

Designing Sustainable Data Archives: Comparing Sustainability Frameworks

Kristin R. Eschenfelder¹, Kalpana Shankar²

¹University of Wisconsin-Madison

²University College Dublin

Abstract

This theory review paper argues that in order to ensure the longevity of data, we need a better understanding of the sustainability of institutions that steward data. The paper considers what sustainability means in relation to data archives. It compares five frameworks that inform the concept of sustainability in order to develop a more complex understanding of the concept of sustainability. The resulting conceptualizations of sustainability can aid data archive stakeholders, designers and analysts in making decisions about how to develop “sustainable” data institutions.

Keywords: data archive; sustainability; organizational change; information infrastructure; organizational resiliency

doi: 10.9776/16243

Copyright: Copyright is held by the authors.

Acknowledgements: This research was funded by the Alfred P. Sloan Foundation

Contact: Corresponding author: eschenfelder@wisc.edu

1 Introduction

The longevity of data archives (DA) is a growing concern as researchers, archival practitioners, and funders of DA projects seek to ensure that resources invested will have benefits that endure beyond the period of original research funding. Although there has been significant research into the preservation of the data themselves, there has been less attention paid to the sustainability of the institutions that curate the data. For data to remain accessible over time, the data repository organization and its services which preserve, organize and provide access to data must themselves be sustainable (Knowledge Exchange, 2014; National Academy of Sciences, 2014).

In this paper, we ask how theoretical perspectives on organizational sustainability can inform the design or analysis of information institutions like data archives. We begin with the field of information science and information practice to set a baseline understanding of how data sustainability is framed. To develop a deeper understanding of data institution sustainability, we then compare five theoretical frameworks. All of the frameworks enrich understanding of the concept of sustainability; however, not all of them use the term “sustainability.” We compare the frameworks’ structures and point out differences in how they depict sustainability as a concept. We also draw out the different emphases of the frameworks including internal skill capacities, environmental monitoring, the turbulence of external environments, governance and relationships, or changes in the scientific communities and their data. For the purposes of this paper, our inquiry is limited to frameworks that focus on the longevity of organizations or services, rather than those that examine ecological sustainability or broader societal issues (e.g. Chowdhury, 2012). We conclude with a brief discussion of how data archive sustainability can be discussed in more nuanced ways than the current literature suggests and brief implications for designing data services that persist over time.

2 How does information science talk about sustainability?

We began with a scan of the literature with the keywords “digital library” and “sustainability” from library and information studies (LIS) databases (“data archives” proved to be too narrow a topic), focusing on the time period between 2001 and Summer 2013 (when this project was concluded). We removed articles that were not from information science journals. We supplemented these with other white papers and reports collected as part of our project. This resulted in 45 articles. We conducted an inductive analysis to identify themes (described in Table 1). Because our analysis was inductive, we do not report inter-coder reliability or frequencies, but Table 1 describes prevalence from more to less common.

Topic	Description
Financial (Most prominent)	Articles discussed possible sources of revenue, possible business models, fundraising and donations, fees for use and pricing, licenses and subscriptions, contracts and educational fee for service.
Relationships	The importance of developing and maintaining relationships. Articles discussed relationships with granting agencies, suppliers of content, journals,

	host institutions, and other partner organizations.
Valued Services	The importance of providing services that are clearly valuable to the community of users and other stakeholders to sustainability. This included remaining up to date with changing user expectations about services.
Standards	The need to remain compliant with all technology standards to ensure content sustainability.
Accountability	The importance of accountability and metrics in order to show success and impact to stakeholders
Knowledge/Skills	Many articles mentioned staffing issues in relation to finances, but some articles discussed sustainability in terms of staff knowledge and capacities including for example the importance of entrepreneurial leadership and the need to “professionalize” staff as projects move away from their startup phase.
Legal	Rights management, legal compliance with intellectual property and privacy laws
Disaster Planning (Least prominent)	Planning for disasters in order to ensure sustainability.

Table 1: Sustainability Themes from the Practice Literature Listed from Most to Least Common

In summary, the most prominent themes from the information science literature were financial resources, the importance of relationships and building services valued by stakeholders. Themes such as standards compliance, accountability and staff knowledge/capacities, legal issues and disaster planning appeared but were not as prominent.

3 Sustainability Frameworks:

In this section we compare five frameworks that enhance our understanding of DA sustainability developed from the brief inductive literature review outlined above. Given space constraints, we only describe each framework’s most important contribution to our understanding of sustainability.

3.1 The Sustainability Index

The Sustainability Index (SI) was developed in part to guide fledgling open-information/data institutions toward long term sustainability (Knowledge Exchange, 2014). It employs a grid framework (see Figure 1) and it depicts five stages of development, with five being the most sustainable.

