Open government data (OGD): challenging the concept of a "Designated Community"

Nathan Moles
Faculty of Information, University of Toronto, Toronto, Canada

Received 23 October 2019
Revised 31 March 2020
Accepted 23 October 2020

Abstract

Purpose – This paper aims to explore the curation of government-produced datasets for release as open government data (OGD) from the perspective of the digital curation and preservation concept of a "Designated Community". Specifically, it explores how digital curation functions when there is no clear Designated Community to which curation services can be targeted.

Design/methodology/approach – The research was conducted through a case study of the City of Toronto's efforts to revitalize their OGD program. Data was collected using three methods: semi-structured interviews, non-participative observation and document analysis.

Findings – The curators of OGD responded to the absence of a Designated Community through two complementary methods. The first was to draw from the discourse that defines the OGD domain. The second was to take a participatory approach that incorporated members of the community surrounding OGD and various other stakeholders into the process of developing a plan for the revitalization of the program.

Research limitations/implications – This study opens new directions for investigating the application of the Designated Community concept and its role in digital curation and preservation.

Practical implications – The approach used by OGD curators in this case has the potential to be used in other curation situations where there is no clearly defined user group.

Originality/value – The findings presented in this paper contribute empirical insights to on-going discussions on the concept of a Designated Community in digital curation and preservation.

Keywords Open data, Digital curation, OAIS, Open government data, City of Toronto, Designated Community

Paper type Research paper

Introduction

Digital curation can be understood as "active involvement in the management, including the preservation, of digital resources for future use" (Becker, 2014). Expanding on this definition, digital curation focuses not just on preserving digital information, but on keeping it functional, supporting its on-going fitness for purpose and adding value (Ross, 2006). Conceptualized in this way, digital curation is an outgrowth or evolution of the field of digital preservation (Yakel, 2007).

One of the central concepts of digital preservation is the "Designated Community". Introduced by the Reference Model for an Open Archival Information System (OAIS), a high-level standard for conceptualizing digital preservation repositories, OAIS defines a Designated Community as "[a]n identified group of potential Consumers [users] who should be able to understand a particular set of information" (CCSDS, 2012, pp. 1–11). In OAIS, the knowledge, skills, and use of digital information by a Designated Community are used as a reference to ensure the information products delivered to users are tailored to their needs

Records Management Journal
Vol. 31 No. 1, 2021
pp. 18-33

0956-5698
DOI 10.1108/RMJ-10-2019-0065

and skills. While OAIS was specifically developed for organizations with long-term preservation responsibilities, its applicability explicitly encompasses "organizations with other responsibilities, such as processing and distribution in response to programmatic needs" (CCSDS, 2012, pp. 1–2).

The emergence of this concept marks a noticeable shift in the how curation and preservation are conceptualized (Bettivia, 2016). In the analogue realm, the artifacts themselves have always been considered the focus of curatorial efforts. With digital information, however, the focus shifts away from the artifacts and towards the users or purpose for which curation and preservation activities are undertaken. In other words, digital curation and preservation efforts are outcome-oriented and targeted at specific users and their uses of the digital information. The concept of a Designated Community enables this shift and makes it explicit.

Today this approach serves as the basis for most digital curation and preservation practices. As noted by Bettivia (2016), however, this conflicts with the mandates of many organizations that serve diverse audiences. Information professionals engaged in digital curation and preservation efforts at such organizations are caught in a conflict. The established means of curating and preserving digital information assume an identifiable user group that is not consistent with their organizational mandates or the domain in which they operate. This conflict motivates the research question for this study:

> How does digital curation function when there is no clearly identifiable Designated Community or user group towards which curation services can be targeted?

The research question was investigated in the context of releasing government-produced datasets to the public as open government data (OGD), a domain of curation that precludes a Designated Community and challenges digital curation orthodoxy. A case study of the City of Toronto, a large municipality engaged in innovative program-building for their OGD operations, was used as a study site.

The following sections provide background on the concept of a Designated Community and previous research in the area, before discussing OGD and the methods used in this research. The specifics of the case are described before a discussion of the findings for the research question.

Background and previous research

OAIS and the Designated Community

The concept of a Designated Community appears to have originated with OAIS (CCSDS, 2012). Its key feature as a theoretical construct is that it is not a group of actual users, but an abstract understanding of a group of potential future users (or Consumers in OAIS terminology). OAIS conceptualizes the Designated Community as a collective, potentially made up of multiple user groups that share an identifiable set of characteristics. Amongst these is the "Knowledge Base", understood as "a set of information, incorporated by a person or system that allows that person or system to understand received information" (CCSDS, 2012, pp. 1–12). In OAIS, a Designated Community is defined by those undertaking the digital preservation efforts as their target user group, although with the recognition that this group can evolve or change over time.

The Designated Community features prominently throughout the Standard and is central to OAIS's definition of an Archive, as an organization with the intent to preserve digital information, and the mandatory responsibilities an Archive assumes in its management or curation of digital information (CCSDS, 2012).

In most discussions of the Designated Community concept, an understanding of the Designated Community, including its Knowledge Base, is regarded as a means to inform the

tasks needed to deliver digital information that is technically and intellectually tailored to users (for example Parsons and Duerr, 2005; Giaretta *et al.*, 2009). Some experts have argued that the Designated Community should be leveraged to define the boundaries of the descriptive information required to make digital information intelligible (Giaretta *et al.*, 2009). Others have observed that it is users who ascribe significance to digital information, or specific characteristics of digital objects, within a set context for a specific goal or purpose (Dappert and Farquhar, 2009). In either case, the Designated Community is presented as a source of requirements for curation and preservation processes.

