of Web Observatories [5-6].

We believe that the notion of the Web as an *ecosystem of interacting social machines* is a useful analytic and design paradigm, providing an alternative conceptualization to models such as crowdsourcing, collective intelligence, and human computation [7]. In other words, we see a community of living, hybrid organisms, rather than a set of machines which happen to have humans amongst their components. Social machines often change in behaviour (the *perpetual beta* principle of the social web) and their successes and failures inform the design and construction of their offspring and successors.

At the outset of our study of social machines we considered the scope of the observation task [8]. While one instinctive approach has been to identify and classify individual social machines as encountered 'in the wild', we argued the need to look at social machines interacting within their ecosystem and throughout their life course. We also felt it important to take a broad definition, so as not to blinker the observation, and to embrace the cyber-physical as well as the socio-technical.

In [9] we considered this multi-faceted space in terms of *trajectories* through (multiple, potential) social machines. Each trajectory is a line of investigation, scoped by a purpose, intersecting both technical and social elements of the ecosystem. While a trajectory, or even a set of trajectories, is insufficient to define a particular social machine fully, we argued that this is a more productive approach to the investigation of social machines than attempting to defend exemplars that are entirely demarcated by internal and external boundaries. Instead, trajectories provide a framework for observing and comparing aspects of social machines for analysis. That we plot a trajectory according to a (social) purpose forces us to recognize that there may be multiple complementary, or even potentially conflicting, motivations for actors within a social machine observatory.

The intervening period has seen multiple observation exercises, conducted by the authors and many others, and this has developed our ideas about the methodology of observation. In this paper we describe and reflect upon the development in our thinking, as we

discuss *how to observe* social machines. We hope this also provides some useful ideas and provocations for those turning to the study of social machines.

The next section relates our observation experiences. We reflect upon these and offer some summary recommendations in Section 3, with closing thoughts about social machines, sociality, and Web Science in Section 4.

2. EXPERIENCES

2.1 Web Observatories

Every social machine can generate useful data, and sharing the data (with appropriate ethical consent), and especially also the analytic methods, will build capacity and insight in the social machine community—or, more broadly, the computational social science and Web Science communities.

This is the ambition of the "globally distributed and collaborative Web Observatory" [6], through which the web "is observable at scale across space and time". The Web Observatory envisions a global data resource and an open analytics environment to nurture Web Science research, providing a distributed archive of data and activity on the Web, as well as methodologies and tools to explore its evolution. Rather than a single independent store, it is seen as a virtual construction over the multitude of emerging observatory activities.

There is a strong argument that Web Science projects should record their data in Observatories to support re-use and also research which is replayable, reconstructable, repurposeable, and ultimately reproducible [10]. But what primary data do we need in order to observe social machines, and how is it to be collected? Our experiences, accounted below, suggest multiple kinds of data, supporting mixed methods.

2.2 Stethoscopes

It is not clear that web analytics are sufficient for comprehensive analysis of social machines. Take, for example, the scientific discoveries that have come about through the Galaxy Zoo [11] citizen science project, such as the 'green peas' galaxies [12]. While part of the Galaxy Zoo social machine involves a crowdsourced data analysis process, citizen scientists also interact by discussion forums and engage with the science itself—they participate in social processes, and indeed they have co-created them. Hence studying Galaxy Zoo requires consideration of both data reduction and discourse.

Galazy Zoo is part of the Zooniverse, the common platform on which new citizen science projects are regularly delivered. It is, therefore, a case study in multiple social machines, and enables us to investigate the factors that affect user engagement across multiple projects. As an example, a quantitative analysis [13] of ten citizen science projects hosted on the Zooniverse platform examined the level of participation of users in Zooniverse discussion forums in relation to their contributions toward the completion of scientific tasks. It used a data set of over 50 million activity records and more than 250 000 users.

