

Giving user-generated content back to the users: Testing data de-centralisation with active content creators

Amy Guy
School of Informatics
University of Edinburgh
amy.guy@ed.ac.uk

ABSTRACT

Our poster gives an overview of on-going work which is part of a broader effort to enhance and facilitate collaborative opportunities for digital content creators. We describe *creative media production social machines*, which gives a perspective on the user-generated content ecosystem that changes how we consider content creators and the technical systems they use. We observe individual content creator behaviour within social machines - understanding their behaviour from the bottom-up rather than top-down - in order to inform better design of systems which support their interactions. We illustrate on-going work in formalising certain observations, to provide a basis for designing and developing systems which can successfully co-operate in the creative media production social machine space.

Author Keywords

Social machines; user-generated content; metadata; creative media; user-centric design; identity; online communities.

ACM Classification Keywords

H.1.2. Information Systems: User/Machine Systems.

INTRODUCTION

Within the online user-generated content ecosystem, there are many *active* content creators. For them - amateur musicians, independent film-makers, artists, storytellers - content creation is a core part of their lives and perhaps a means to an income. They are actively seeking new audience members, and are likely to understand the ones they've got fairly well. They collaborate with other creators, work to improve their output, and cultivate one or more online personas through which they present their content.

They do not confine themselves to one website or community, but spread versions of themselves across several. Their success depends on both social and technical factors; how they present themselves, how their audience receives their content, how the platform(s) on which they

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author(s).

Published in: van Leeuwen, JP, Stappers, PJ, Lamers, MH, Thissen, MJMR (Eds.) *Creating the Difference: Proceedings of the Chi Sparks 2014 Conference*, April 3, 2014, The Hague, The Netherlands.

publish permit their content to be consumed, shared and re-used, and how different content-host and social networking platforms may interact with each other.

We therefore take the perspective that these content creators are participants in one or more *social machines*: critical parts of the systems in which they operate, rather than merely users. Their desires and needs transcend individual websites, applications or communities, and we argue that this should be taken into account when improving existing systems or designing new ones.

Social Machines

Social machines are systems for which human and computational components are complementary and equally critical: "processes in which the people do the creative work and the machine does the administration" [1]. They exist already on many different scales and in many different domains, often evolving, responding to technological and social developments, and interacting with each other [8, 3, 5, 12].

Creative Media Production Social Machines

In [7] we describe *creative media production social machines* as "a broad class of systems where:

- "humans may use a purely digital, or combination of digital and analogue methods, and a degree of creative effort, to produce media content;
- "the content is published and publicly accessible on the web;
- "a global audience may consume, curate and commentate on this content in technologically-mediated environments."

When we realise that users of a system implicitly adopt one or more online personas across multiple systems (websites, applications, communities), and intermix their activities and interactions between systems whether the technology explicitly allows it or not, we can cease to consider them as users. Rather, they are *components*, without which the systems wouldn't function; they are part of a larger socio-technical picture.

MOTIVATIONS

At present, metadata about content creator personas, activities and interactions, not to mention the content they produce, is typically hosted by a central content-host site (like YouTube). Content-host sites, whether they host the same types of media or not, typically do not allow cross-site interactions, or easy transfer of data from one site to another.

Reliance up on these silos is disempowering to content creators. Their success is usually hinged around their

metadata, for example how many videos they have uploaded, how many views their content has received, who their connections are. If a content-host site disappears, a large part of their creative portfolio and online networks may disappear with it. This is similarly the case if a creator's online profile is banned by a particular site (many creators on YouTube have problems with false copyright claims [11]).

They are also limited by the functionality provided by the site(s) they choose to use. Developers seeking to extend content-host site functionality are restricted by the APIs (or lack thereof) available. Despite the ability to share or embed content through social networks or personal websites, data about the content is still confined to the environment of a particular content host, and difficult to mash-up with content or data in other systems.

Availability of open web standards provides an opportunity to return ownership of data and metadata around content creation activities and interactions to the content creators themselves.

GOALS

Our goals therefore are to define a model for representing content creators and their content which will allow them to be referred to independently of the website on which content is hosted. We will implement this model using open web standards, namely linked data, to allow the creation of useful applications which can communicate with each other.

This outcome goes beyond simply facilitating the management of disparate content by its creator(s), but aims to allow them to have complete ownership of any and all data associated with a creative work, independently of their chosen creation and distribution platforms. They can then permit third-party applications to access the data needed to perform useful functions. Next, we describe some examples of when this will be useful.

Example scenarios

Changing content distribution platform: A successful amateur filmmaker publishes their work on YouTube. They have 100,000 subscribers, and millions of views. At some point, YouTube makes some visual or functional change to the website that the filmmaker doesn't like. Currently, the filmmaker's reputation and audience are tied to YouTube, so they have little choice but to tolerate the changes (even if the changes meant that they were more readily exploited by advertisers, or their content was less often recommended to people because of algorithm changes). Assuming instead that all content creators' data is available as linked data, and content host platforms which can consume linked data have been developed, they are free to publish elsewhere whilst retaining their reputation and links to their audience.

