Draft economic sustainability reference model

Chris Rusbridge and Brian Lavoie

1. Scope

This reference model addresses the sustainability of digital assets in which there is a long-term public interest. The reference model aims to help key decision-makers build a sustainability strategy, which could be thought of as a case for the continuing flow of resources. The model is particularly targeted at situations more complex than a simple business case, taking non-financial benefits and value into account. The model encourages consideration of economic risks across the whole economic lifecycle of the digital assets.

Many ideas and concepts in this reference model are based on the findings and recommendations in the final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. The reference model does not describe an architecture, nor any specific solution.

Version history

<table>
<thead>
<tr>
<th>Version No.</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>Version for external reviewing</td>
<td>12 October 2011</td>
</tr>
<tr>
<td>0.55</td>
<td>A couple of text changes; Figure 4 fixed; Brian’s responses to Neil’s comments</td>
<td>11 October 2011</td>
</tr>
<tr>
<td>0.54</td>
<td>First pass at responding to NG’s comments</td>
<td>7 October 2011</td>
</tr>
<tr>
<td>0.52</td>
<td>First comments incorporated</td>
<td>5 October 2011</td>
</tr>
<tr>
<td>0.51</td>
<td>First combined</td>
<td>21 September 2011</td>
</tr>
<tr>
<td>0.1 etc</td>
<td>Individual sections</td>
<td>10 May 2011</td>
</tr>
</tbody>
</table>

Copyright: This document is draft and may be changed at any time without notice. This version is copyright Chris Rusbridge and Brian Lavoie, and may not be cited or quoted without permission, other than for the purposes of discussion, and developing and improving the draft. The final version will probably be licensed under CC-BY. This reference model was produced with support from JISC and OCLC-Research.
Draft economic sustainability reference model

Contents

DRAFT ECONOMIC SUSTAINABILITY REFERENCE MODEL............................................. 1

1. SCOPE .......................................................................................................................... 1

2. INTRODUCTION ......................................................................................................... 2

3. SUMMARY OF THE REFERENCE MODEL................................................................. 4

4. AN ECONOMIC LIFECYCLE ...................................................................................... 6

5. SUSTAINABILITY STRATEGY .................................................................................... 8
   5.1. FIVE CONDITIONS FOR ACHIEVING ECONOMIC SUSTAINABILITY ............... 8
   5.2. TAKE-AWAY POINTS: ......................................................................................... 12

6. ECONOMIC RISKS, ASSESSMENT AND REMEDIES ........................................... 12
   6.1. HOW DO RISKS APPLY IN THIS CONTEXT? ..................................................... 13
   6.2. WHAT KINDS OF THINGS ARE AFFECTED BY ECONOMIC RISKS? .................. 13
   6.3. RISK TREATMENTS AND REMEDIES ............................................................... 15

7. KEY ENTITIES ........................................................................................................... 17
   7.1. DIGITAL ASSETS ............................................................................................. 17
   7.2. PROCESS ............................................................................................................ 19
   7.3. STAKEHOLDERS AND THE STAKEHOLDER ECOSYSTEM ................................ 20

8. SUSTAINABILITY CONCEPTS, GLOSSARY, DEFINITIONS (CR) ......................... 24

9. TOOLS (BL) .............................................................................................................. 25

10. EXAMPLES AND CASE STUDIES ........................................................................... 25
   10.1. COMMERCIALLY-OWNED CULTURAL CONTENT ............................................. 25
   10.2. SUSTAINABILITY IN CONTEXT: COLLECTIVELY-PRODUCED WEB CONTENT ...... 29
   10.3. SUSTAINABILITY IN CONTEXT: RESEARCH DATA ........................................ 33
   10.4. SUSTAINABILITY IN CONTEXT: SCHOLARLY COMMUNICATIONS .................... 36

11. REFERENCES ............................................................................................................ 41

2. Introduction

Much of the information that society produces and consumes today is in digital form, with an increasing proportion existing solely as digital information throughout its life. Digital content and technologies confer many advantages, but also introduce new and grave risks that threaten the long-term availability of valuable digital assets. There is a growing understanding that threats to the persistence of digital assets are often economic rather than technical in nature. Economic risks arise from the problems of allocating scarce resources to sustaining digital assets.