Motivated by our desire to study trajectories through multiple social machines, we have adopted a slightly different observation methodology which we describe as the *stethoscope metaphor*. Modelled around a doctor's bag of diagnostic medical instruments, this enables researchers to engage with an individual social machine, perhaps to assess its state of health. The instruments therein are not just probes for remote data collection by telemetry, but methodological templates for mixed methods analysis of the machine in the wild. We can deploy such probes around the ecosystem of interacting machines, avoiding assumptions of social machine boundaries.

In [14] we described the stethoscopes available in the Zooniverse citizen science environment, considering how the available telemetry can inform the requirements of a wider Web Observatory. The focus on a purpose—here a specific system design for a citizen science platform—can, if sufficiently instrumented, provide as powerful an environment for observations as a generic web observatory that has been conceived as such.

While this suggests that every social machine is its own observatory, ultimately every social machine is also its own laboratory, since it provides the apparatus to conduct configuration experiments such as A-B testing with independent citizen cohorts.

Combining data analytics from the Zooniverse platform with consideration of purpose, Segal *et al*. [15] demonstrate the measurable utility of a 4-step iterative method derived from an identified trajectory through the citizen science social machine: (i) qualitative understanding of motivations or purpose of participants in the platform, combined with (ii) quantitative profiling of populations, both informing (iii) a modification to the platform operation that addresses the motivational factors for the identified cohort, which is then (iv) evaluated through quantitative and qualitative methods.

Looking at logs tells us who and not why, yet the subjects are available for survey. It is also the case that citizens are engaged with multiple social machines, so a citizen-centric perspective reveals more about the ecosystem than examining an individual machine.

2.3 Plot and prosopography

In order to emphasize the importance of considering sociality alongside the prevalent mechanistic approaches to social machine analytics, we have conducted work adopting the analogy of stories as a provocative methodological example [16]. This analogy identifies stories within and about social machines transversally, from analytical and design perspectives. The success of a social machine's narrative has been used to evaluate its health in three respects: evaluating the potential and realization of its storytelling assesses its sociality; evaluating its interactivity and reactivity assesses its sustainability; and evaluating the collaboration between authors assesses its emergence.

This methodology uses the characteristics of narratives to uncover the social machine as story. Particularly as we observe interactions between and through social machines and their ecosystem, the need to specify the type of narrative to observe becomes apparent. By systematically gathering elements from biographies of various types of unitary and composite entities within and across social machines, we can create homogenous data from heterogeneous sources. Prosopography is a method, used in particular by historians, that allows systematic study of a collection of biographies, be they of persons, artefacts, infrastructures, or groups of these. Thus we develop a *prosopographical* methodology for observing social machines in which we investigate their common characteristics by means of a collective study of their biographies [17].

Our work on narratives and biographies has highlighted the importance of understanding the vantage point of the observer, or equally the narrative point of view, within the ecosystem. Some of the observers mentioned in this paper are depicted in Figure 1.

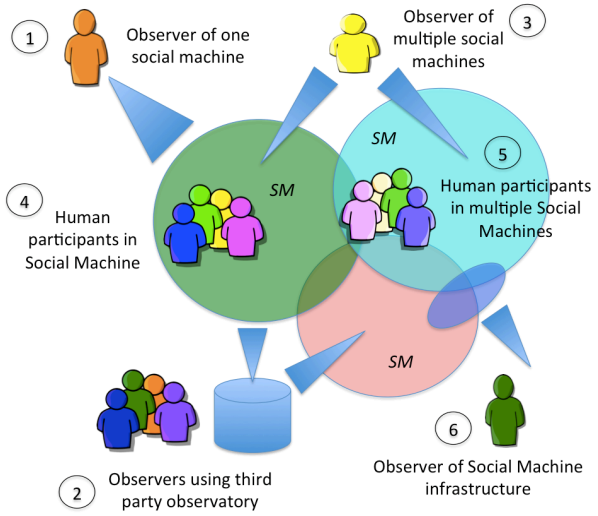

Figure 1. Six illustrative points of view of observers or narrators.