	Low sustainability				High sustainability
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Funding management skills	Fundraising expertise to source seed capital	Operational funding in place. Budget planning and development expertise	Development funding in place. Budget deployment and management skills	Resilient funding plans to cover all business functions for the medium term	Resilient funding plans to cover all business functions for long term (5-10 years)
Business planning skills	Initial business plan in place	Market awareness; horizon scanning	Embedding adaptability to market changes	Formalised strategic planning process	Visionary management alert to internal and external opportunities
Business operational management skills	Basic facilities management	Appropriate human resources management	Crystallising a suitable organisational structure and culture	Settled, effective organisational workflows/systems	Professional management of facilities and human resources
Business development skills	Initial market research	Accurate market segmentation and development of value proposition	Marketing mix planning (including price, distribution, promotion and service development)	Continued development of effective sales approach, systems and workforce	Robust and adaptive sales and marketing strategies with achievable targets
Financial management skills	Basic financial management	Appropriate financial systems, including cost control	Provision of quality management accounting	Professional-level financial management	Appropriate financial strategy and systems with relevant targets
Technical development skills	Minimum skills required to provide the initial service/product	Higher order depth and breadth of skills	Appropriate management skills for managing development and expansion, including skills in outsourcing	Manage capacity for skills-based organisational growth	Settled as a forward-looking, innovative learning organisation capable of managing creativity
Legal skills	Ad hoc purchase of external expertise or rely on skills of host organisation (e.g. university)	Contracts for customers and staff become normalised	Processes are in place to adhere to relevant laws and regulations governing the enterprise	Settled arrangements are in place with a regular source of legal expertise	Periodic reviews are undertaken to ensure legislative compliance and efficacy of legal processes and paperwork
Policy awareness	Basic research to understand government, funder and other policies relevant to the enterprise	Invest in desk and field research to maintain current awareness of relevant policies	Sufficiently well established to be invited to policy development events	A network of contacts to enable early awareness of relevant policy shifts	Effective policy influencing capabilities
Governance system	At this stage governance systems are often ad hoc and informal	The need for appropriate systems of governance become clear	Effective governance systems evolve as the organisation shows it has the potential to grow	A formal governance system is established and effective in monitoring and advising	Periodic reviews of the appropriateness and effectiveness of the governance system are conducted
Organisational structure and interdependencies	Initial structure may not be expedient, not optimal	Early growth may highlight limitations of current structure	Structure is flexible enough to permit growth	Settled and effective organisational structure	Review effectiveness of structure regularly

Figure 1: Sustainability Index (Knowledge Exchange, 2014)

The SI lists target skills needed at each stage from basic skills (level 1) through advanced skills (level 5). For example, under funding expertise, level 1 is depicted as “fundraising expertise to source seed capital” and level 5 is “resilient business plans to cover all organizational functions for the long term.”

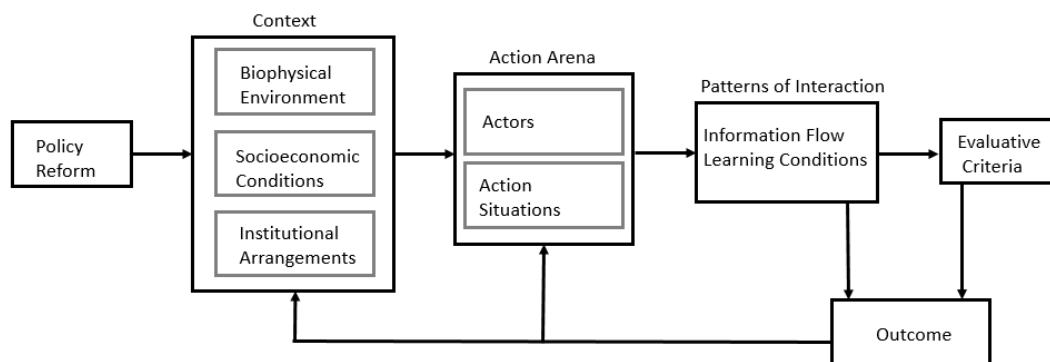
One contribution of the SI to our understanding of sustainability is its focus on internal capacities that organizations ought to cultivate to achieve “High Sustainability.” Four of the SI’s ten skill areas are related to financial/business management; and as we noted, this was a dominant area of concern from our analysis of the information sciences practice literature. However, the SI’s skill list also draws attention to topics that were not as prevalent in our literature analysis including the need to develop capacities for governance and legal and policy knowledge.