The literature on digital preservation tends to focus on actions once digital assets move into an archive. When considering the economic sustainability of digital assets, it is important to look at the whole economic lifecycle of the assets. There are critical stages in the lifecycle of many digital assets where the economic risks can become very significant. Because of the need to think about economic risks to digital assets
throughout their lifecycle, this document uses the term digital curation rather than
digital preservation. Digital curation involves maintaining, preserving and adding
value to digital information throughout its lifecycle\(^1\).

Until recently, the economics of digital curation was a neglected topic in the
literature. This gap was filled by the final report of the Blue Ribbon Task Force on
Sustainable Digital Preservation and Access (BRTF), *Sustainable Economics for a
Digital Planet*. This report explained how achieving economic sustainability is
complicated by the difficulty in coordinating benefits, incentives to preserve, and
roles and responsibilities across the community of stakeholders attached to a
particular set of digital assets. The report provided a wealth of findings and
recommendations to assist decision-makers in thinking through the issues associated
with achieving economic sustainability in regard to their own digital preservation
activities.

In imagining ways to build on the foundations established by the BRTF report, an
obvious path forward is to translate the concepts, findings, and recommendations of
the report into a resource that is of practical value to economic decision-makers. Such
a resource should help them discharge their responsibility of building sustainability
strategies for their digital curation activities. The report, in its original form, is a
lengthy narrative that requires a significant investment of time and effort to fully
digest. It is also pitched more at the strategic level in terms of understanding the key
issues of economically sustainable digital curation. It would be useful for a successor
to distil out the most essential concepts and reformulate them into a practical, “on the
ground” guide for practitioners designing a sustainability strategy for a digital
curation activity.

This paper lays out a *reference model for economically sustainable digital curation*. It
defines the notion of a sustainability strategy; highlights the key concepts planners
must take into account when designing that strategy; and enumerates the kinds of
economic risks that the sustainability strategy should defend against. The reference
model also includes an economic lifecycle model that assists in thinking through
sustainability issues over the complete lifecycle for digital assets. The reference
model aims to be domain-independent, and therefore applicable across a wide range
of digital curation contexts.

A sustainable digital curation activity requires sufficient resources to achieve its long-
term goals over the complete digital lifecycle. This reference model is intended to be
a tool to aid planners in building, clear, persuasive arguments to help unlock those
resources. It is not a cost model, or a taxonomy of business models. It is a way for
planners to organize their thinking about the features of the digital assets,

\(^1\) Adapted from the definition by the UK Digital Curation Centre [http://www.dcc.ac.uk/digital-curation/what-digital-curation](http://www.dcc.ac.uk/digital-curation/what-digital-curation)
the costs will be real, financial costs, much (although not necessarily all) of the societal value will be non-financial. The problem of keeping resources flowing to meet those costs then changes to convincing the decision-makers who control those resources of that societal value. A successful sustainability strategy must address all of these elements if it is to be compelling enough to attract resources.

Because these economic sustainability decisions are made by people, with their own values, background and opinions, there can be no certainty that a sustainability strategy built using this reference model will be successful, or will continue to be successful even if it succeeds at first. Libraries, archives or other services that seek to manage digital assets in the long-term public interest must therefore pay continuing attention to the value their collections can unlock for society, and ensure they collect and can present persuasive evidence of that value.

3. Summary of the reference model

Economically sustainable digital curation is a complex topic even when discussed only in generalities. It becomes even more complex when applied to the particular circumstances of a specific digital curation activity. The reference model helps planners navigate this difficult terrain by breaking it down into four primary components: the economic lifecycle; the sustainability strategy; economic risks and remedies; and key entities.