While trajectories and stethoscopes enable us to consider social machines across their boundaries, the plot and prosopography approaches enable us to consider their interleaved lives and to explore the temporal dimension. Our experience is that no single method gives a comprehensive analysis; understanding the ecosystem will require many observers, and this is facilitated by this approach.

2.4 Scenarios

Another strand to our methodological thinking, again transcending individual machines, has been the adoption of motivating scenarios. Our previous paper [8] considered two motivating scenarios which continue to be salient:

1. In the "machines of spam" we described a social website attacked by a spam social machine and modified by incorporating a blacklisting social machine to defend it. This narrative, featuring social machine as first class citizen, continues to be a persuasive example of social machines thinking which emphasizes interaction and evolution. It also exemplifies trajectories as a useful framework to tease apart the different observable elements.
2. In "the befriending of Raspberry Pi" we described a cyber-physical system demonstrating accidental assembly—each social machine behaving according to its purpose, but the combination unanticipated. This scenario has grown to be more important than ever, as the electronics industry risks delivering the devices of the Internet of Things with a benign world assumption that takes little or no account of accidental emergent behaviour nor malicious exploitation.

An additional scenario, *scholarly social machines* [18], has proven an effective subject of study, engagement and discussion. Scholarship itself is an in-the-wild experiment in social machines, an ecosystem in which we all participate—by communicating knowledge through conference papers, for example. The process of research has long involved a sense-making network of scholars and citizens exchanging scholarly writing, but we are seeing the co-constitution of an increasingly democratized scholarly communications system, intermediated by new social machines, constructing a sense-making network of humans and computers.

As libraries and publishers re-invent themselves for digitized and 'born digital' content, we see a plethora of new websites emerging, from repositories for data sharing to new models of peer review—a vital ecosystem in which new experiments are conducted, and social machines come and go. Looking ahead we see a set of concerns in the face of increasing automation. Who gains credit and who owns the intellectual property generated when research runs automatically? Who is liable for any damage that arises? What are the implications of unintended or accidental assembly of research methods and outcomes? What are the consequences of automated research that occurs at very high speed, possibly speculatively, without human intervention? Will methodologies be locked into our knowledge infrastructure without critical or creative thought?

In our latest work we look at the social machines of the music sphere, in an industry that has gone digital (almost) end-to-end. The traditional chain of recording, production, distribution, and consumption is being reconstructed through democratization: consumer as producer and distributor of new material via the Web. Apart from the individual social machines that we might identify, like the websites for sharing and annotation of recordings [19], we can look at the overall production-distribution-consumption ecosystem as a social machine with humans empowered in creativity.

As a case study this touches on democratization, empowerment, and automation, themes which occur in parallel in scholarly social machines and elsewhere. It has also highlighted human creativity by inspiring a discussion about the music of social machines, as demonstrated in the work of Julie Freeman who has visualized Zooniverse data [20].

3. REFLECTIONS

The study of individual social machines continues to be important. Faced with the necessity of defining a field of study, researchers have tended towards social machines that are popular and successful. However, there are other social machines in the broader ecosystem that must not be neglected. This includes failed social machines, those deployed or repurposed with malicious intent, or those that simply 'go wrong'. The ecosystem is rich and diverse.

However, to understand fully the theory and practice of social machines we must also look beyond the individually demarcated engines, towards their context and their interactions. Our methods are a step towards an ecosystem approach because they avoid assumptions about the boundaries of individual machines, and fully recognize that the ecosystem evolves—that co-constituted behaviour is emergent and might not be as intended, and social machines may be transient or persistent.

Throughout our work we have been concerned that researchers seem to fall into assumptions that hide the essential sociality of social machines. Crucially we must recognize that a human has a choice of technologies and social machines, and technology now comes with social machines built in, or at least preconfigured. Humans are empowered through choice and through influence on design: they appropriate the technology and interpret it creatively and subversively.

