# OPEN DATA PLATFORMS FOR A DATA COMMONS: HOW TO SUPPORT THE LIFE OF DATA BEYOND RELEASE

Hannah Hamilton
University of Stirling

Stefano De Paoli
Abertay University

Anna Wilson
University of Stirling

Greg Singh
University of Stirling

## Extended Abstract

As the theme of this conference suggests, the digital technologies that pervade and shape our lives have life of their own. Many of these technologies are, in turn, animated by and give life to the data they consume, process and release. Yet the majority of people are excluded from the production and evolution of both digital technologies and data, with control of these processes resting in the hands of large corporations and governments. Our relationship with data in particular is characterised by an imbalance of power, and the liveliness of data can seem a threat rather than an opportunity. In this paper, we suggest that in order to achieve a more productive, symbiotic relationship with data, people need to be enabled to engage with it both critically and creatively. We describe a project aiming to design a system that facilitates just such engagement with Open Data, using waste data as a case study.

The Open Data movement has long sought to make data more lively by making it more accessible. There is an established consensus that the availability of digital data and Internet-based communication technologies should foster economic and social well-being (Shirky, 2010), as well as business innovation and productivity (Jenkins et al., 2013). Indeed, this has been a key expectation of Open Data policies since the G8 Open Data Charter of 2013. Open Data are expected to create new opportunities for economic activity, improve transparency and governance, and empower people to live in more creative and sustainable ways through increased knowledge. Indeed, the Open Knowledge International links data and knowledge

through their definition of Open Data: "Knowledge is open if anyone is free to access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness" (Open Knowledge Foundation, n.d.). Large quantities of data are being produced by many organisations and released openly online for other actors to use. Yet despite several years of effort, the extent to which open data are animated by critical and creative interaction remains limited. Some authors have argued that it is too simplistic to assume that people and other actors will simply take open data and infuse in it a new life beyond the original use (Janssen et al. 2012). Even if data is made available via the Internet for others to reuse, it does not necessarily follow that this data will be easy to find, use or indeed be trustworthy (Meijer et al. 2012). Indeed, it is often unclear whether open data has a life beyond the one it was initially collected for, to the extent that some authors have described the public reuse of government data as no more than a "myth" (Hellberg & Hedström, 2015).

One way to tackle some of the challenges surrounding open data use is via the creation of Internet-based platforms designed to promote and facilitate data reuse by the public (broadly defined to encompass not only individuals but also all sorts of organisations), establishing open data as a commons, aggregating the data and offering recommender systems capable of supporting the life of data beyond its initial use and production. However, designing such platforms is not easy for a variety of reasons. Firstly, it will require the platform producers to adopt at least a symmetrical approach (Latour, 2005) addressing the technological problem of creating an internet-based infrastructure, collecting the data and ensuring it is interoperable whilst also and at the same time addressing the social problem of creating a healthy and active community around the platform. Secondly the platform producers must tackle the challenge of facilitating education about the use of data for the many, rather than a limited number of individuals or organisations that possess specialised skills (e.g. in data analysis of visualisation) (Mandinach & Gummer, 2013).

*Data Commons Scotland* is a project that aims to create a prototype Internet-based platform that not only enables access, but actively encourages increasingly sophisticated and critical use of open data, thus facilitating a life for data beyond the one of the original producer. During initial scoping research, the project team identified three major problems with existing systems: disjointedness; single-level use design, and inconsistency. The project is using the case study of waste data in Scotland, with the idea that data produced by many actors (e.g. government, companies, public authorities) could then have a new life in the hands of other stakeholders, including individuals, schools, businesses and other organisations. The successful reuse of waste data could support behaviour change in a number of fields (such as circular economy or Open Energy), fostering collaborative, sustainable environmental awareness through data literacy. The challenge, then, is to design a system that meets not only the existing needs of multiple users, but also recognises and builds on their capacity for learning.

To ensure that a range of potential users are represented in the design process, we have conducted qualitative interviews (n=31) with a number of data owners (e.g. public authorities, recycling companies) and potentially interested stakeholders (e.g. teachers, individuals) as well as data collected from two workshops. The first workshop targeted the insight of stakeholders working within the area of waste, the second on those interested in open data. The focus of both the interviews and workshops was to discover needs, values and capacities in relation to a new

platform supporting the reuse of data in Waste Management and how such a platform needs to function, look and feel to offer maximum utility to a range of audiences. Dialogues included explorations of the usefulness of certain features, e.g. recommender systems, chatbots, tools for assessing data reliability, as well as discussions on how common barriers to opening up data within the area of waste and beyond may be reduced or removed.

These primary data were analysed using the methods of phenomenography (Åkerlind 2012), which explicitly embraces variation rather than seeks to define typical or average experiences and understanding. The results were used to create personas and scenarios that reflect the complexities of potential users' lives and, in particular, their values and interests (Wilson et al. 2018). The scenarios emphasise that engagement with the platform might be driven by more than one interest, and that the platform itself might be designed to encourage increasingly critical and creative engagement with data. They describe how different audiences may hope to interact with an open data platform on waste, as well as describing encounters they may have that would make this process easier or more of a challenge.

These scenarios have been constructed not only as a research tool but as a mechanism for communicating the needs of users to the technical partners on the project, who will be responsible for shaping the look, feel and functionality of the platform. They identify immediate challenges, particularly around issues of data reliability and confidence, and the development of mechanisms to enable non-experts to engage with data uncertainties. Our presentation thus raises questions as much as it illustrates our findings.

## References

Åkerlind, G.S. (2012). Variation and commonality in phenomenographic research methods. *Higher Education Research & Development*, 31(1), 115-127.

Hellberg, A. S., & Hedström, K. (2015). The story of the sixth myth of open data and open government. *Transforming Government: People, Process and Policy*. 9(1), 35-51.

Janssen, M., Charalabidis, Y. & Zuiderwijk, A., (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, *29*(4), 258-268.

Jenkins, H., Green, J. & S. Ford (2013). *Spreadable Media: Creating Value and Meaning in a Networked Culture*. NY: NYU Press.

Latour, B. (2015). *Reassembling the Social—An Introduction to Actor Network Theory*. Oxford: Oxford University Press.

Mandinach, E. B. & E. Gummer (2013). Defining data literacy: A report on a convening of experts. *Journal of Educational Research and Policy Studies*. 13(2), 6-28.

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of theoretical and applied electronic commerce research*, 9(3), 32-44.

Open Knowledge Foundation (n.d). Open Definition. Available at https://opendefinition.org/od/2.1/en/.

Shirky, C. (2010). *Cognitive Surplus: Creativity and Generosity in a Connected Age*. London: Allen Lane.

Wilson, A., De Paoli, S., Forbes, P., & Sachy, M. (2018). Creating personas for political and social consciousness in HCI design. *Persona Studies*, 4(2), 25-46.