Figure 1 Summary of reference model

The economic lifecycle can be understood as the background against which a digital curation activity operates. It is intended to convey the notion that economic decision-making does not take place in a static environment, but rather a dynamic one that continually cycles through a progression of events, each of which potentially may impact economic sustainability. The first step in designing an effective sustainable
strategy is to understand the economic lifecycle over which that strategy is expected to operate. The economic lifecycle is described in Section 4, and graphically depicted in Figure 2.

A sustainability strategy is the means by which a digital curation activity orchestrates the economic factors by which necessary to ensure that the activity has adequate resources to achieve its long-term goals. More specifically, a sustainability strategy is a plan for meeting the BRTF’s five necessary conditions for economic sustainability: value, selection, incentives, resources, and organization/governance. These conditions are described in detail in Section 5.1.

Meeting these conditions requires a close understanding and coordination of the key entities relevant to economic decision-making for digital curation: digital assets; the curation process; and stakeholders. Digital assets are what we invest in when we allocate resources to digital curation; the curation process is the mechanism by which our resource allocation is transformed into a “return on investment” (i.e. the ongoing availability of valuable digital assets); and stakeholders are the individuals or organizations that reap the benefits of sustainably curated digital assets, and/or contribute toward the curation process which maintains them. The concept of a sustainability strategy is described in Section 5, and graphically depicted in Figure 3.

Economic risks and remedies are, in the case of the former, potential obstacles to achieving sustainability, and in the case of the latter, potential solutions for mitigating or overcoming these obstacles. In the most general interpretation, economic risks impact digital assets and the curation process in a variety of ways; a general list of the ways economic risk impacts digital curation is provided in Section 6.1. More specifically, economic risks manifest themselves as potential challenges to achieving one or more of the five conditions required for economic sustainability. The reference model provides a simple framework of possible “reactions” to economic risk that a sustainability strategy can employ; the framework is described and illustrated with examples in Section 6.3.

Finally, identifying economic risks (and appropriate remedies for these risks) in the context of a given digital curation activity requires a deep understanding of the key entities that constitute the economic environment for that activity. As mentioned above, the key entities for digital curation are digital assets, the curation process, and stakeholders. Understanding the properties of these entities, and how they manifest and relate to one another in a particular digital curation context, is a necessary step for identifying the significant economic risks relevant to that context. Key properties of digital assets are described in Section 7.1; key properties of the curation process are described in Section 7.2. Understanding stakeholders presents a special challenge for sustainability planners; the key concept idea here is the stakeholder ecosystem, described in detail in Section 7.3. The stakeholder ecosystem consists of three layers of abstraction that, taken together, constitute a useful sketch of the way stakeholders are organized within a particular digital curation context, and how that organization might harbor potential economic risks impacting the prospects for achieving sustainability. The first layer is the types of stakeholder roles associated with most digital curation activities; these roles are graphically depicted in Figure 4. The second layer describes the relationships existing between stakeholder roles in a given digital curation context; an example of these stakeholder relationships is provided in Figure 5. The third layer completes the
stakeholder ecosystem by describing the distribution of stakeholder roles across distinct individuals or organizations. Understanding this distribution of roles is critical to identifying potential economic risks inherent in a particular configuration of the stakeholder ecosystem. An illustration of how analysis of the stakeholder ecosystem can expose potential economic risks (and suggest potential remedies) is provided at the end of Section 7.3.

* * *

In summary, the reference model for economically sustainable digital curation maps out the key elements of the problem space planners face when designing a sustainability strategy for their digital curation activities. The reference model simplifies the process of designing an effective sustainability strategy by breaking the process down into the key components that should be considered, and drawing planners’ attention to the properties of those components most relevant for economic sustainability. While the reference model does not “solve” any digital curation activity’s sustainability issues, it provides a framework within which the nature of these issues can be understood, and appropriate solutions devised.