It follows that *how* we observe should also be human-centric: not just weblogs and examining the digital exhaust of the social machine, but engaging with the humans who are part of its operation and design. The proverbial explorer in the jungle counting the legs of an elephant has less opportunity to speak to

the elephant directly or to its designer: in the social machines ecosystem we can—and should.

The narrative and prosopographical approaches are proving fruitful in providing a new lens which is explicitly aware of the viewpoint and time dimension of the narrator, acknowledging that there is rarely or never an omniscient narrator (even though many papers take that voice). Biographies can be of people or indeed any part of a machine, and interleaved stories offer rich analysis and insight. Significantly, our stethoscope and biography methodologies allow us to consider the social machine itself as the living organism.

Conducting observations, analyses and experiments in the wild has raised several ethical issues and we note the need for a broader discussion. This is only partly about responsible use of data, as we also need to consider the purpose of social machines, as in the framework for responsible innovation which highlights process, product, and purpose [21]. While we are asserting that humans are autonomous and have agency and creativity, there is a discussion still to be had as to whether social machines exhibit such aspects. Reminiscent of Asimov's laws of robotics [22], it might be informative to consider the 'roboethical' principles that have been established in the field of robotics, arguably another kind of social machine [23].

In summary, our reflections suggest the following advice to the social machines researcher:

1. **Beyond Twitte**r. Explore more broadly than the large and sustained social machines characterized by successful platforms;
2. **Multiple machines**. Avoid overprescribing the boundaries of a social machine, but rather consider all its interactions, throughout its life course;
3. **Talk to humans**. Ultimately we must be able to take a human-centric view within the ecosystem, not just a social machine perspective—the citizens are us;
4. **Listen to stories** about social machines, compare them, and question alleged omniscience—observing the observers. This emphasizes sociality, viewpoint and time.

4. WEB SCIENCE AND SOCIAL MACHINES

Tim Berners-Lee provides a popular definition of social machines in the phrase "...abstract social machines on the Web: processes in which the people do the creative work and the machine does the administration" [1]. This sets a vital principle for socio-technical systems at scale and under automation, acknowledging a distinct role for humans—we see it as a principle of empowerment. However a more complete definition follows in the same chapter: "The stage is set for an evolutionary growth of new social engines. The ability to create new forms of social process would be given to the world at large, and development would be rapid." This second part better captures the spirit of our work, encompassing social process, democratization, and acceleration.

While arguments as to what is and is not a social machine are set to continue for some time, we believe people could do worse than look back to this starting point. In contrast to Web 2.0, social machines give us *user generated process* as well as *user generated content*, and there is an emphasis too on automation and computational process. This leads to the co-created behaviour of the human-machine collective that characterizes social machines, and in turn demands citizen-centred methods.

In 2008, Halford *et al* [24] responded to the 'clarion call' of Web Science with a manifesto drawn from social theory. Our methodological experience in studying social machines has a strong resonance with the opening concepts:

1. The **co-constitution** of technology and society; i.e. the ways in which people and the Web make each other.
2. The importance of **heterogeneous actors**, human and non-human, as these are constituted in the networks that produce the Web.
3. The significance of **performativity**, whereby people build 'the web' moment by moment, so it is an unfolding, enacted practice.

These are equally applicable to social machines, which we also see as a temporarily stabilized, and eminently revisable, set of socio-technical relations.

While it has been easy to see social machines as a subset of the Web ("on the Web", according to Berner-Lee's definition), the suggestion here is stronger: Web as social machine. Each social machine is perhaps then identified by a set of co-constituted processes, social and automated, in the evolving Web. This notion of a very dynamic Web, fundamentally in flux, also underpinned previous work around Web Science methods [25], where the emphasis was on flow of data.

Our work continues to embrace the cyber-physical—where the social machine intersects with the physical world through sensors and tangible objects. This intersection could be seen to occur in the Internet of Things. It follows that as Internet of Things develops as an area of study, we already have a set of concepts and methodologies for its design, development, and analysis. This social perspective is poorly acknowledged at present, perhaps again due to the emphasis on the technological artefact than on its use and its interpretative flexibility. It seems likely that the methodological lessons of social machines and Web Science can be usefully applied as we move forward into deeper cyber-physical deployment.

