InterPARES Trust

InterPARES Trust
NA08 Status Report
November 15, 2015
Vancouver, BC

NA08: Open Data, Open Government and Big Data: Implications for the Management of Records in an Online Environment

Researchers:
Jim Suderman, Katherine Timms, Valerie Léveillé, Grant Hurley, Kelly Rovegno, John McDonald

This status update briefly reviews the key findings and conclusions of the first two phases of this study. The report goes on to set out the framework of analysis for the current, third phase, which builds on the previous two, and summarizes the information gathered to date within that framework.

First Phase

The findings from the first phase of the study were published in an article by John McDonald and Valerie Léveillé entitled "Whither the retention schedule in the era of big data and open data?" in *Records Management Journal*, Vol 24, #2 (2014), pp. 99-121.

The article explores the conceptual evolution of open data and big data, noting some of the differences between the two, such as:

- Open data initiatives generally focus on existing datasets while big data initiatives often create new datasets (mashups);
- Communities supporting each type of initiative differ, with open data communities more oriented to public policy-making and governmental transparency, while big data communities seek new knowledge or an economic edge.

The article asserts that both big data and open data initiatives share the characteristic of being based on some form of business process, which is essential to setting appropriate retention and disposition rules, a task which is itself an evolving process as a result of the transition from physical to digital recordkeeping. In addition, records managers have a leading role in establishing retention and disposition rules that are implementable and relevant within the organization and to the communities served by data both big and open.

The article uses a fictitious business organization to illustrate the organizational dynamics affecting retention management of data, ranging from the priority of publishing open data to a 'Why not keep everything?' approach. Concluding that retaining everything was neither desirable nor technologically feasible, the illustration points to the ISO Technical Report, "Information and documentation – Work process analysis for records" (ISO/TR 26122:2008) as a guide to analyze the relevant business processes.

The study noted that a business process does not start and stop with the creation and dissemination of a dataset. It extends back into the processes generating the source data (which could come from a variety of sources) that contributed to the final data set.

It also includes the steps involved in administering the processes themselves such as the approval of data collection methods, conduct of the collection/survey, measurement of data quality, etc., all of which generate records in the form of emails, reports, etc.

The study concludes that retention and disposition specifications need:

- To account for those records that are necessary to ensure the accuracy and authenticity of the data generated to support big data and open data initiatives; these can include project management documentation, data collection procedures, metadata models and related data classification documentation, survey design documentation, data quality assessment procedures and tests, and related documentation, as well as key versions of data used to generate master files, public use data files, and other data made available through big data and open data initiatives.
- To have been developed in an integrated fashion as part of the steps involved in designing and implementing the systems, processes and technologies that support big data and open data initiatives.

As a result, it was not seen as sufficient for the records manager to simply identify the retention period, but also to set out the specifications for retention in a comprehensive manner that accounts for all related processes. Concluding observations emphasized the need for a more robust recordkeeping framework that accommodates data generated both internally and externally.

Second Phase

The findings from the second phase of the study were published in an article by Valerie Léveillé and Katherine Timms entitled "Through a Records Management Lens: Creating a Framework for Trust in Open Government and Open Government Information" in *Canadian Journal of Information and Library Science*, Vol 39, #2 (June, 2015), pp. 154-190. The article proposes a universal framework to ensure the accuracy, authenticity, and trustworthiness of information through the analysis of business processes, documentation, and structures supporting open government initiatives.

The second phase of the study began by defining the concepts of open government and open government data and information in the Canadian context. The study concluded by considering the impact of these two concepts on the creation, management and control of government records through business process analysis as defined by ISO/TR 26122:2008(E), the same standard relied upon in the first phase of the study, detailed above.

Open government, a concept evolved from e-government, emphasizes service delivery and the public's right to government information. The objectives of the open government ideology are therefore to increase transparency of government processes and decision-making, increase accountability of government to the public, and increase citizen participation or engagement.

Open government information, (for example, government data sets, open information, metadata components, etc.) falls under a broader 'open' umbrella, including open source and open access, all of which broadly contribute to open knowledge: "any content, information, or data that people are free to use, re-use and redistribute – without any legal, technological or social restriction."[1] The goal of open government information initiatives, and thus of open government initiatives, is to foster increased trust between government and citizens.

Following a survey of the Canadian open government landscape at the municipal, provincial and federal levels, the article considers open government as a three-staged business process:
1. Initiation;
2. Identification and distribution;
3. Promotion and evaluation.

