Social science data archive business models : A historical analysis of change over time

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

UCD DUBLIN

RESEARCH QUESTION

In what ways have social science data archive (SSDA) business models changed from the 1970s to the 2000s?

MOTIVATION

Concerns for sustainability of data archives/digital libraries

Interest in different business models to improve sustainability

What is a business model in the SSDA context?

What changes should SSDA expect to make over time?

METHODOLOGY

1. Literature review of science infrastructure business models

2. Historical case studies of long lived social science data archives:
ICPSR (1964- 2004)
UKDA (1967 -2004)
LIS Cross National Data Center (1985-2005)

also UC Berkeley, EDINA, Irish Social Science Data Repository and IPUMs at U Minnesota.

SOURCES
Charles Beagrie Ltd. (2017). The Archive Development Canvas. Retrieved from https://cessda.net/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP4/Cost-Benefit-Advocacy-Toolkit/Archive-Development-CanvasDaSilva, C. M., & Trkman, P. (2014). Business model: What it is and what it is not. Long Range Planning, 47(6), 379–389.
Dillo, I., Hodson, S., & Waard, A. de. (2016). Income Streams for Data Repositories. https://doi.org/10.5281/zenodo.46693
The Knowledge Exchange. (2014). Report on Knowledge Exchange workshop Sustainable Business Models for Open Access Services; Crow, R. (2013). Sustainability of Open Access Services Report. Phase 3: The Collective Provision of Open Access Resources.
Morris, M., Schindehutte, M., & Allen, J. (2005). The entrepreneur's business model: toward a unified perspective. Journal of Business Research, 58(6), 726–735.
Osterwalder, A.; Pigneur, Y.; Tucci, C. L. (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of the Association for Information Systems, 15.

What is a business model? Frameworks from the business literature.

"Business model": term emerged in the 1990s as internet increased options how to offer services. Little agreement on what a business model actually consists of. A combination of two ideas: (1) the idea of organizations as a bundle of resources and capabilities that can be arranged in different ways, and (2) the question of how best to arrange those resources and capabilities to conduct transactions.

SSDA Business Model Framework (combination of Morris and Osterwald frameworks)

1: How does the data archive create value?
2: For whom does the data archive create value?
3: What is the data archive's source of competence or internal capacity strengths?
4: How has the archive competitively positioned itself vis a vis alternatives over time?
5: How does the data archive obtain revenue?
6: What have the time, scope, and size ambitions been of the data archive and its governance bodies?
7: How does the archive create and maintain communications loops with its users, funders, governing bodies and other stakeholders?
8: What kind of relationships does the data archive establish between itself and its different customer segments/stakeholders?
9: How are work activities arranged and resources organized to support activities?

Business models: Activity in the Data Professions

CESSDA Toolkit: Archive Development Canvas
Brainstorming tool - Includes many of the above elements "use the Canvas to model how the elements would change if you adopted a different approach."

European Commission Research Infrastructures Project
(2017, Feb) **EU Commission Final Report on Stakeholder Workshop** Recommends increased attention to industry as a possible user group/ financial supporter, and that the EC provide a framework for long term funding commitments.
(2016, Nov) **EU Commission. Stakeholder Workshop Long-Term Sustainability of Research Infrastructures – Exploring RI Full Potential** (Brussels) Emphasizes the importance of partnerships with commercial industry as a user and as a supporter.
(2015) **EU Commission Horizon 2020 Advisory Group on European Research Infrastructure** (Nov) Emphasizes how "flexible business models" are "essential to keep [research infrastructure] sustainable in the long run.

RDA-WDS Cost Recovery Interest Group Project
(2016, Dillo, Hodson, & Waard) explored current and potential future alternative income streams via structured phone Interviews with representatives from 22 mostly science data repositories globally.

OECD Global Science Forum (GSF) and CODATA
(2016, Geneva) **Report of Workshop on Strengthening the Sustainability and Effectiveness of Research Infrastructure** Confusion about what the term business model means, note that it implies profit, where the mission of infrastructure is not to create profit but to create knowledge.
(2015) **Sustainable Business Models for Data Repositories** documents current revenue streams, assess attitudes to revenue diversification, explore what new income streams are being trialed.
SciDataCon (2016) **Session on Sustainable Business Models for Data Repositories** witness statements from 14 repositories.

Knowledge Exchange
The Knowledge Exchange. (2014). Report on Knowledge Exchange workshop Sustainable Business Models for Open Access Services and the Sustainability Index; Crow, R. (2013). Suggests combining fee and free services, but assumes a service collective and open data.

Others projects: APARSEN, US National Academy of Sciences Workshop Session 2 (2014) "Business Models and Economics of Sustainable Data Infrastructures", Sustaining Domain Repositories for Digital Data White Paper (2013-Ann Arbor); Ithaka S+R Projects 2008-2011

Change Over Time: ICPSR, UKDA, LIS

1: How does the data archive create value?

ICPSR: Curation for all deposits, then release of "as is" data, then addition of contract services, education classes, educational materials

LIS: Take data from national statistical agencies, harmonize it so that it can be compared, maintain codebooks, offer access, education, production of white papers. *Considering serving as platform for other groups' data collections.*

UKDA: Curation for UK government agency datasets (sought to expand), broker international exchange, education classes, educational materials, easy ordering, qualitative expertise, started information studies courses to maintain SSRC funding (1970),

2: For whom does the data archive create value?

ICPSR: Expanded from focus on political science to broader set of researchers. Develop contractor service to government agencies, granting agencies. Heavy educational use in 1970s.

LIS: Income researchers and policy analysis internationally, granting agencies, media, organizations interested in inequality

UKDA: UK government agencies (sought to expand to increase support), expanded from focus on political science to broader set of data, consideration of commercial users, qualitative researchers ('99), pressure to provide services to host university.

4: How has archive positioned itself in relation to alternatives?

ICPSR: Alternatives - free data put up by individual researchers via Internet. Other projects such as repositioning in relation to IPUMS free Census Data

LIS: No direct competitors as no one else does the data harmonization, but people do confuse their microdata with other cross national economic data and wonder why do we need microdata?

UKDA: 1970s comparison of UKDA to other data banks to argue UKDA's value to SSRC; archive remained at Essex in part due to free computing cycles (1980s); Rebidding of ESRC contracts led to greater coordination via the UK Data Service model.

8: What kind of relationships does the data archive establish between itself and its different customer segments/stakeholders?

ICPSR: Historical formal structures of Council, committees, Organizational Representatives, attend conferences as vendor, summer training, sometimes difficult host university support, sought relationships with EU and American data banks

LIS: Attend major international income statistical conferences, cultivate and maintain contacts in member nations formally (via boards) and informally, host universities support in Lux and NYC

UKDA: Sought relationships with EU and American data banks, data exchanges, effort to extend length of ESRC contracts to 5 year or more, host university support weaker when not seen as prestigious, bringing in resources or providing services to the host.

Kristin Eschenfelder, University of Wisconsin-Madison | Kalpana Shankar, University College Dublin | and Rachel Williams and Allison Langham of the University of Wisconsin-Madison