4. An economic lifecycle

Being clear about the nature of the lifecycle of the information that is being curated or preserved is important when assessing economic sustainability. One of the great ideas from the Blue Ribbon Task Force was “the case for preservation is the case for use”. Use brings value, and value justifies preservation. It works the other way, too; if you separate use too far from the preserved content, then the value is reduced and the argument for preservation is diminished. This is not to say that there might not be other motivations for curation investment: e.g., a “bequest to future generations”, or even a sense of “sentimental value” derived from knowledge that a digital asset continues to exist (see for instance Coyne 2007ii). But demonstrable value from use is likely to make the strongest case for ongoing investment in digital curation.

Quite often in digital preservation circles the model is one of “post-use”. The functional model of OAISiii takes a resource, ingests it safely into the digital preservation box, where it is looked after until someone asks for it, when it is disseminated out to the consumer. Putting emphasis purely on the archival stage of digital assets (effectively after the economic life of the resource) makes sustainability hard to achieve. In contrast, the economic sustainability of digital assets can depend on consideration of the whole lifecycle. This document therefore uses the term digital curation rather than digital preservation. Digital curation involves maintaining, preserving and adding value to digital information throughout its lifecycle2.

An economic lifecycle model is included as a key component of the reference model. Several lifecycle models were assessed, including the Digital Curation Centre’s lifecycleiv (described in Higgins 2008v). Generally these models were too fine-grained for the purposes of this reference model, and none expressed the important economic realities. Therefore a draft economic lifecycle model has been prepared (see Figure 2 below).

---

2 As noted, adapted from the definition by the UK Digital Curation Centre
http://www.dcc.ac.uk/digital-curation/what-digital-curation
The diagram is an economic view of the digital asset lifecycle. This lifecycle is intended for an individual data asset, for a data or information service (i.e. a service that makes data assets available) or for a data archive.

Digital assets are created and some are selected and acquired for the service or archive under consideration. The selected digital assets have to be prepared for use; this is the “ingest” phase in OAIS, the editing phase in publishing, etc. It includes adding relevant metadata required for use. This stage has been identified as one of the highest cost stages in archiving (Beagrie et al 2008). Once usable, the digital objects have to be made available; running costs might continue indefinitely.

Available digital assets can be used (disseminated in OAIS terms). The assets create value through use (or availability for potential use). The economic case must be that the aggregated value over many resources and significant time exceeds the aggregated costs (even if the two are expressed differently).

As long as there is a reasonable perception of value, this situation can continue indefinitely. But eventually, some sort of problem (risk event) will arise. This could be a technical problem to do with the digital assets (e.g. obsolescence); it could be a technical problem to do with the service (needs some kind of significant upgrade); it could be a business problem to do with the service (bankruptcy, change of ownership or focus, etc). And it could also well be that the projected value of the curated digital assets failed to materialize, thereby calling into question the wisdom of continuing to allocate resources toward their curation. There is a question as to whether this resource has reached the end of its economic life. Decisions then must be made, as
further significant investment may be required. Some (or all) digital assets will be retained, perhaps transformed to make them usable in the new environment. Some (or all) digital assets will be disposed of, de-accessioned, etc. The best outcome for these latter assets is that they are handed off to another service or archive; without that critical step, they will very soon be lost. Similarly, assets might have arrived at the current service having been handed off from an earlier service.

5. Sustainability Strategy

Digital curation activities adopt a technical strategy to orchestrate the technical processes and workflows needed to maintain the long-term accessibility of digital materials: e.g., redundant storage, regular media refreshment, format migration scheduling, and so on. In the same way, a digital curation activity must develop a sustainability strategy to orchestrate the economic factors necessary to ensure that the activity has adequate resources to meet its long-term curation goals. A sustainability strategy is a plan that organizes the key entities associated with long-term sustainability – i.e., digital assets, the curation process, and stakeholders – in such a way that the curation activity becomes a sustainable economic activity. The sustainability strategy is enacted over the economic lifecycle described in the previous section; that is, the sustainability strategy is intended to guide the curation activity through the significant events that occur during the economic lifecycle.