Hence we see the social machines methodology to be of broad applicability on the Web, in the Web, and in the future of the Web, even to be paradigmatic in the Kuhnian sense [26]. Social machines can be seen as a paradigm of Web Science.

In part 3 we will reflect on the evolution of our thinking after this paper, as we embark on the next chapter in the biography of social machines research.

5. ACKNOWLEDGMENTS

The work of the Oxford authors is supported by SOCIAM: The Theory and Practice of Social Machines, funded by the UK Engineering and Physical Sciences Research Council (EPSRC), under grant number EP/J017728/1, also FAST EP/L019981/1, and Smart Society: Hybrid and Diversity-Aware Collective Adaptive Systems: When People Meet Machines to Build a Smarter Society, funded under the European Commission FP7-ICT FET Proactive Initiative: Fundamentals of Collective Adaptive Systems (FOCAS), Project Reference 600854. The work of the Southampton author is supported by the EU FP7 EINS project under grant agreement No. 2880821. We are grateful to our colleagues in these projects, the fruits of these collaborations being the foundation for much of what is summarized in this paper.

6. REFERENCES

[1] Berners-Lee, T. and Fischetti, M. 1999. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper, San Francisco.

[2] Hale, S.A. 2014. Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM conference on Web Science* (WebSci '14). ACM, New York, NY, USA, 99-108. DOI=http://doi.acm.org/10.1145/2615569.2615684

[3] Noy, N.F., Mortensen, J., Musen, M. and Alexander, P.R. 2013. Mechanical turk as an ontology engineer?: using microtasks as a component of an ontology-engineering workflow. In *Proceedings of the 5th Annual ACM Web Science Conference* (WebSci '13). ACM, New York, NY, USA, 262-271. DOI=http://doi.acm.org/10.1145/2464464.2464482

[4] Xu, J., Compton, R., Lu, T. and Allen, D. 2014. Rolling through tumblr: characterizing behavioral patterns of the microblogging platform. In *Proceedings of the 2014 ACM conference on Web science* (WebSci '14). ACM, New York, NY, USA, 13-22. DOI=http://doi.acm.org/10.1145/2615569.2615694

[5] Hall, W. 2014. Observing the web. In *Proceedings of the 2014 ACM conference on Web Science* (WebSci '14). ACM, New York, NY, USA, 1. DOI=http://doi.acm.org/10.1145/2615569.2618143

[6] Tiropanis, T., Hall, W., Shadbolt, N., De Roure, D., Contractor, N. and Hendler, J. 2013. The Web Science Observatory, *IEEE Intelligent Systems* **28**(2) pp 100–104. DOI= http://dx.doi.org/10.1109/MIS.2013.50

[7] Shadbolt, N., Smith, D., Simperl, E., Van Kleek, M.,Yang, Y. and Hall, W. 2013. *Towards a classification framework for social machines.* In *Proceedings of SOCM2013: The Theory and Practice of Social Machines*, Rio de Janeiro, Brazil, International World Wide Web Conferences Steering Committee, pp. 905-912.

[8] De Roure, D., Hooper, C., Meredith-Lobay, M., Page, K., Tarte, S., Cruickshank, D. and De Roure, C. 2013. Observing Social Machines Part 1: what to observe? SOCM2013: The Theory and Practice of Social Machines, Rio de Janeiro, Brazil, International World Wide Web Conferences Steering Committee, pp. 901-904.

[9] Page, K.R. and De Roure, D. 2013. Trajectories through Social Machines. *Building Web Observatories Workshop*, ACM Web Science 2013, Paris, 2013. Available on: https://sites.google.com/site/bwebobs13/bwebobs13_Trajectories%20through%20Social%20Machines.pdf

[10] De Roure, D. Replacing the Paper: The Twelve Rs of the e-Research Record, SciLogs blog. Available on: http://www.scilogs.com/eresearch/replacing-the-paper-the-twelve-rs-of-the-e-research-record/ (visited 29 May 2015).