A records management framework emerges from an open government policy that includes records management components, either integral to the policy or by reference to relevant records management policies. To give strength to the policy, responsibility for its implementation must be appropriately assigned organizationally or overseen by records professionals collaborating with authorities administering access to information, internal regulation, and audit functions.

An enterprise-wide inventory of government information provides a starting point for determining what information – including relevant contextual metadata – is eligible for publishing. While new technologies may have to be deployed to make this information open and accessible to the public, 'feedback loops' should also be established to allow the open government authority to determine whether citizens are aware of available information and whether or not this information is useful to them.

Six categories of 'information objects' result from open government initiatives, including those that:
1. are generated to initiate and sustain the program, e.g., action plans;
2. serve as inputs to the process of releasing information, e.g., candidate data sets;
3. are outputs of the processing of releasing or publishing information;
4. are platforms for accessing released information, which may contain, for example, metadata of released information;
5. are inputs from citizens, e.g., completed feedback forms;
6. are related to the initiative(s), e.g., reports, presentations.

Maintaining this documentation is an important records management task that provides an authoritative, complete, and accurate source to demonstrate the integrity and reliability of information published as part of the open government or open data initiative.

The article concludes with a brief assessment of Canadian jurisdictions against the framework outlined, highlighting:

---

[1] Open Knowledge, "What is Open?" (https://okfn.org/opendata/, checked 8 November 2015), quoted in CJILS article on p. 158.

- a general need for more substantive integration of open government with records management;
- better citizen engagement mechanisms, including feedback loops;
- meaningful metrics to evaluate open government initiatives.

Third (Current) Phase

With the term 'open government' in danger of meaning anything to anybody, the study adopted the four core principles of the Open Government Partnership as the scope for the term:
- Accountability
- Technology & Innovation
- Citizen Participation
- Transparency[2]

The study also adopted the five engagement categories described by the International Association for Public Participation's (IAP2) engagement spectrum:
- Inform
- Consult
- Involve
- Collaborate
- Empower.[3]

This current phase of the study explores citizen engagement or participation initiatives in the Canadian context with the purpose of gaining a better understanding of how citizen engagement affects approaches to records and information management. Citizen engagement actively involves citizens in policy or program development at any level of government. Citizen engagement processes and initiatives seek to establish a more collective basis of decision-making power by seeking input from citizens much earlier in the process of drafting legislation, for example. Though the level and opportunities for citizen input and technologies used may differ between different engagement initiatives, knowing the contexts, functions, and technologies deployed will enable more informed recordkeeping.

Information was gathered from the Canadian federal government, the provincial governments of Alberta, British Columbia, and Ontario, and the city governments of Toronto and Vancouver through semi-structured interviews with open government leaders in each of those jurisdictions. We also asked interviewees to point us to specific citizen engagement initiatives they felt were particularly effective, innovative, collaborative or empowering.

Findings have been grouped under five context-based headings identified by the InterPARES research project as essential to preserve the identity and integrity of digital records.[4]

---

[2] http://www.opengovpartnership.org/about/open-government-declaration
[3] http://c.ymcdn.com/sites/www.iap2.org/resource/resmgr/imported/IAP2%20Spectrum_vertical.pdf

1. The juridical or administrative context, defined by InterPARES as the "legal and organizational system in which the [record-]creating body belongs";
2. The provenancial context is that of the "creating body, its mandate, structure and functions";
3. The procedural context refers to the processes that result in created records (the "business procedure in the course of which a record is created");
4. The technological context are the "characteristics of the hardware, software, and other components of an electronic computing system in which records are created";
5. The documentary context is the "archival fonds to which a record belongs, and its internal structure".

Juridical / Administrative context

Generally speaking, there are three common juridical contexts within Canada:
- the federal or national context;
- the context of each of the provinces or territories;
- the context of every municipality.

These contexts influence and are influenced by each other as well as by other juridical environments, for example, international organizations. And although most Canadians act simultaneously in all three juridical contexts, the governmental entities in each case operate their own open government initiatives, with limited integration across jurisdictions.

