

# DNA: From Search to Observation Revisited

Ian Brown  
Web Science Institute  
University of Southampton  
Southampton, SO17 1BJ, UK  
+44 (0)23 8059 5000  
ian.brown@soton.ac.uk

Lisa Harris  
Web Science Institute  
University of Southampton  
Southampton, SO17 1BJ, UK  
+44 (0)23 8059 5000  
lj.harris@soton.ac.uk

Wendy Hall  
Web Science Institute  
University of Southampton  
Southampton, SO17 1BJ, UK  
+44 (0)23 8059 5000  
wh@soton.ac.uk

## ABSTRACT

In this paper, we describe extensions to the process model first described in the paper “From Search to Observation” based on additional field interview work. This process model forms part of a triad of perspectives under the banner of a methodology known as **DNA**, which looks at structure (**D**efinition), process (**N**ature) and motivations of actors (**A**rchetypes) for Web Observatories (hereafter WO) and more generally the class of Social Machines. We discuss the rationale for the model enhancements, enumerate and summarise the changes and close with an introduction to future work around use of open source tools and languages for implementing and analyzing social machine processes using this model. The additional perspectives we are now considering are an *extensive* revision to the model (which now addresses more than three times the number of factors in the previous model) and hence a revised paper is called for in this space.

## Categories and Subject Descriptors

H.5.3 [Group and Organisation interfaces]:

## General Terms

Design, Human Factors, Standardization, Theory.

## Keywords

Web Science, Social Machines, Web Observatory, Observatory models.

## 1. INTRODUCTION

We have previously argued that, whilst WOs *share* certain architectural features with systems in the Search and Web Analytics space, the inherent nature of “Observation” in the Web Science sense is more than simply a sub-set of these tools. We will not further rehearse this argument, which is outlined in the earlier paper [1], save to summarise that in order to establish the differences in *usage/application* of WOs rather than the *structure* of Observatories [2] (which might physically be largely indistinguishable from other forms of repository or analytical platform) the research started with a discourse/narrative analysis which was performed on the current literature for Observatories

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org). *WebSci '15*, June 28 - July 01, 2015, Oxford,

and *Observatory-shaped objects*.<sup>1</sup> This resulted in an initial vocabulary of 17 processes focused around the implementation/execution of WOs, which we have gone on to validate and expand through a process of individual and focus-group interviews of academic and business users during the period 2012-2015.

## 2. THE NEED FOR REVISION

In our initial work the focus was purely on elicitation of usage patterns (and to some extent problems/challenges) of WO systems and resulted in a basic workable model (see fig 1.) of processes for discussion/validation with practitioners.

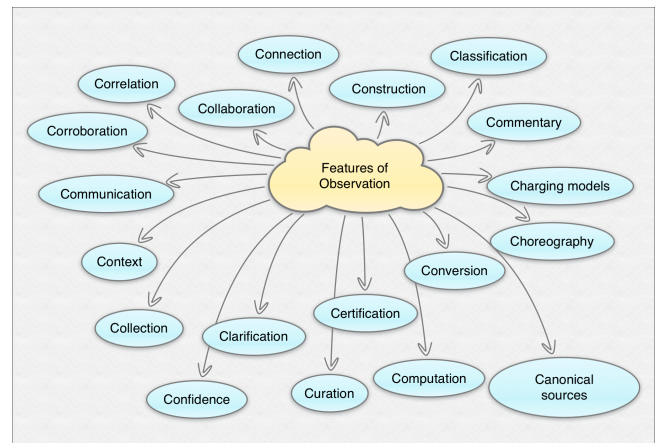


Figure 1 - Unstructured early model of WO processes

This initial view however did not capture any grouping/structures/sequencing of the processes and also suffered with a number of fundamental issues that we address in this paper:

- Firstly that the WO is treated as existing in a socio-technical vacuum with external factors (whilst perhaps implied) are not elucidated/enumerated.
- Secondly that the model does not make social elements of the social machine sufficiently explicit
- Thirdly the non-linear, emergent effects resulting from these process flows were ignored
- Finally that no method of implementing this type of analysis is suggested

<sup>1</sup> We assume not all relevant systems will choose to use the term *Web Observatory* but may be recognizable as similar/related to WOs.

Thus we offer here a substantially improved version over the initial model which we have termed  $e^5$  and which forms part of the larger DNA model.