A sustainable economic activity is one that addresses each of the five sustainability elements set out in the BRTF definition of economic sustainability:

5.1. FIVE CONDITIONS FOR ACHIEVING ECONOMIC SUSTAINABILITY

- **Value:** This condition simply requires some stakeholder or set of stakeholders to have a clear interest in (or to see value from) the long-term curation of a particular set of digital assets. A digital curation activity is unlikely to attract funding – and thereby sustain itself – if no one sees value in the digital materials it is curating! But more than this, those that do have an interest in long-term curation need to be able to express that interest or value in compelling, specific ways. What kinds of valuable activities would be possible if the digital materials in question remained accessible over time? Or conversely, what kinds of valuable activities would not be possible if the materials disappeared? What might certain stakeholders be willing to pay to ensure the digital assets do not disappear?

Since a sustainability case is particularly hard to build where the costs are real money but the returns are non-financial, sustainability can be greatly affected by perceptions of value. Attempts have been made to express value of curatorial efforts in financial terms (for example the British Library carried out such an exercise in 2004\(^{vii}\)), but they are not necessarily convincing either to a hard-nosed accountant or to those who see value in more societal terms.

- **Selection:** One of the fundamental principles of economics is that resources are scarce; we usually cannot achieve everything that we might like with the resources available to us. For digital curation, this means that the slogan “curate everything for all time” is a non-starter. The resources available for curation will always be limited, and therefore we must prioritize and make
choices: to the extent we can, we must proactively select digital assets for curation that are likely to promise the greatest value through use over time. And we should also keep in mind that the process of selection works in two directions. In particular, we must be prepared to “de-select” curated digital assets when the value from continuing to curate them no longer justifies the cost.

- **Incentives:** It is one thing for a stakeholder to articulate an interest in having ongoing access to a set of digital materials (see “Value” above). It is quite another for a stakeholder to step forward and accept responsibility for curation. Sustainable digital curation requires stakeholders who not only recognize the value of curation, but who are also willing to sponsor or carry out the curation process. In short, there needs to be robust motivation or incentives to curate. Cultivating these incentives often means identifying and leveraging an institutional self-interest in curation: for example, curation as a business opportunity; curation as part of an institutional mission; curation as a means of fulfilling a mandate, and so on. Often, curation involves stakeholders with differing incentives to curate; these incentives need to be orchestrated over the full digital life cycle. For example, a media company may perceive a revenue incentive to curate a digital movie over a limited period while the asset has economic value. When this period expires, mechanisms need to be in place to transfer the asset to another institution with a different curation incentive, such as a library or archive.

- **Resources:** Naturally, no discussion of economic sustainability is complete without talking about resources. Curation activities, like any other activity, require sufficient resources to achieve long-term goals. Meeting this condition often boils down to developing mechanisms to transfer funding and other resources from those who benefit from and are willing to pay for digital curation, to those who are willing to provide curation services. There is a variety of market and non-market mechanisms for doing so, such as pricing models, compulsory fees or taxes, volunteer efforts, and philanthropic donations. Whatever mechanism is chosen, it must support an ongoing flow of resources such that long-term curation goals can be achieved. But it is not enough to simply make resources available for curation. These resources should be used as efficiently as possible. Efficiency in this sense does not mean cutting corners, but rather getting the most value out of the resources allocated to curation. For example, we should strive to leverage economies of scale by spreading costs over higher volumes of curation activity. We can also attempt to leverage economies of scope by spreading costs over different yet related services: e.g., locating curation and end-user access services on the same repository platform.

- **Organization/governance:** Economic sustainability also requires that planners choose an appropriate organizational form for digital curation activities. A variety of organizational forms are possible: for example:
  - An organization with no private interest in curation curates on behalf of others (e.g. a third-party curation service provider)
  - An organization with a private interest in curation curates on behalf of itself and others (e.g. a research library)
An organization with a public mandate to curate on behalf of society (e.g. a national archive).

To the degree there is discretion to choose, an organizational form for curation should be appropriate given the conditions prevailing in a particular context. In addition to an appropriate organizational form, a good governance mechanism is needed to ensure that curation goals are clearly articulated, a strategy is formulated for achieving these goals, curation responsibilities are appropriately allocated, and metrics and benchmarks are in place to evaluate outcomes.