For example, we noted that similar but not identical open government licenses were prepared by the federal government and by the provincial and municipal governments we contacted. For example, *Open Government Licence – Toronto* is "based on version 1.0 of the Open Government Licence – Ontario, which was developed through public consultation and a collaborative effort by the provincial and federal government." [5] Besides replacing references to the Province of Ontario with the City of Toronto, the latter includes "a provision for the Ontario Personal Health Information Protection Act, 2004."[6]

We also noted that leadership at all levels of government is assigned to senior elected officials suggesting that open government initiatives are seen as important to legitimizing the authority of elected governments. For example, the President of the Treasury Board, an elected official,[7] has leadership at the federal level, while the Deputy Premier, also an elected official, has leadership of the province of Ontario's open government initiatives. Toronto's City Council established "Open Government by Design" and "Engage the Public" as two strategic directions for the City, assigning

---

[4] Duranti, Luciana, and Pat Franks, eds., "Context" entry *Encyclopedia of Archival Science* (2015), p. 152. Quoted definitions are taken from the InterPARES 2 Glossary, (accessed 8 November 2015).
[5] Toronto. "Open Government Licence – Toronto" (accessed 8 Nov 2015).
[6] *Ibid.*
[7] Assignment of this responsibility may change following the recent election.

leadership of the former to the City Clerk and of the latter to the City Manager, both central officers of the City administration.[8]

Provenancial context

Interviewees from the federal government observed that that most engagement initiatives are conducted by individual departments. They were aware of few multi-departmental, collaborative consultations, commenting that there is no culture of sharing between departments, partly as a result of privacy concerns and partly because doing so requires planning and guidance which is absent. One result of this situation is that separate engagement initiatives conducted by different departments may go to the same stakeholders with similar requests. Many Canadians are not aware of the roles and responsibilities of different levels of government. One interviewee commented that web renewal plans include 'scoping down' online content to be more citizen-centric and less organization dependent. This has advantages for the citizen but may obscure provenance.

It also appears that provenancial context is complex at least for engagements falling into the 'Collaborate' and 'Empower' categories of the IAP2 spectrum. In these cases there is more than one creating body, each with its own mandate, structure and functions presumably engaging as more or less equal partners and where recordkeeping requirements and practices may vary substantially. We were asked who the steward was for information collected by both government and non-government groups. Establishing clear stewardship for such records is crucial to determining what legal obligations and rights pertain to them.

Procedural context

Preliminary study findings indicate a recent trend towards centralization of organizational policy, performance measures, and practices, very likely linked to the centralization of responsibility for open government initiatives, including their citizen engagement components, noted previously. Nevertheless, citizen engagement policies and procedures appear to be still very much a work in progress.

- The liquor policy review in British Columbia was the only instance encountered that had a clearly established engagement process, although it appears that process was unique to this consultation and did not follow a well-established, common process.
- "Growing Conversations," a Toronto engagement initiative to enhance citizen engagement practices:
    - instituted a "Conversation Corps" (drawing from practice in Austin, Texas);[9]

---

[8]

https://www1.toronto.ca/City%20Of%20Toronto/City%20Manager's%20Office/City%20Manager%20Profile/City%20Initiatives/Strategic%20Actions%20for%202013%20to%202018_FINAL.pdf

[9] This program allows individuals (e.g. private citizens) to sign-up as facilitators to lead consultations on topics pertaining to government business.

- produced a guide prepared by the City Planning division.
- Communication and consultation guides exist in some departments in the Canadian government, and may be sufficient to govern simpler types of engagement.
- In Vancouver there is currently no suite of policies for open government, nor are there formal, mapped-out workflows for open data. The development of policy pieces, guidelines, etc. – including those for citizen engagement – will be addressed in the next phase of the program.
- The absence of whole-of-government guidance on conducting consultations inhibits coherence, consistency and the development of standards. The absence of standards, guides, etc. was recognized as a major issue by several interviewees (e.g. TBS and PCO).

Interviewees identified even fewer performance measures.
- One jurisdiction, having conducted an environmental scan, concluded that there were not many open government jurisdictions with mature performance measurement frameworks against which to benchmark.
- It is hard to establish effectiveness measures even for the simplest type of engagement on the IAP2 spectrum, 'Inform'. Measures may be increasingly difficult to establish for the more complex types of engagement.

Technological context

All citizen engagement initiatives identified used some kind of web-based technology to inform citizens, collect input, or offer resources related to open government, such as open data. Besides conventional technologies, such as meeting rooms and snail mail, citizen engagement technologies ranged from:
- Uses of commonly available web technologies, such as blogs, e-mail and surveys;
- Third-party social media and collaboration platforms, such as Twitter, LinkedIn, and Google Docs;
- Adapted third-party social media platforms built for citizen engagement, such as MindMixer (now mySidewalk) which supports Toronto's IdeaSpaceTO – a place where "you can apply your ideas, knowledge & expertise to help shape the future of Toronto [and] explore and comment on the ideas of others within the community."[10]
- Data access platforms built in-house, such as the CKAN-based (open source) portal planned by Alberta, where people will be able to search for information and data over many databases.