### 3. METHOLODOLGY

The general approach to this work has been the development of a grounded theory according to the constructivist variant preferred by Charmaz [3] and thus, by definition, this model is substantive, focusing largely (though not exclusively) on academic research subjects. The wider research will also engage with business and government/non-profit participants in the future. It should be noted that the A section of the DNA methodology focusses specifically on this issue of differing perspectives and Archetypes and will form the subject of future papers/articles.

#### 3.1 Restructuring the model

Our initial action was to group the original 17 processes into a notional “black box” comprising three activities around the discovery, assembly and delivery of data and sources.

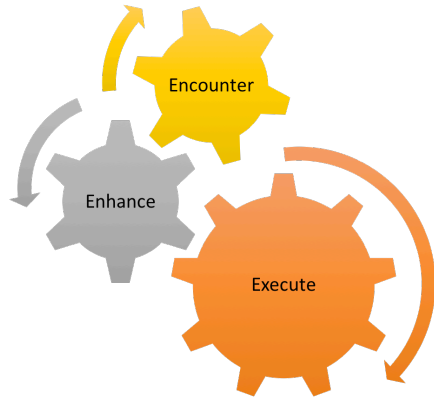


Figure 2 - WO processing model

Following the analysis of additional theoretical concepts from focus groups and individual interviews our focus was drawn to the idea that many of the challenges and factors brought up by participants were **exogenous** (either external to and/or unexplained by) this pure “processing” model of WO. We therefore include both external factors that might influence WO from an eco-system perspective and also those emergent factors that might result from the use of WO that would in turn feed back into the eco-system.

#### 3.2 Exogenous factors

The study of additional (non-process) facets of external forces and resulting (emergent) properties more than triples the size of the original model and also caused a number of the original processes to be refined/re-assigned.

Under the external model we elected to assess/categorise the processes according to the established PESTLE<sup>2</sup> model which covers respectively:

Political factors, Economic factors, Socio-Technical factors (sometimes considered as two factors: Social and Technical), Legal factors, Environmental factors

<sup>2</sup> <http://pestleanalysis.com/pest-analysis/>

Under the emergent factors we elected to assess/categorise the processes according to the idea that an effect should be exogenous in the sense of not-explained-by-the-model and may also broadly decompose into a PESTLE structure.

e.g. We might consider that revenue/profit resulting from license fees for WO services would **not** be an unexplained/exogenous factor (i.e. it results directly from the processing model), whereas the formation of community-lead *de jure* processes/standards potentially subverting the original specifications of the WO designers **would** be considered exogenous as they are **not** directly explained by the mere existence of a processing model.

### 3.3 The $e^5$ model

Adding in these two additional perspectives, the  $e^5$  model presents a more flexible view of WO operation as it recognizes the potential impact of inbound factors beyond the control of WO builders as well as outbound factors and effects (both intended and unintended), which may result from WO interactions and interoperations. The model extends the processing section as follows:

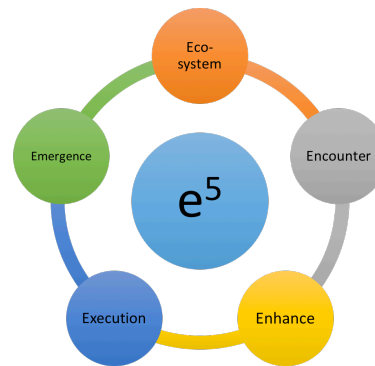
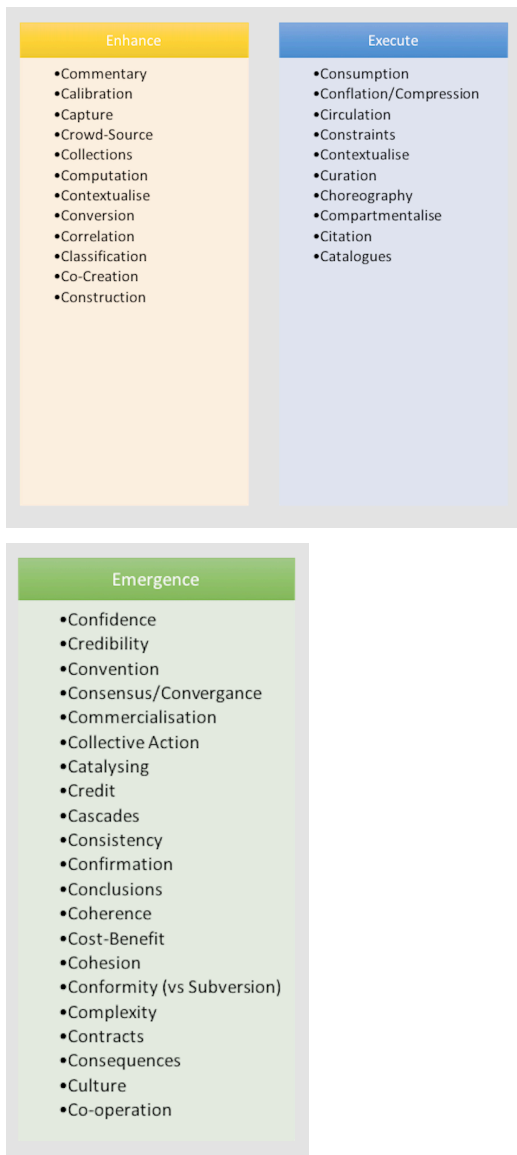