One limitation of the SI’s stage model structure is that it implies linear progress via stages. The framework’s grid format cannot easily accommodate depiction of self-learning feedback loops or relationships with external actors (although these are implied in the text of the Index).

3.2 Institutional Analysis and Design Framework

The Institutional Analysis and Development Framework (IAD) delves more deeply into how actors organize themselves to manage commons resources to ensure their sustainability (Ostrom & Hess 2011). Hess and Ostrom define commons resources as “shared by a group of people subject to social dilemmas.” They stress that knowledge commons are not synonymous with open access, and that in many cases their resources are only shared by some people for some uses (2011). Sustainable knowledge commons are those that meet user needs without “compromising the ability of future generations” to use the resource and (Ostrom & Hess pg 63). Sustainability is a process that requires monitoring, evaluation, and adjustments.

The IAD begins to change understanding of sustainability. As seen in Figure 2, the IAD is arranged as a feedback loop process. This shifts our understanding of sustainability from something that can be achieved by progressing in a linear fashion from stage to stage (i.e., the SI) to something that must be maintained through monitoring and constant adjustment.



Institutional Analysis and Development Framework from Ostrom, E., Hess, C. (2011). A Framework for Analyzing the Knowledge Commons, in Hess and Ostrom (Eds.) *Understanding Knowledge as a Commons* (pp. 41–82). Cambridge MA: MIT Press.

Figure 2: Institutional Analysis and Development Framework

The IAD asks how the interaction of stakeholders through governance influences the maintenance of sustainable commons resources. The IAD fleshes out the SI’s call for formal governance systems that provide “advice and monitoring,” and provides structure for DA designers to create or analyze governance structures including rules, processes, sanctions and evaluation criteria. As depicted in Figure 2, stakeholders in “action arenas” interact create, evaluate and modify governance structures including rules, the processes by which use rules are created, and rule enforcement and sanctioning practices.

From an IAD perspective, in order to remain sustainable, DA would be continuously adjusting their policies and practices in light of performance against agreed upon criteria, stakeholder’s use of and contribution to the pooled resources, and changing environmental, socioeconomic and institutional conditions.

3.3 Organizational Resilience

Scholars of organizational resilience are interested in how organizations maintain functionality over time by detecting threats and adapting to changing conditions. Resilience is an organization's ability "to return to a stable state after a disruption." (Bhamra & Burnard, 2011: p 5376), and it is a function of both an organization's level of exposure to disruptive events, and the capacity of an organization to make internal adjustments to cope with disruptions.

Similar to the IAD, the Organizational Resiliency (OR) framework (see Figure 3) depicts resilience in terms of processes with feedback loops involving organizational learning. But similar to the SI, the OR framework draws attention to the need for internal capacities required to: detect potential threats (detection and activation), be self-aware about weaknesses in relation to potential threats (response detection), and to quickly respond.

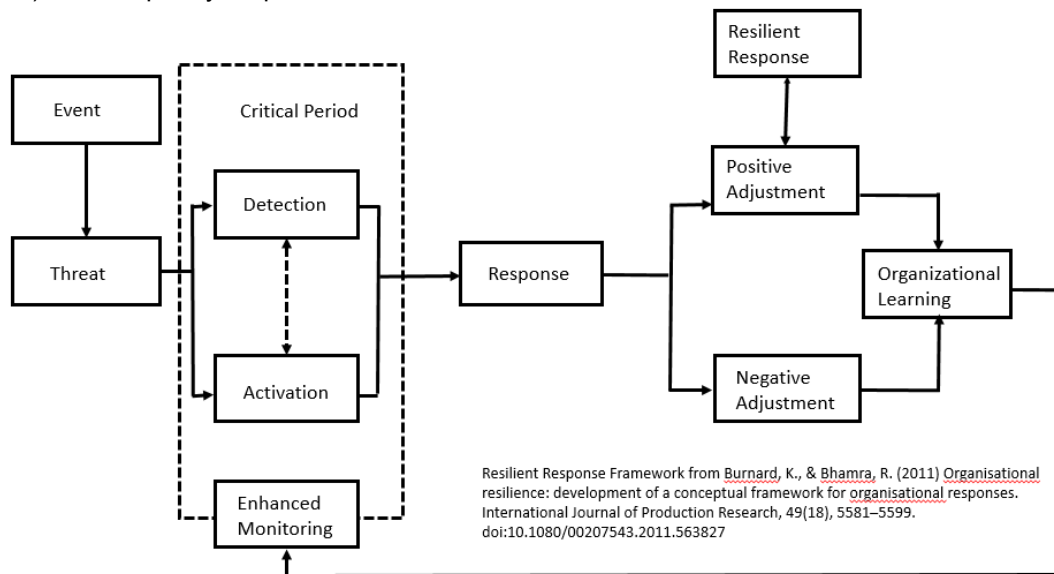


Figure 3: Resiliency response framework (Burnard and Bhamra, 2011)

The OR framework points to how DA environments may vary in the degree to which they present disruptions – both across DA and over time. Sustainability is influenced by the capacities of a given DA to scan the environment to detect changes, recognize threats and opportunities and take action in response. The SI similarly called for development of skills in “horizon scanning” and “adaptability to organizational change,” but the OR framework links the capacity to scan and take action to an organization's ability to adapt and therefore be resilient.