Creating a portfolio of creative work: A musician and sound designer who has recorded their own album (available on SoundCloud), collaborated with another musician to produce an EP (the other musician released it on BandCamp), made backing music for a short film (published by the film's director on YouTube) and created

the soundscape for someone's animation (on Newgrounds) must manually collate this information onto a personal website to effectively build their portfolio. They have no web development knowledge and no resources to pay someone else to help. With all of the relevant data instead available as linked data, it is trivial for third-party portfolio-creation application to access and visualise it in a logical and attractive way. Metadata such as listens/watches, downloads or comments can be automatically integrated too.

Finding collaborative partners: In the current ecosystem, content creators must rely on personal networks or cold approaches to meet others to work with. For many who are just starting out, this is especially difficult. With a system that can access and interpret linked data to determine creators' skills, availability, interests and needs, partnerships can easily be automatically suggested, and perhaps initiated by the indication of automatically detected mutual friends.

APPROACH

We use ethnographical studies of content creators to inform the process of formally representing their identities and activities. The formal representation, an ontology for linked data, is in preliminary draft stages. We use this ontology to annotate content creator activities across a number of different platforms. We can then demonstrate the utility of this linked data by building applications that complement content creator activities regardless of the websites they use.

Studies

We interviewed 39 amateur content creators, selected at random during two days of *VidFest*, part of the 2013 London ComicCon. We learnt that two thirds work collaboratively with others but none used specialist tools to facilitate collaboration or improve their creative process.

We also observed current behaviour of a small but diverse sample of ten content creators 'in the wild', with the goal of reaching a better understanding of their online identities, connections, activities and interactions. For the ten creators, 93 online profiles were discovered. Through these profiles, we discovered that many creators represent different versions of themselves online by adopting multiple personas. One persona may have many online profiles across different websites. Personas sometimes even interact with each other as though they are completely separate people. More details of this study can be found in [7].

It is important to note that we have no desire to attempt to disambiguate the identities of content creators, to connect their online personas to an offline individual behind them, or to otherwise infringe upon web users' right to pseudonymity or anonymity. Rather, we argue that it is crucial to acknowledge when, how and why creators use different presentations of themselves, in order to improve the systems they use.

Ontologies

Many existing ontologies are suitable to some degree for representing aspects of digital content creators. We summarise the key ones here:

- FOAF: Friend-of-a-Friend can be used to represent connections between people [6]. We extend this to allow more fine-grained types of connection.
- SIOC: Semantically-Interlinked Online Communities provides core concepts for linking textual exchanges of messages to each other and to their creators [2].
- OntoMedia: this provides generic content description for various media types, and allows descriptions of relationships between different media types [9].
- PROV-O: The W3C Provenance Ontology allows annotation of an activity, the inputs and outputs, and the roles that people play [10]. We will extend this to provide more domain-specific detail, where the activity described is the 'creation process' of a piece of media content.

CONCLUSION

We have outlined some of the limitations experienced by content creators in the current user-generated content ecosystem. Our preliminary studies show that the complex and nuanced identity behaviours of content creators makes it imperative to consider their experience within creative media production social machines, rather than on individual websites and services, in order to design and develop improved systems.

REFERENCES

1. Berners-Lee, T. Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. London. (1999). Texere.
2. Breslin, J., et al. SIOC: an approach to connect web-based communities. *International Journal of Web Based Communities*, 2(2), 133-142, (2006).
3. Byrne Evans, M., O'Hara, K., Tiropanis, T., and Webber, C. Crime Applications and Social Machines: Crowdsourcing Sensitive Data. In *WWW 2013 Companion*, Rio de Janeiro (2013). ACM.
4. Dalton, B. Pseudonymity in Social Machines. In *WWW 2013 Companion*, Rio de Janeiro, (2013). ACM.
5. De Roure, D., et al. Observing Social Machines Part 1: What to Observe? In *CHI'13*, Paris, 2013. ACM.
6. Graves, M., Constabaris, A., and Brickley, D. Foaf: Connecting people on the semantic web. *Cataloging & Classification quarterly*, 43(3-4), 191-202, (2007).
7. Guy, A. and Klein, E. Constructed Identity and Social Machines: A Case Study in Creative Media Production. In *WWW 2014 Companion*, Seoul (forthcoming). ACM.
8. Hendler, J. and Berners-Lee., T From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*, 174(2):156–161, (2009).
9. Jewell, M.O., Lawrence, K.F., Pruegel-Bennet, A. and schraefel, m.c.. Annotation of Multimedia Using OntoMedia. In *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*. Edinburgh, 2006.
10. Lebo, T., et al. PROV-O: The PROV Ontology. Working Draft, W3C, 2012.
11. Power, M. "Looks like this is goodbye.." [Facebook update] Retrieved from <https://www.facebook.com/powerm1985/photos/a.267403889986893.65529.177637432296873/656515964409015/?type=1&theater> on 16 February 2014.
12. Strohmaier, M. A Few Thoughts on Engineering Social Machines. In *WWW 2013 Companion*, Rio de Janeiro, 2013. ACM.