Most of the previous research explicitly focused on the Designated Community in digital curation and preservation has been instrumental in its orientation and assumed the existence of a clearly defined user group or groups towards which curation and preservation efforts could be directed. In an early theoretical contribution to the topic, Parsons and Duerr (2005) argue that the Designated Community concept is only useful if the Designated Community is well understood. To this end, they provided guidance for the application of the concept. Parsons and Duerr advise scoping the Designated Community as broadly as possible to capture variation in the use of digital information but limiting this definition to those users who share common characteristics, so that the Designated Community can be described in rich detail. Amongst their recommendations was that the concept of a Knowledge Base be expanded to include discursive aspects such as conceptual metaphors, common knowledge, or generally accepted opinions that often contextualize digital information within communities of users. Baker *et al.* (2016) would later illustrate this recommendation by showing how the creation of "data products" at the National Snow and Ice Data Center (NSIDC) expanded and co-evolved with the NSIDC's Designated Community.

Other scholars have sought to operationalize the Designated Community concept through tools and methods that draw from the interactions of actual users with repository systems. Giaretta (2011) developed an analytical approach he referred to as "DC Profiles" which involved the identification of modules of information resources and their dependencies. By assigning modules to communities, structuring them hierarchically, and formalizing them. This approach would allow for a comprehensive mapping of the Knowledge Base of the Designated Community and provide a robust reference for curation and preservation actions. Kärberg (2014) developed a method for gathering user observation data from the interactions of users with the National Archives of Estonia website using web analytics and user surveys. From this data, users were divided into three groups based on the frequency of their visit, their path to the website, and the duration of their engagement. The resulting groups could then be modelled using formal matrices of binary triples. Working in a similar direction, Kim (2015) used data from web crawls to profile a Designated Community in a web environment, by their HTML content, the structure of their knowledge organization and webpage links. Using this method, Kim discovered that online communities follow recognizable patterns and that these can be based on information such as the file extensions they use and user-generated tags.

An alternative take on the Designated Community comes from the work of Huvila (2008). In a case study describing the creation of two digital archives, Huvila introduced the concept of an open Designated Community. The archives in his study were constructed using an approach adapted from the literature on participatory archives (notably Shilton and Srinivasan, 2008; Yakel *et al.*, 2007). This approach is characterized by three principles: *decentralized curation* in which responsibility is shared between curators and participants; *radical user orientation* in prioritizing discoverability and usability over preservation; and multiple dimensions of *contextualization* in addition to provenance. In this participatory

context, the self-defining open Designated Community is one in which users hold responsibility for providing contextual information for themselves and their colleagues. Like the curating organizations of concern in the present study, these archives faced users with varying technical skills and research interests. However, the archives and the data they hold reflect a specific topic which provides a common thread across users and the open Designated Community, from whom contributions to contextual information are expected, is comprised of authenticated users with subject expertise.

The most in-depth consideration of the Designated Community to date has come from Bettivia (2016). Through a series of interviews with OAIS creators, and practitioners from the cultural heritage sector, Bettivia examined the dynamics of power between an OAIS Archive and its Designated Community. She noted that the concept of a Designated Community inherently forces curating organizations to favour some users over others. In the process, Bettivia raised the concern that the Archive is empowered by its ability to define its Designated Community to the detriment of the communities it serves. In part this is due to a lack of mechanisms for including input from user communities and methods for applying the concept in inclusive ways. As a response, Bettivia mirrors Huvila's (2008) innovations by suggesting that Archives define their Designated Communities in collaboration with actual users.

Several of Bettivia's (2016) other findings are also significant for the present study. Amongst these, Bettivia found that interview participants from organizations that were more oriented to curation than preservation felt the concept of a Designated Community was important in their work. However, a common theme heard from her interview participants was that aligning their curation and preservation processes with their Designated Communities and the objectives of their organizations was itself a complicated challenge. Many of these institutions did not have formally defined Designated Communities and their understanding of their users was frequently based on tacit knowledge.

Bettivia's (2016) study did not examine the curation process of any one of these institutions in detail and left open the question explored here of how curation functions when the situation precludes a clearly defined Designated Community. Like the cultural heritage institutions that were the focus that research, OGD curation has no set or pre-defined user group, which makes it an ideal situation in which to explore these issues, as the following section explains.

Open government data

OGD are datasets created or collected by government bodies or agencies and released to the public. As a form of open data, OGD meet the terms of the Open Knowledge Foundation's *open definition* in that they can be "freely used, modified, and shared by anyone for any purpose" (Open Knowledge Foundation, n.d.). OGD are frequently released to share their social and commercial value, to facilitate transparency in government, and enable participatory governance (Open Knowledge Foundation, n.d.). Examples of datasets released as OGD include data related to transportation, government services, and public finances.

Several attempts have been made to create and iterate principles that define what constitutes OGD (for example Malamud *et al.*, 2007; Group of Eight (G8), 2013; International Open Data Charter, 2015). These have focused predominately on barrier-free accessibility, releasing data in formats that enable diverse use, and the removal of restrictions on use, including for commercial purposes. Sets of principles such as those included in the International Open Data Charter (2015) acted as requirements that datasets must meet to fulfill the objectives of the larger social movements in which OGD participates.

OGD curation occurs within a complex context at the intersection of two broad social movements (Yu and Robinson, 2012). The first is the technology-oriented open data movement, which is characterized by advocacy for freely available data with minimal restrictions or access barriers. Often this is done with an eye towards innovation and the commercial potential of insights generated by data analysis. The second is the open government movement, defined by politically oriented attempts at accountability, transparency, and administrative efficiency in governance. However, Yu and Robinson (2012) have observed a blurring of the distinction between open data and open government, meaning that OGD writ large participate in both enterprises simultaneously. The current discourse surrounding OGD is a merger of these two movements that depoliticizes OGD and presents data as ideologically "neutral", while conflating the interests of citizens, governments, and the private sector around the push for data release (Currie, 2016). Both the open data and open government movements encourage participation by a wide range of people and seek to enable a variety of forms of engagement. OGD are usually presented as a way of realizing open government ambitions while simultaneously providing data that can drive technological innovation and economic prosperity. Advocates emphasize a wide range of potential uses of OGD that include combining OGD with other datasets, both public and proprietary, data mining, remixing, mashups, visualization and the range of operations facilitated by data analysis tools or social media platforms.