Another complication about sustainability is brought out by the OR framework's assumption that organizations can return to a “steady state” if they successfully overcome a disruption. Is there a steady state of sustainability which organizations can achieve (i.e., Stage 5 of the Sustainability Index?) Can organizations “slip” from being sustainable, but regain this steady state if they are resilient? We return to this question at the end of the paper.

3.4 Infrastructure Studies: Project Flexibility

The Project Flexibility (PF) framework was developed with the goal of encouraging development of sustainable long-term science infrastructures and ask what it means for a scientific project to be “flexible” in order to remain viable over the long term. Ribes and Polk suggest that flexibility is not an attribute of a project, but rather an attribute of the project's relationship to something else (2014). Said in another way, one should not state that a project is flexible, but rather that some aspect of the project and its relationship to thing X is flexible. For example, the project's metadata may be flexible in relation to the changing nature of data. The same logic can be applied to thinking about sustainability. Sustainability is not an attribute of a digital archive, it is an attribute of a digital archive's relationship to something else.

The PF framework considers science data project flexibility in relation to a number of changes that are similar to the “disruptions” suggested by the OR framework. The flexibility framework points to: changes in the objects of investigation or study participants, innovations in methodologies and instruments, changes in research priorities and emphasis, changes in which research fields may be more

or less interested in a particular data institution's services, changes in collaborative practices and the nature of research teams, changes in the nature of data and the organization of work required to steward data and documentation, new expectations created by collaboration and coordination technologies, the size and makeup of the data institution and how it organizes work, and changes in funding and regulatory environments including, expectations about justification and broader impacts and expectations about data sharing or deposit.

The PF framework shifts our understanding of sustainability away from sustainability as an absolute attribute or state of the data institution (a condition implied perhaps by the earlier frameworks), whether the sustainability attribute is achieved from linear progress (the SI) or from continuous adjustments via feedback loops (the IAD and OR frameworks). It suggests sustainability is relational, or an attribute of a relationship between an element of data institution and something else. To give an example, a particular data service of a DA may be sustainable with respect to its user base (i.e., it is heavily used), but not to its funding streams (i.e., if seed funding has expired and no new sources of revenue have been developed).

3.5 Socio-Technical Concepts and Sustainability

In this final section, we draw on socio-technical frameworks to further develop how we conceptualize data archive sustainability. The principle of symmetry from science and technology studies (STS) suggests equal attention to failure and success of technological systems. Claims that a technology "works" is the result (and not the cause) of the system being successful (Pinch and Bijker, 1984). Further, something can be working, then not be successful and no longer work (Wyatt, 2008). Perceived success is contingent and transient and thus merits further investigation. Applying this line of thought to claims of organizational sustainability, we begin by acknowledging that we should treat claims of sustainability as claims meriting investigation. We should also see sustainability as a fragile state that changes over time. The concept of interpretive flexibility of relevant social groups points out that while one group interprets a technology as successful, another may interpret it as a problem. Culturally, success is achieved when powerful stakeholder groups interpret a technology favorably (Bijker 1997). The STS literature would also caution us to pay attention to whose claims and definitions of "sustainable" are invoked, how and for what purpose. Who are the actors who are declaring sustainability or lack thereof? What are they invoking to support their claims?

4 Discussion: How should we conceptualize sustainability

The sustainability frameworks described above complement and expand our initial understanding of data archive sustainability drawn from our inductive analysis of the information science practice literature. Table 2 below summarizes the most important contribution of each framework to our expanded understanding of data archive sustainability.