What constitutes an effective technology depends on the type of engagement initiative. A blog may provide better support for a collaborative engagement than an email blast, for example. However, engagement may depend more on a suite of technologies, as is evident from the statistics for the consultative review of BC's Liquor Policy, which used e-mails, Twitter, and blog posts to encourage feedback. The trend appears to be moving toward more substantial uses of technologies to encourage feedback and conversation from many different sources, or as in the case of Toronto's immersive engagement platform, the deployment of technologies specifically intended for citizen engagement processes.

---

[10] Toronto. Ideaspace home page. (Accessed 8 November 2015).

Interviewees observed that:
- "Technologies work best if you're human and thoughtful – trust is mediated by technology, not created by it."
- "We need to be a little more human with our data. Open data and civic engagement is more about a conversation, being frank."

The study also noted that while a range of technologies may play a key role in engaging citizens, initiatives that encourage inputs from many sources may have the effect of fragmenting the record of engagement, where records documenting inputs may be spread across many platforms, some of them outside of the control of recordkeepers. Furthermore, it may not be clear how citizen feedback is actually used to inform the end result, and if further technologies have been used behind the scenes for analysis and decision-making.

Documentary context

That records are essential to effective citizen engagement initiatives is hardly a surprising conclusion. Using technologies that enable citizens to create records themselves seems to contribute positively to engagement. Individual choice, and the limitations set by the engagement initiative as to the channel(s) influences the type, e.g., a free text comment to a blog post, a 're-tweet', can significantly affect the context of the records created, and whether these records are accessible into the future.

As noted in the above, the record of an engagement initiative may be fragmented as a result of different provenancial or technological contexts. Detailing that contextual information in such a way that is intelligible to individuals examining the records subsequent to the initiative will be a significant challenge to records keepers. But they may not even be aware of related contexts, which may create uncertainty regarding where the complete record resides and whether fragments held in different places can be understood as a cohesive whole.

A further consideration may be the existence of records that are not relevant to the engagement initiative but that come into existence as a result of a technological platform that allows 'free' posting. For example, even if the technological context is that of the government, not all records held on that technological platform may be addressed to or authored by it; two participants may conduct a sidebar conversation in the same discussion forum intended for a particular issue.

Even basic records like a report resulting from an engagement initiative may have significant variations. One interviewee advised that the composition of reports on consultations will range from verbatim to highly synthesized – even sanitized – content.

Questions arising

It is clear that citizen engagement activities within open government initiatives are still at an early stage of maturity. This is an important consideration for recordkeeping, as recordkeeping strategies have to be mapped according to the level of maturity.

Records are created to fulfil one or more purposes within open government initiatives, but not all those purposes may have equal weight. Where establishing accountability is paramount, strong custody and control mechanisms relating to both record creation and maintenance activities are in place. By contrast, it seems that where broad citizen engagement is paramount, custodial and control mechanisms for records may end up being looser. This may be due to flexibility in terms of technology and type of engagement, in perhaps an inverse relationship to the ability to demonstrate accountability since the immediate priority is on citizen engagement rather than demonstrating accountability through documentary evidence.

Under what circumstances will assembling a comprehensive record of past engagement initiatives be required? Will inconsistencies in terms of procedural norms and the wide range of technological platforms underlying the records make it impossible to determine whether the records are complete and authentic? Will that affect trust? What do the answers to these questions say about the completeness and authenticity of records generated in current engagement initiatives?

The Historical Study of Cloud-based Services (ITrust, NA11) concluded that "trust in the service provider is a far greater consideration than trust in the technology." Building trust beyond a basic level is challenged due to differences in priorities:
- Elected officials put politics first
- Public servants put policy first
- Citizens tend to see no difference between the political side and the public service. And what will they put first?

Next Steps

This phase of the study will prepare at least two products:
1. A report of research conducted to date, addressed to the ITrust researchers.
2. A primer or guide specific to citizen engagement initiatives addressed to recordkeeping professionals (and, possibly, leaders of citizen engagement initiatives).

Current work towards these products is being conducted as follows:
- Development of 'the 5 x 5', an assessment of the five types of engagements set out in the IAP2 spectrum against the five diplomatic contexts identified in earlier phases of InterPARES research work: Juridical/Administrative, Provenancial, Procedural, Technological, and Documentary.
- Examination of the specific role of technology, starting from information collected in this phase of the study. It is possible that this will merge with the preceding work.
- Analysis of two citizen engagement initiatives identified by the BC government (a provincial services card; a review of provincial liquor policy). Depending on our success in these two

cases – over which study participants have limited control – additional cases will be examined. These in-depth studies will be used to assess priorities and observations collected through interviews as well as providing an excellent source of illustrative examples for both of the study products, detailed above.