By embracing a diverse user base and encouraging a wide range of data uses, the OGD community face a challenge very similar to that confronting the cultural heritage institutions studied by Bettivia (2016). Users have differing capacities to access and use OGD, and these capacities are tightly linked to the benefits that can be obtained from data release (Davies and Bawa, 2012). Within this context, OGD initiatives are tasked with releasing data that is usable by an audience with varying knowledge, skills, interests, and expertise, while simultaneously positioning that data so that it contributes to the larger objectives of open government. There is a concern here that OGD initiatives may inadvertently focus on serving the interests of a small technically sophisticated group of users at the expense maximizing the accessibility and usability of data for the widest possible audience (Janssen, 2012). Insights from various streams of research support this framing of OGD curation.

Initially, the focus of OGD initiatives was almost exclusively on making data available. However, there has since been a recognition that releasing OGD alone is not adequate for achieving open government (Evans and Campos, 2013). Gurstein (2011), for example, notes an important distinction between the accessibility of data and its use as part of open government or other endeavors. Likewise, Janssen *et al.* (2012) identified the assumptions that releasing data will automatically yield benefits, that it inevitably leads to open government, and that OGD can be readily used by the general public, as some of the "myths" surrounding OGD. These points were affirmed in a case study by Hellberg and Hedström (2015), which showed that ordinary citizens had limited ability to use or engage with OGD, indicating that guidance and support were required to enable broad use. Further support for this position comes from the insight that OGD curation processes and initiatives themselves play a role in shaping datasets, and by extension, influencing the ways in which they can be used. In revealing case studies, Davies and Frank (2013) and Denis and Goëta (2014) called attention to the role of curation in constructing OGD, while questioning its usability and relationship to the activities the data supposedly reflect. In analysis at a higher level, Bates (2014) linked the selection of datasets for release and their prioritization to the financial interests of the private sector and neo-liberal ideologies in government.

As the open data and open government movements have matured, much of the focus has shifted towards positioning OGD as a mechanism to facilitate participation in open government and its usability in that capacity. One approach has been to incorporate the involvement of citizens and data users into OGD initiatives. This is not, however, a simple matter of inclusion. Instead there are degrees and varieties of participation. Linders (2012) describes one form of participation as "citizen coproduction", in which the citizens served by OGD initiatives are treated as active contributing partners in addressing social challenges and share some of the responsibility for outcomes. Building on the work of Linders (2012), Sieber and Johnson (2015) outlined conceptual models for OGD initiatives that indicate potential directions for growth and ways in which initiatives can incorporate citizen participation. Amongst these are the promotion of OGD to specific communities or for specific purposes, crowdsourcing and the use of OGD as a tool to enable citizen participation in decision-making or policy developed. In creating these conceptual models, Sieber and Johnson drew distinctions between the provision of OGD, more active and collaborative approaches to its use, and a deeper level of engagement involving a bidirectional exchange of information and cooperation in decision-making and the development of policy.

Examples exist of OGD initiatives making efforts in this direction. In the case described by Hellberg and Hedström (2015), OGD curators organized a working group that included stakeholders from the private sector and a university, along with staff from a municipal government, to encourage OGD release and subsequent use in an innovation competition. A different approach is described by Both (2012), where citizens in Berlin were asked to provide input regarding the prioritization of datasets for release by the municipal government through an online poll. These endeavors experienced limited success but generated enthusiasm and support for their general approach.

There has been some research on OGD release conducted from a records management perspective. This research has produced complementary insights but has not approached OGD curation as problematic in relation to the concept of a Designated Community. Research into the shared conceptual foundations of OGD and records management was undertaken by Borglund and Engvall (2014). They examined the discourse surrounding OGD and digital archives initiatives in Sweden and demonstrated that the central information construct in each area are synonymous. This suggests that OGD are records being referred to in different terms by the global open data and open government movements, which in turn supports the application of records-related concerns to OGD.

Empirical research conducted as part of the InterPARES Trust explored the role of records managers in OGD release. Case studies in the UK examined OGD initiatives in a local government (Page *et al.*, 2014), the National Health Service (NHS) in England (Harrison *et al.*, 2015), an institution of higher education (Brimble *et al.*, 2016), and a semi-autonomous hospital trust (Chorley *et al.*, 2016; Chorley *et al.*, 2017). These studies revealed that although records management staff were a consistent presence across all cases there was significant variation in the organizational structures of OGD initiatives, the roles and responsibilities involved in OGD release, and the extent of guidance available (Shepherd *et al.*, 2019).

Another branch of the InterPARES Trust investigated the record-keeping requirements of OGD initiatives in a Canadian context. McDonald and Léveillé (2014) considered the applicability of retention and disposition specifications to records and datasets. Amongst their contributions was the recognition that OGD are the product of a distinct business process separate from the organizational functions through which data are initially created and used before their release to the public. Léveillé and Timms (2015) extended this analysis by identifying three sequential stages of the OGD release process. However, the three stages

of initiation, identification and distribution, promotion and release, downplay the construction of datasets by not assigning this work a distinct stage. As part of this research, the five engagement categories of the International Association for Public Participation's (IAP2) engagement spectrum were connected to the five contexts identified by diplomatics (Suderman *et al.*, 2015; Suderman and Timms, 2016).