[11] Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., Murray P. and Vandenberg, J. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society* **389** (2008), 1179–1189.

[12] Cardamone, C., Schawinski, K., Sarzi, M., Bamford, S.P., Bennert, N., Urry, C.M., Lintott, C., et al. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies, *MNRAS* **399** (2009), 1191–1205.

[13] Luczak-Rösch, M., Tinati, R., Simperl, E., Van Kleek, M., Shadbolt, N. and Simpson, R. 2014. Why won't aliens talk to us? Content and community dynamics in online citizen science. In Proceedings of *Eighth International AAAI Conference on Weblogs and Social Media*, Ann Arbor, US, 1-4 Jun 2014. 10pp.

[14] Simpson, R., Page, K.R. and De Roure, D. 2014. Zooniverse: observing the world's largest citizen science platform. In *Proceedings of the companion publication of the 23rd international conference on World Wide Web*, 1049-1054. DOI=http://dx.doi.org/10.1145/2567948.2579215

[15] Segal, A., Simpson, R.J., Gal, Y., Homsy, V., Hartswood, M., Page, K.R. and Jirotka, M. 2015. Improving Productivity in Citizen Science through Controlled Intervention. In *Proc. Web Science track, WWW 2015*, Florence, Italy.

[16] Tarte, S.M., De Roure, D. and Willcox, P. 2014. Working out the Plot: the Role of Stories in Social Machines. *SOCM2014: The Theory and Practice of Social Machines,* Seoul, Korea, International World Wide Web Conferences pp. 909–914.

[17] Tarte, S. Willcox, P., Glaser, H. and De Roure, D. 2015. Archetypal Narratives in Social Machines: Approaching Sociality through Prosopography. ACM Web Science 2015. DOI=http://dx.doi.org/10.1145/2786451.2786471

[18] De Roure, D. 2014. The future of scholarly communications. *Insights: the UKSG journal*, 27, (3), 233-238. DOI=http://doi.org/10.1629/2048-7754.171

[19] Yee-King, M., d'Inverno, M., and Noriega, P. 2014. Social machines for education driven by feedback agents. In *Proceedings First International Workshop on the Multiagent Foundations of Social Computing*, AAMAS-2014, Paris, France, May 6 2014.

[20] Karen Eng, *Julie Freeman launches new online artwork, We Need Us*, TED Fellows blog, 5 October 2014. Available at: http://fellowsblog.ted.com/2014/10/05/fellows-at-tedglobal-2014-julie-freeman-launches-new-online-artwork-we-need-us (visited 29 May 2015).

[21] Stilgoe, J., Owen, R. and Macnaghten, P. 2013. Developing a framework for responsible innovation, Research Policy, Volume 42, Issue 9, November 2013, Pages 1568-1580, DOI=http://dx.doi.org/10.1016/j.respol.2013.05.008

[22] Asimov, I. 1942. Runaround. *Astounding Science Fictio*n, New York: Street and Smith Publications, Inc. March 1942. pp 94–103.

[23] EPSRC and AHRC, Principles of robotics. Available at: http://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/ (visited 29 May 2015).

[24] Halford, S., Pope, C. and Carr, L.A. 2010. A manifesto for Web Science. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, US, April 2010.

[25] De Roure, D. and Goble, C. 2010. Anchors in Shifting Sand: the Primacy of Method in the Web of Data. In *Proceedings of WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, North Carolina, US. April 2010.

[26] De Roure, D. 2014. The Emerging Paradigm of Social Machines, Digital Enlightenment Yearbook 2014 227 K. O'Hara et al. (Eds.) IOS Press, 2014. pp 227-234. DOI=https://dx.doi.org/10.3233/978-1-61499-450-3-227