A sustainability strategy is not unlike the perhaps more familiar concept of a business plan. According to Wikipedia, a business plan is a “formal statement of a set of business goals, the reasons why they are believed attainable, and the plan for reaching those goals”\textsuperscript{viii}. In the same way, a sustainability strategy articulates a realistic set of curation goals (that is, goals which address the Value condition, tempered by the limits imposed by the Selection condition) and the means for achieving them (including Incentives, Resources, and Organization/Governance). Like a business plan, a sustainability strategy functions like a roadmap in helping decision-makers move curated digital assets smoothly across the economic lifecycle. A sustainability strategy may be more generalized, potentially expressing both value and costs in different (non-monetary) terms. However, for the particular case where a sustainability strategy is built on expected revenue, the sustainability strategy could collapse into a business plan.

Achieving sustainability requires planners to take into account the properties of digital assets, the curation process, and stakeholders, and align them in such a way that each of the sustainability conditions is addressed. The result is a sustainability strategy (see Figure 3). It should be noted that a successful sustainability strategy is not one that guarantees long-term economic sustainability. Regrettably, such a strategy does not exist. Instead, a successful sustainability strategy is one that maximizes the prospects of achieving sustainability by cultivating a thorough understanding of the conditions in the economic environment relevant to the five sustainability conditions.
There can be no “off-the-shelf” sustainable strategy as no two digital curation contexts are exactly alike. One reason is that stakeholders (and by extension, curation decision-makers) are humans, who will make judgments based on perceptions of value and cost. The properties of the three entities often represent a conglomeration of conditions that can operate either to promote or to discourage sustainability. And since the nature of Digital Assets, the Curation Process, and Stakeholders will differ from context to context, planners will need to think through the conditions relevant to their particular context and design a sustainability strategy that is the best fit (or at least a reasonable fit) for those circumstances.

So in reality the design of a sustainability strategy will first need to account for the properties of the key entities (digital assets, the curation process, and stakeholders) as they manifest themselves in the context in question (see Key Entities/Properties section). Next, the implications of these properties for meeting the five sustainability conditions must be considered, and in particular the key risks that arise in meeting those conditions must be identified and remedies sought to overcome or at least mitigate them (see Economic Risks section). The result is a sustainability strategy that fits the circumstances of the particular digital curation activity in which it is expected to operate.

It is important to emphasize again that any sustainability strategy is necessarily an approximation, in the sense that it cannot guarantee long-term sustainability for the digital curation activity. Not all economic risks can be identified or foreseen; for those that are identified, not all of them can be resolved or mitigated. Moreover, solving some problems may only be possible through trade-offs. For example, restricting access to paying users in order to strengthen incentives to contribute to a digital curation activity may also have the effect of diminishing the societal value proposition for curation, since the benefits from curation would be distributed over a smaller user base.
Planning a sustainability strategy is an ongoing activity. No sustainability strategy is likely to remain effective without alteration indefinitely. Planners must be alert to changing conditions in the economic environment, and be prepared to re-evaluate, adjust, or even completely re-design the sustainability strategy as needed. More generally, a sustainability strategy needs to be cognizant of and address the key economic risks present in the context of a particular digital curation activity.

5.2. **Take-away points:**

- A sustainability strategy orchestrates the key entities in order to ensure a curation activity has sufficient resources to meet its long-term goals.
- Achieving sustainability means meeting the BRTF’s five sustainability conditions.
- To design a successful sustainability strategy, planners must understand the properties of the key entities; identify significant economic risks associated with the properties; and identify appropriate remedies to address the risks.
- No sustainability strategy is perfect; it can only maximize the prospects of achieving sustainability, not guarantee it.
- A sustainability strategy must evolve as conditions evolve.

6. **Economic risks, assessment and remedies**

A key function of any effective sustainability strategy is to aid a curation activity in navigating the myriad economic risks that exist throughout the economic lifecycle. Kevin Knight, Chairman of the ISO Working Group on Risk Management Standards has pointed out “You do not have to manage risk!! Survival is not compulsory”.