In light of the background on the concept of a Designated Community in the section above, OGD can be seen as presenting a unique challenge for digital curation. The aspirations of the open data and open government movements provide a motivation for the release of OGD, but also preclude any narrowly defined user group. Janssen's (2012) point about the inadvertent exclusion of some users echoes the concerns raised by Bettivia (2016) regarding inclusivity and the power imbalance between an OAIS Archive and its Designated Community. The need to counterbalance this concern means that OGD curation stands in contrast to established ways of conceptualizing digital curation and preservation on this point.

Research design and methods

This study was conducted using a case study research design (Yin, 2009). Data collection took place between September 2016 and January 2018 using three methods: document analysis, non-participative observation, and semi-structured interviews. The document analysis was conducted first to establish an understanding of the background and context of the case, identify events for observation, and provide a basis for interview questions. Documents for the study were obtained by searching the City's website using three terms related to OGD ("open data", "open government" and "information management"). Any documents to which this initial set referred, including those from sources outside of the City, were subsequently obtained. In total, 462 documents were analyzed, including policy documents, meeting minutes and agendas, and written submissions to public-facing committees. The vast majority of these documents were publicly available, although a small number of documents were obtained through a freedom-of-information request. The document analysis itself was done following a technique articulated by Hodder (2000) that involved reading, considering the documents within the broader context of the case, looking for patterns, and following the course of events through chronological documents.

Non-participative observation was selected as a method to complement document analysis and provide information that would not have been captured in documents or revealed in interviews. 21 public events were observed. Those events that took place within the study period were observed in person, while relevant meetings of elected officials prior to September 2016 were observed through the City of Toronto's YouTube channel [1]. As part of the observation process, the author noted any details that appeared significant or relevant for the study without becoming involved in the proceedings. During the group activities in the public consultations, however, the author did participate in the group discussions, but attempted not to steer the conversations in specific directions.

The semi-structured interviews were done concurrently with the observations of events. Study participants were recruited with the assistance of a primary contact at the study site, who reached out to prospective participants through introductory emails. Over the course of the data collection period, further study participants were approached directly. In total, 18 semi-structured interviews were conducted with curation participants and OGD advocates external to the City administration. The interviews followed the advice of Brinkmann (2014) and Galletta (2013). An initial set of interview questions, derived from the documents analyzed and the research question, was tailored to the specific roles of the interview participants in OGD curation and evolved over the course of the study. Participants were

asked open ended questions as well as questions focused on details of the curation process. They were encouraged to express their thoughts freely and were not held rigidly to the structure of the interview protocol. The interviews ranged in duration from 30 min to roughly two hours, with most taking approximately 45 min to complete. The interviews were recorded and transcribed by the author. Together with the documents and notes from the observations these formed the data for analysis.

Analysis of the data was guided by Cultural-Historical Activity Theory (CHAT) (Engeström, 1987). CHAT is a body of theory that emerged from Soviet psychology in the 1920s and since the 1990s has been used in diverse fields such as human-computer interaction (Nardi, 1996), information systems (Ditsa and Davis, 2000), information seeking behaviour (Wilson, 2006), and digital curation (Lee, 2015). Based on the work of Vygotsky (1962, 1971, 1978), its focus is on understanding goal-oriented individual and collective social activities. As articulated by Engeström (1987), CHAT contains concepts that define activities, including the subjects who undertake an activity, the objectives they aim to achieve, and the social, material, and conceptual components that make up activities. In this modern form, CHAT provides researchers with a theoretical lens to dissect activities into their constituent components and identify the relationships between those components. The benefit of CHAT for studies such as this one is that it provides a way of understanding complex social activities in relation to the intentionality of their participants and within the contexts in which they occur.

The process of analysis itself involved interpreting the events in the case as CHAT activities. Data was coded using CHAT concepts through several iterations of qualitative coding with the NVivo [2] software package. During this process further codes were created as necessary based on an approach derived from ground theory (Glaser and Strauss, 1967). This coded data formed the basis for the insights that speak to the research question.

OGD curation at the city of Toronto

The City of Toronto was an early adherent to the open data and open government movements. Initial efforts to release OGD began in late 2009 at the behest of an enthusiastic mayor, and in the following years, the City actively collaborated with other Canadian municipalities in developing OGD programs (Giggey, 2012). However, after changes in political leadership and tumultuous political scandals, the City's progress in releasing valuable datasets and building their OGD program stalled. By the end of 2015 there was a growing perception amongst the open data community in Toronto that the development of the OGD program had reached a plateau.

Starting early in 2016, the topic of OGD began to appear as agenda items for public committees instigated by an advocate on City Council. These agenda items included deputations from the public, during which OGD advocates and members of civic tech community groups expressed dissatisfaction over the slow pace of data release, the limited functionality of the web portal, the limited description of the datasets available, the low value of the data that was being released, and the lack of a real mandate for the program. A frequent argument during these deputations was that the OGD the City was releasing had little value for innovation, political accountability, or commercial application.

The extent of the dissatisfaction amongst the public and the forcefulness of their advocacy appears to have caught elected officials off guard. In May 2016 following a series of public meetings, reports from civil servants, and deputations from advocates, a committee of elected officials issued executives in the administration with a mandate to develop a strategic plan to revitalize the City's OGD program.

This initiative forced the City administration to confront the shortcomings of their existing OGD curation efforts. To mitigate the dissatisfaction of their users, OGD curators endeavored to incorporate consideration for users' knowledge, skills, and use of OGD into the release process. In doing so, they had to navigate the central conflict of a Designated Community in this domain: the rationale behind the release of OGD precludes the existence of an identifiable group of users to whom curation services can be targeted.