Figure 3 - Revised  $e^5$  process model

## 4. RESULTS

Following the collection and analysis of the additional focus group and interview material the process vocabulary (now expanded to a vocabulary of various 30 factors and 32 processes) was re-cast as follows:

Eco-System	Encounter
<ul style="list-style-type: none"> <li>•Community</li> <li>•Commercial</li> <li>•Charitable</li> <li>•Corp Structure</li> <li>•Collegiality</li> <li>•Celebrity</li> <li>•Convenience</li> <li>•Cost</li> <li>•Compulsion</li> </ul>	<ul style="list-style-type: none"> <li>•Communication</li> <li>•Canonical Sources</li> <li>•Clarification</li> <li>•Connection</li> <li>•Certification</li> <li>•Collections</li> <li>•Charging Model</li> <li>•Conflict (+resolution)</li> <li>•Confidentiality</li> <li>•Consignment</li> <li>•Conspicuity</li> </ul>



**Table 1 - e5 factors and processes**

Space constraints do not permit the reproduction of formal definitions/diagrams of each process in a short paper format but a visual representation and brief description of each process/factor is available for review/comment on-line at <http://bit.ly/1bjQf9P>.

#### 4.1 Future Work

As the grounded theories for this work continue to evolve we propose to consider the instrumentation and implementation of WO processes (just 3 of the 5 e5 groups since exogenous factors are not something that can be *implemented* but perhaps can be measured/evaluated). Whilst we can visualize process definitions using existing notations such as BPMN, UML and others we are particularly keen to look at *mobilizing* these process in a way that can measured and interwoven both with (semi) manual processes and also cross-platform interactions. To this end we are starting to consider platforms, which combine both work-flow and manual steps and so the next step is implement a number of processes as a

proof of concept including inter-system communication using lightweight protocols such as LSC [4].

In the wider context, as intimated above, the process perspective is only one of three in the **DNA** method and so there is further work to look for correlations and groups between functional aspects, process aspects and motivational aspects.

## 5. CONCLUSIONS

A model of socio-technical systems, social machines and Observatories needs a broad model of interactions comprising both a technical/processing vocabulary as well as a social/ecosystem context from which to operate. With so much research focusing on the emergent properties of the Web it seems inadvisable to propose a model, which does not allow both elements to be included in the analysis.

Whilst we are not suggesting that *every* social machine or observatory will be materially affected by *every* external factor, nor that each system will employ *every* processes listed, we do suggest that a core vocabulary of processes is useful for broader social machine analysis and that these findings represent a substantial improvement over the original WO process model. Ultimately we are looking for a set of useful techniques and perspectives that move us closer to an evidence-based (grounded) theory of the operational factors for a broad range of social machines.

## 6. ACKNOWLEDGMENTS

This work is supported under SOCIAM: The Theory and Practice of Social Machines. The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1 and comprises the Universities of Southampton, Oxford and Edinburgh and also the EPSRC grant number EP/K503150/1.

## 7. DATA ACCESS STATEMENT

Due to ethical concerns, supporting data cannot be made openly available. Further information about the data and the conditions for access are available from the University of Southampton repository.

## 8. REFERENCES

- [1] Brown, I.C., Hall, W. & Harris, L., 2013. From Search to Observation. Available at: <http://www2013.org/companion/p1317.pdf>.
- [2] Brown, I.C., Harris, L. & Hall, W., 2014. Towards a taxonomy for web observatories, International World Wide Web Conferences Steering Committee.
- [3] Charmaz, K., 2014. Constructing Grounded Theory, SAGE.
- [4] Murray-Rust, D. et al., 2014. A collaboration model for community-based Software Development with social machines. In Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2014 International Conference on. IEEE, pp. 84–93.