Framework Reviewed	Contribution to understanding Data Archive Sustainability
Sustainability Index	Roadmap of skills/competencies that DA should develop as they move from Grade 1 (low sustainability) to Grade 5 (high sustainability)
Institutional Analysis and Development Framework	DA and stakeholders develop, maintain and evaluate rules for provision and use of shared data resources in a continuous feedback loop of evaluation and learning. Decisions about actions taken based on others' contributions to/use of the shared resource developed within specific physical, economic and institutional contexts.
Organizational Resiliency	The degree to which a DA is resilient, or returns to a steady state, depends on the volatility of its environment and on its internal capacities to scan the environment for threats/opportunities and take action based on that information.
Infrastructure Theory: Project Flexibility	Sustainability is not an attribute of data archives, but of the relationship between a DA and something else.
Socio-technical	Analysts should treat claims about sustainability as phenomena to investigate. Interpretive flexibility leads relevant social groups to have different interpretations of a DA's sustainability.

Table 2. Frameworks that Inform Data Archive Sustainability

The information science literature gave us starting points by pointing out common topics of concern insuring financial stability, creating and maintaining robust relationships with a multitude of stakeholders, developing services and products valued by stakeholders, standards compliance, accountability and staff knowledge/capacities, legal issues and disaster planning. The reviewed frameworks expand on and enrich some of these topics. The Sustainability Index and Organizational Resiliency Frameworks emphasize skills especially those related to environmental monitoring and building capacities for change. The Institutional Analysis and Development Framework illustrates how governance relationships, rules, monitoring and enforcement influence stakeholder's decisions about support of and use of pooled resources. The OR framework draws out how the turbulence of external environments might impact sustainability. The Project Flexibility approach describes changes in the scientific communities and their data that influence sustainability.

More importantly, the five frameworks suggest new ways of conceptualizing sustainability that may be helpful to designers and analysts of information services like DA. For example, the stages of the Sustainability Index suggest an ideal state of "High Sustainability" resulting from linear progress in obtaining skills and competencies. Alternatively, the cyclical feedback loops of the IAD and the OR frameworks draw attention to the processes needed to sustain an "ideal" sustainability state. In contrast, the project flexibility approach argues that we shouldn't think of sustainability as a steady attribute of an organization like a DA. Sustainability is only an attribute of the relationship between a particular part of a DA or data service and something else. Finally, concepts from STS suggest we analyze sustainability not as an attribute of an organization or a relationship, but rather as a claim made by particular stakeholder groups at certain points in time, employing particular strategies for varied purposes.

In order to better ensure the longevity of data and promote data sharing, information science needs a more nuanced understanding of what "sustainable" means in relation to information services such as data archives. In pursuit of the goal of a more complex conceptualization of sustainability, we first analyzed how the current information science practice literature frames sustainability. We then compared five frameworks that added new dimensions. The resulting more complex conceptualizations of sustainability can aid data archive stakeholders, designers and analysts in making decisions about how to develop or critique data institutions that important stakeholders perceive as sustainable over the long term and in light of unexpected disruptions.

5 References

- Bhamra, R., Dani, S., & Burnard, K. (2011). Resilience: the concept, a literature review and future directions. *International Journal of Production Research*.
- Bijker, W. (1997). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge MA: MIT Press.
- Burnard, K., & Bhamra, R. (2011). Organisational resilience: development of a conceptual framework for organisational responses. *International Journal of Production Research*, 49(18), 5581–5599. doi:10.1080/00207543.2011.563827
- Chowdhury, G. (2012). Sustainability of digital information services. *Journal of Documentation*, 69(5), 602–622.
- Hess, C.; Ostrom, E. (2011) Introduction: An Overview of the Knowledge Commons. in Hess and Ostrom (Eds.) *Understanding Knowledge as a Commons* (pp. 3–26). Cambridge MA: MIT Press.
- Knowledge Exchange Project (2014). Report on Knowledge Exchange Workshop Sustainable Business Models for Open Access Services, 1–4. <http://knowledge-exchange.info>
- National Academy of Sciences. (2014). Workshop: Strategies for Economic Sustainability of Publicly Funded Data Repositories Asking the Right Questions. Retrieved June 25, 2015, from http://sites.nationalacademies.org/pga/brdi/pga_087151
- Ostrom, E., Hess, C. (2011). A Framework for Analyzing the Knowledge Commons, in Hess and Ostrom (Eds.) *Understanding Knowledge as a Commons* (pp. 41–82). Cambridge MA: MIT Press.
- Pinch, T.J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science* 14(3), 399-441.
- Ribes, D., & Finholt, T. A. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, 10(5), 375–398.
- Ribes, D., & Polk, J. B. (2014). Flexibility Relative to What ? Change to Research Infrastructure. *Journal of the Association for Information Systems*, 15(January 2013), 287–305.

Wyatt, Sally (2008) Technological Determinism is Dead; Long Live Technological Determinism. In: Hackett, Amsterdamska, Lynch and Wajcman (Eds) *the Handbook of Science and Technology Studies*. Cambridge MA: The MIT Press.