From the start of the process, the development of the new strategic plan was guided and shaped by four principles that emerged during the public deputations. In brief, these principles were (City of Toronto, 2018): *co-development*, that the plan would be developed in partnership with the public; *civic issues*, priority for release would be given to datasets that spoke to civic challenges; *efficiency*, OGD release could and should facilitate collaboration across organizational units within the City administration; and *inclusivity*, efforts to release OGD and revitalize the City of Toronto's OGD program should be broadly inclusive and aim for data to usable by as wide and diverse an audience as possible.

On the basis of these principles, over the course of 2017 OGD curators at the City of Toronto undertook an elaborate process to develop what would be referred to as the Open Data Master Plan. The development process involved staff at the City beyond those directly involved in OGD release, along with members of the public and the communities with interests in OGD.

The first public events related to the Open Data Master Plan took place in March 2017 and drew heavily from the input received during the public deputations. Members of the open data community, academics, civic tech groups and civil servants from across the administration were invited to a day-long workshop to discuss the process of OGD release and how it could exist in the future. Much of the workshop was conversational, with some breakout sessions and visual aids used for brainstorming. Topics such as data quality and metrics for evaluating the program were discussed in depth. Towards the end of the month a second workshop brought together a larger number of people from the same groups and focused more heavily on interactive activities. These activities were designed to gather feedback on priorities for the new plan, the principles underpinning both its development and its content, OGD precedents set by other jurisdictions, and to identify civic challenges.

By June, the City of Toronto had partnered with Open North [3], a non-profit consulting firm and a steward of the International Open Data Charter, to co-develop the Open Data Master Plan and coordinate the involvement of stakeholders. The following month, the City took a further step to formalize its collaboration with the open data community by starting an Open Data Advisory Group. The group was composed of volunteers from outside of the administration and tasked with helping to steward input from the consultations into actionable components of the Open Data Master Plan.

Open North conducted a further public consultation in July, in which challenges, potential solutions, priorities, and barriers for OGD were discussed. The process of consultation continued through to September involving internal City staff in various roles, government representative from neighboring jurisdictions and the Province of Ontario, and external stakeholders from the non-profit sector, academics, and tech startups. These events were run like workshops and City staff used the International Open Data Charter to frame the conversations. In total 20 consultations were conducted, involving 125 stakeholders.

In conjunction with these consultations, and as a component of the Open Data Master Plan, City staff worked to create a new OGD portal. The focus of these efforts was on making OGD usable from the point of access, rather than merely accessible. This meant accommodating users with widely varying knowledge, skills, and interests. To achieve this

objective, a user-centered design approach was taken, and the portal was created through an iterative process based on the Lean methodology for start ups from Ries (2011).

Inspired by lean approaches used in manufacturing, the Lean methodology (Ries, 2011) was designed to provide entrepreneurs with a rigorous foundation for building and developing start ups. It centers on creating minimum viable products, for which testable predictions about their reception can be made and their performance in relation to these predictions evaluated. Products can then be iterated based on user feedback and behaviour. Through this process, the Lean methodology helps entrepreneurs to identify customer needs and to target their products and businesses to meet those needs. The methodology is intended to work independent of the domain or context in which the start up operates, but the methodology and its presentation are strongly oriented towards for-profit businesses in the technology sector. In its use in this case, the Lean methodology allowed City staff to begin with a fairly basic portal that focused on the essentials of providing access, developing it iteratively and building in additional features to meet the needs of various users as they evolve or emerge.

Interviews with members of the community, an online survey, and input from Open North were used to create a set of user personas that covered a wide range of potential portal users from businesses and non-profits to the public sector and interested individuals. These were organized into beginner, intermediate, and advanced groups based on their expected technical skills, knowledge, and familiarity with data analysis. The personas formed the basis for user stories from which specific portal features were derived. A beta version of the portal was available in September. It featured built-in data analysis and visualization tools from CKAN [4], Esri [5], and Tableau [6]. These tools were complemented by content that presented the data in a narrativized context (referred to as 'data stories') and advice to novice users on data analysis. For advanced users the portal supported on-demand format conversion for downloads and the generation of APIs on command. The Python scripts used in the built-in analysis were made available and City staff encouraged further contributions from users.

In December 2017 a draft of the Open Data Master Plan was released for a final round of consultation before it was submitted and approved by City Council in January 2018. It included provisions for revising the release process to make it more streamlined and automated, integrating it with other work processes in the municipal administration, and engagement and collaboration with external communities to promote access. Action points in the Open Data Master Plan were linked to specific performance measures and an implementation roadmap that outlined priorities and deliverables through to 2022. The new OGD portal, which was included as a deliverable in the roadmap, was launched in May 2018. All documentation and code related to the portal and the Open Data Master Plan were made available to the public so that the process could be transparent and other governments would have the opportunity to build on their work.

Findings and discussion

The research question that initiated this study asked:

> How does the curation of digital information function when there is no clearly identifiable Designated Community or user group towards which curation services can be targeted?

The curation of datasets for release as OGD fits this description, as the targeting of specific user groups is highly problematic (Janssen, 2012).

Those responsible for the curation of OGD responded to this situation through two complementary methods. The first was to draw from the discourse that permeates and defines the OGD domain. In the absence of a clear Designated Community and any comprehensive empirical information about them, OGD curators looked to the values and concepts in this area for guidance. These included the sets of principles for OGD, most notably those in the International Open Data Charter (2015). Collectively, these acted as a proxy for the requirements of a Designated Community and were used in place of an understanding of users more typically gained from direct input through methods such as those of Kärberg (2014) and Kim (2015).

The OGD discourse, in turn, had a profound impact on curation processes in this case. The concepts, principles, and values within the broader OGD discourse oriented City staff to collaborate with the open data community in the development of the Open Data Master Plan and framed the process of OGD release as an effort to make data as widely accessible and usable as possible. From this perspective, the revitalization of the OGD program, in particular the creation of a new OGD portal with much greater functionality, can be interpreted in part as an attempt to bring the program into greater alignment with evolving conversations in the open data and open government movements.

Extending the first method, the second method was to incorporate OGD advocates, civic tech groups, members of the open data community, and the public into the process of revitalizing the OGD program. From the initial public deputations, members of these groups were instrumental throughout the process. Contributions from the community were made through the various public consultations, the work of the Open Data Advisory Group, and by providing feedback at all stages. Working in this capacity, OGD advocates introduced and reinforced themes from the broader discourse and acted as a proxy for a wide spectrum of potential users.

There are several advantages to the two methods used in this case. Referencing the domain discourse builds a common understanding of the domain that can be shared between curators and users. This connects curators with larger communities of prospective users than they would have access to using previous empirical methods. As this case study has shown, the discourse in a domain can be used to help identify requirements for datasets that curation processes can then be tailored to fulfill. The evolution of the discourse will be easier to monitor than patterns in actual data use and will provide curators with a way of maintaining awareness of developments in the domain. Likewise, the on-going involvement of members of the community served by curation services will provide insights into data practices and increase buy-in from the wider public.

Seen from the perspective of this study, the challenge of curating digital information without the aid of a Designated Community and the specifics of the OGD domain led curators to an approach that echoes Huvila's (2008) participatory archive and Bettivia's (2016) suggestion of incorporating user input. Like in Huvila's case, and that of Shilton and Srinivasan (2008), this approach has at its core the involvement of users in curation processes, to which this case adds the involvement of data creators within the City administration. As a participatory initiative, it has parallels to some of Huvila's principles. For example, decentralized curation exists in the sharing of responsibility for the Open Data Master Plan amongst the curators, Open North, and the participating members of the community. While as an OGD initiative, it displays radical user orientation in that preservation is a secondary concern to facilitating access and usability. However, limited attention is paid to incorporating multiple dimensions of context, largely due to the heavy emphasis on describing the features and provenance of datasets to aid in their usability.

Although the curation in this case follows a participatory approach broadly in line with Huvila's (2008), it does not have an open Designated Community. That concept was developed in a tightly defined domain and in a case where contributions were restricted to authenticated expert users. This contrasts sharply with the open data and open government movements which aspire to a broader and more diverse potential audience and a case in which curators sought to enable barrier-free participation. The key aspect of this concept, placing the burden of providing contextual information on the community, conflicts with the demands from the public for better described and contextualized datasets which initiated the Open Data Master Plan development process. Shifting the responsibility for providing contextual information to the community would realizing Janssen's (2012) concern about releasing data for a small group of tech-savvy already-informed enthusiasts and undermine the inclusive aspirations behind OGD release.

At the same time, this case also differs slightly from Bettivia's (2016) suggestion of incorporating user input into the definition of a Designated Community. The Open Data Master Plan does not define a clearly delineated Designated Community in the conventional sense. Its conception of users is intentionally broad and drawn from the concepts and principles of the domain. Community input during the development process served to reinforce the concepts and principles from the OGD discourse and this inclusive conception of potential users for OGD.

These differences can be accounted for by seeing the development of the Open Data Master Plan as a reflection of the open data and open government movements. The co-development dimension has overtones of the "coproduction" described by Linders (2012) in that the community was involved at a fundamental level and shares responsibility for the outcome of the project. Likewise, the revitalization of the City of Toronto's OGD program along these lines brings it closer to the bidirectional information exchange described by Sieber and Johnson (2015). Lessons have been learned by the open data and open government movements from projects like that described by Both (2012). For example, the use of data stories for narrativized context, the visualization and preview functionality, and the provision of data analysis supported in the new OGD portal make progress in providing the guidance to users that Hellberg and Hedström (2015) found missing in their study. These portal features also speak to the distinction between accessibility and usability made by Gurstein (2011) with the emphasis placed on the latter.

There are, however, important criticisms to be made of the methods used in this case. Janssen (2012) raised the concern that OGD initiatives can become focused on the interests of a small subset of sophisticated and already engaged users, in effect undermining the ideals behind OGD release. The approach to creating the Open Data Master Plan was intended to be iterative and be repeated as part of the on-going development of the OGD program, together with the openness of the process, these factors mitigate, but do not nullify, Janssen's concerns. Counterbalancing the tendency to focus on elite advanced users will require a continuous push for openness and inclusivity, along with outreach to communities of potential users who may not already be engaged with OGD. Furthermore, there is a criticism to be made of the use of the Lean methodology for start ups (Ries, 2011). Currie (2016) has expressed concerns about the introduction of a neo-liberal ideology reminiscent of Silicon Valley into the public sector and this is a clear example of that trend. There was a noticeable lack of critical reflection on this point amongst the participants in the case.

Finally, this case presents interesting directions for further inquiry in research and for practice. In addition to providing an example of successful participatory program development and community collaboration, it raises the question of how the Designated Community concept can or should be used in participatory initiatives. The approach described here has the potential

to be used in other curation situations. This is one possible direction for organizations that have no clear Designated Community. However, the methods outlined here are resource intensive and require an identifiable domain discourse and close engagement with prospective user communities to be applicable. This may not be viable for some organizations. It is not clear, for example, how these methods could be applied by a national archives, unless it was done for a small or themed subset of their digital holdings. The involvement of external users may also face challenges of institutional resistance and professional hostility similar to those provoked by participatory archives initiatives (Flinn, 2010).

Conclusions

Organizations with mandates to curate or preserve digital information for diverse audiences present a challenge to conventional ways of conceptualizing digital curation and preservation. Exploring how this conflict was navigated in OGD curation at the City of Toronto revealed a dynamic participatory approach that has the potential to be applicable in other situations. The research described in this article contributes to the emerging discussion on the concept of a Designated Community and considers OGD curation from a new perspective.

Notes

1. www.youtube.com/channel/UCfe2rzOnQzgEDvNzRRPUJsA (accessed October 8, 2019).
2. www.qsrinternational.com/nvivo/home (access October 8, 2019).
3. www.opennorth.ca/ (accessed October 8, 2019).
4. https://ckan.org/ (accessed October 8, 2019).
5. https://esri.ca/en (accessed October 8, 2019).
6. www.tableau.com/ (accessed October 8, 2019).

References

Baker, K.S., Duerr, R.E. and Parsons, M.A. (2016), "Scientific knowledge mobilization: co-evolution of data products and designated communities", *International Journal of Digital Curation*, Vol. 10 No. 2, pp. 110-135.

Bates, J. (2014), "The strategic importance of information policy for the contemporary neoliberal state: the case of open government data in the United Kingdom", *Government Information Quarterly*, Vol. 31 No. 3, pp. 388-395.

Becker, C. (2014), "What is digital curation?", available at: http://dci.ischool.utoronto.ca/what-is-digital-curation/ (accessed 8 October 2019).

Bettivia, R.S. (2016), "The power of imaginary users: designated communities in the OAIS reference model", *Proceedings of the Association for Information Science and Technology*, Vol. 53 No. 1, pp. 1-9.

Borglund, E. and Engvall, T. (2014), "Open data?: data, information, document or record?", *Records Management Journal*, Vol. 24 No. 2, pp. 163-180.

Both, W. (2012), "Open data: what the citizens really want", *The Journal of Community Informatics*, Vol. 8 No. 2.

Brimble, S. Flinn, A. and Shepherd, E. (2016), "Team Europe EU32 project 2015-2016: the role of the records manager/records management in an open government environment in the UK: higher education (final report)", InterPARES Trust, Vancouver.

Brinkmann, S. (2014), "Unstructured and Semi-Structured interviewing", in Leavy, P. (Ed.), *The Oxford Handbook of Qualitative Research*, Oxford University Press.

Chorley, K. Flinn, A. and Shepherd, E. (2016), "Team Europe EU27 project 2015-2016: the role of the records manager/records management in an open government environment in the UK: the national health service 2 (local records) (final report)", InterPARES Trust, Vancouver.

City of Toronto (2018), "Open data master plan 2018-2022", available at: www.toronto.ca/legdocs/mmis/2018/ex/bgrd/backgroundfile-110740.pdf (accessed 8 October 2019).

Consultative Committee for Space Data Systems (CCSDS) (2012), "Reference model for an open archival information system (OAIS): recommended practice, (magenta book)",

Currie, M.E. (2016), "The data-fication of Openness - The practices and policies of open government data in Los Angeles", PhD Thesis, University of California, Los Angeles, Los Angeles, CA.

Dappert, A. and Farquhar, A. (2009), "Significance is in the eye of the stakeholder", *Proceedings of the 13th European Conference on Digital Libraries*, Corfu.

Davies, T.G. and Bawa, Z.A. (2012), "The promises and perils of open government data (OGD)", *The Journal of Community Informatics*, Vol. 8 No. 2.

Davies, T. and Frank, M. (2013), "There's No Such Thing As Raw Data': exploring the Socio-technical Life of a Government Dataset", *Proceedings of the 5th Annual ACM Web Science Conference*, ACM, New York, NY, pp. 75-78.

Denis, J. and Goëta, S. (2014), "Exploration, extraction and 'rawification'. the shaping of transparency in the back rooms of open data", The Shaping of Transparency in the Back Rooms of Open Data.

Ditsa, G.E.M. and Davis, J. (2000), "Activity theory as a theoretical foundation for information systems research", in Khosrow-Pour, M. (Ed.), *Challenges of Information Technology Management in the 21st Century*, IGI Global, pp. 240-244.

Engeström, Y. (1987), *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*, Orienta-Konsultit, Helsinki.

Evans, A.M. and Campos, A. (2013), "Open government initiatives: challenges of citizen participation: professional practice", *Journal of Policy Analysis and Management*, Vol. 32 No. 1, pp. 172-185.

Flinn, A. (2010), "An attack on professionalism and scholarship? Democratising archives and the production of knowledge", Ariadne 62, available at: www.ariadne.ac.uk/issue62/flinn

Galletta, A. (2013), *Mastering the Semi-Structured Interview and beyond: From Research Design to Analysis and Publication*, New York, NY University Press, New York, NY.

Giaretta, D. (2011), *Advanced Digital Preservation*, Springer, New York, NY.

Giaretta, D., Matthews, B., Bicarregui, J., Lambert, S., Guercio, M., Michetti, G. and Sawyer, D. (2009), "Significant properties, authenticity, provenance, representation information and OAIS information", *Proceedings of the Sixth International Conference on Preservation of Digital Objects (iPRES 2009)*, San Francisco, CA.

Giggey, R. (2012), "The G4: Setting city data free", Canadian Government Executive 17.

Glaser, B.G. and Strauss, A. (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, AldineTransaction, Piscataway, NJ.

Group of Eight (G8) (2013), "G8 open data charter and technical annex", available at: www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex (accessed 8 October 2019).

Gurstein, M. (2011), "Open data: empowering the empowered or effective data use for everyone?", *First Monday*, Vol. 16 No. 2, available at: http://firstmonday.org/ojs/index.php/fm/article/view/3316 (accessed 8 October 2019).

Harrison, E., Shepherd, E. and Flinn, A. (2015), *Team Europe EU19 Project 2014-15: A Research Report into Open Government Data in NHS England*, InterPARES Trust, Vancouver.

Hellberg, A.-S. and Hedström, K. (2015), "The story of the sixth myth of open data and open government", *Transforming Government: People, Process and Policy*, Vol. 9 No. 1, pp. 35-51.

Hodder, I. (2000), "The interpretation of documents and material culture", *Handbook of Qualitative Research*, Sage Publications, Thousand Oaks, CA, pp. 703-716.

Huvila, I. (2008), "Participatory archive: towards decentralized curation, radical user orientation and broader contextualisation of records management", *Archival Science*, Vol. 8 No. 1, pp. 15-36.

International open data charter (2008), "International Open Data Charter", available at: https://opendatacharter.net/ (accessed 8 October 2019).

Janssen, K. (2012), "Open government data and the right to information: opportunities and obstacles", *The Journal of Community Informatics*, Vol. 8.

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012), "Benefits, adoption barriers and myths of open data and open government", *Information Systems Management*, Vol. 29 No. 4, pp. 258-268.

Kärberg, T. (2014), "Digital preservation and knowledge in the public archives: for whom?", *Archives and Records*, Vol. 35 No. 2, pp. 126-143.

Kim, Y. (2015), "Designated communities': through the lens of the web", *International Journal of Digital Curation*, Vol. 10 No. 1, pp. 184-195.

Lee, D.J. (2015), "Research data curation practices in institutional repositories and data identifiers", PhD, The Florida State University.

Léveillé, V. and Timms, K. (2015), "Through a records management lens: creating a framework for trust in open government and open government information", *Canadian Journal of Information and Library Science*, Vol. 39 No. 2.

Linders, D. (2012), "From e-government to we-government: defining a typology for citizen coproduction in the age of social media", *Government Information Quarterly*, Vol. 29 No. 4, pp. 446-454.

McDonald, J. and Léveillé, V. (2014), "Whither the retention schedule in the era of big data and open data?", *Records Management Journal*, Vol. 24 No. 2, pp. 99-121.

Malamud, C. O'Reilly, T. Elin, G. Sifry, M. Holovaty, A. O'Neil, D.X. Migurski, M. Allen, S. Tauberer, J. Lessig, L. Newman, D. Geraci, J. Bender, E. Steinberg, T. Moore, D. Shaw, D. Needham, J. Hardi, J. Zuckerman, E. Palmer, G. Taylor, J. Horowitz, B. Exley, Z. Fogel, K. Dale, M. Hall, J.L. Hofmann, M. Orban, D. Fitzpatrick, W. and Swartz, A. (2007), "8 Principles for open government data", available at: https://public.resource.org/8_principles.html (accessed 8 October 2019).

Nardi, B.A. (1996), *Context and Consciousness: Activity Theory and Human-Computer Interaction*, MIT Press, Cambridge, MA.

Open Knowledge Foundation (2019), "Open government data", available at: http://opengovernmentdata.org/ (accessed 8 October 2019).

Page, J., Flinn, A. and Shepherd, E. (2014), *Team Europe EU03: The Role of the Records Manager in an Open Government Environment in the UK (Final Report)*, InterPARES Trust, Vancouver.

Parsons, M.A. and Duerr, R. (2005), "Designating user communities for scientific data: challenges and solutions", *Data Science Journal*, Vol. 4, pp. 31-38.

Ries, E. (2011), *The Lean Startup: how Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Crown Business, New York, NY.

Ross, S. (2006), "Approaching digital preservation holistically", in Tough, A. and Moss, M. (Eds), *Information Management and Preservation*, Chandos Press, Oxford.

Shepherd, E., Bunn, J., Flinn, A., Lomas, E., Sexton, A., Brimble, S., Chorley, K., Harrison, E., Lowry, J. and Page, J. (2019), "Open government data: critical information management perspectives", *Records Management Journal*, Vol. 29 Nos 1/2, pp. 152-167.

Shilton, K. and Srinivasan, R. (2008), "Participatory appraisal and arrangement for multicultural archival collections", *Archivaria*, Vol. 63, pp. 87-101.

Sieber, R.E. and Johnson, P.A. (2015), "Civic open data at a crossroads: dominant models and current challenges", *Government Information Quarterly*, Vol. 32 No. 3, pp. 308-315.

Suderman, J. and Timms, K. (2016), *NA08 the Implications of Open Government, Open Data, and Big Data on the Management of Digital Records in an Online Environment (Final Report No. Version 1)*, InterPARES Trust, Vancouver, Canada.

Suderman, J., Timms, K., Léveillé, V., Hurley, G., Rovegno, K. and McDonald, J. (2015), *NA08 Status Report*, InterPARES Trust, Vancouver.

Vygotsky, L.S. (1971), *The Psychology of Art*, MITPress, Cambridge, MA.

Vygotsky, L.S. (1978), *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, MA.

Wilson, T.D. (2006), "A re-examination of information seeking behaviour in the context of activity theory", *Information Research*, Vol. 11 No. 4.

Yakel, E. (2007), "Digital curation", *OCLC Systems and Services: International Digital Library Perspectives*, Vol. 23, pp. 335-340.

Yakel, E., Shaw, S. and Reynolds, P. (2007), "Creating the next generation of Archival Finding aids", *D-Lib Magazine*, Vol. 13 Nos 5/6.

Yin, R.K. (2009), *Case Study Research: Design and Methods*, (4th ed.), Sage Publications, Los Angeles, CA.

Yu, H. and Robinson, D.G. (2012), "The new ambiguity of open government", *UCLA L. Rev. Discourse*, Vol. 59, pp. 178-208.

Corresponding author
Nathan Moles can be contacted at: n.moles@mail.utoronto.ca