The first formal international standard on risk management is the new ISO 31000 series, published in 2009 and based around an earlier Australian standard. However, these are expensive standards. The UK Institute of Risk Management had published an earlier informal standard. In theory this is replaced by ISO 31000, but their website says

> “IRM has decided to retain its support for the original risk management standard because it outlines a practical and systematic approach to the management of risk for business managers (rather than just risk professionals). It also free to download and also available in 15 languages.”

If access to the formal standard is not available, this is a reasonable place to start from.

The common language we use to discuss risk can be ambiguous and confusing; in particular, the word “risk” is used in two different senses. It can mean simply the probability that something might happen (e.g. “there is a high risk of fire”). But in common speech “risk” is also extended to include the impact (e.g. “jaywalking is a big risk”, “nuclear power is too risky”, etc). To be clear, we aim to use the term “risk” as a combination of likelihood (or probability) and impact.

The new ISO standard defines risk as “the effect of uncertainty on objectives”, whereas the IRM informal standard uses the older definition “risk can be defined as the combination of the probability of an event and its consequences”. It is wrong to
think of risks as always being bad: consequences can be positive or negative. Economists refer to positive (yet uncertain) outcomes as upside risk, and negative (yet uncertain) outcomes as downside risk. We take an upside risk when we invest in something (whether with money or personal time or effort), which is exactly what someone has to do for sustainable archives or information services! We must take risks; if we are too risk-averse we will stay in bed all day and die of bedsores.

6.1. How do risks apply in this context?

In the context of economically sustainable digital curation, economic risks are uncertainties that could affect sustainability. Risks will vary from case to case. It is therefore vital to assess and address risks for the specific digital assets and environment when devising a sustainability strategy.

Risks could affect many different aspects. Assuming some sort of service for the management or curation of digital assets over time, risks could affect

- The creation of the service
- The continuation of the service
- The termination of the service
- The succession or transformation from one incarnation of the service to another, whether in succession or in parallel, e.g. through the handoff process (including legal and other agreements), through migration of the assets, identifiers etc, through revised technology for the underlying service, or new owners or policies
- The quality of the service
- All or some digital assets in the service.

This list is inclusive rather than exhaustive.

The economic lifecycle implies risks at various points, but particularly at the stage labeled “risk event”. Effectively, economic risks will most likely come into play when significant changes are expected, for example to the cost structure (where cost may include effort, time etc as well as money), the group of interested stakeholders, or to the underlying resources. As such they are more likely to apply to the whole service rather than part of it.

6.2. What kinds of things are affected by economic risks?

Any of the 5 sustainability conditions described in section 5.1 above could be affected by economic risks.

Value issues:

- uncertainty of value, or "soft" value;
- insufficient value, or poor return on investment (however calculated);
- value perceptions, including returns on investment, tangible versus intangible value, differing "currencies" of value (versus cost), and the problem of public good versus private good are all susceptible to re-interpretation; or
- difficulties in ensuring usable incentives.
Selection issues:
- where selection does not match demand;
- unmeasured demand, or
- uncertain future demand (or on the up-side, excessive demand causing server problems);
- perceptions of long-term value are not enough on their own, and selection processes are never final; there is often a hidden “value for money” proposition that can come in to play if resources become too tight, or the resource demands of part of a collection (perhaps for technical preservation reasons) become too great.

Rights issues:
- many, including orphan (unknown) rights;
- legal dispute on rights;
- rights needed but not available (e.g. request for rights denied)
- where rights are owned by different parties than those interested in preservation, the incentives to act in the public interest may be out-weighed by the downside risks of market erosion.

Organizational issues, including political/policy change.
- decision-makers change, and policies can change over time, meaning that benefits once recognized may become deprecated (e.g. de-funding of AHDSxii)
- Changes in funding priorities, such as decisions by the National Science Foundation to spend a bit less on basic research and a bit more on ensuring data from earlier research becomes and remains re-usablexiii. (An upside risk.)
- where organizations would need to spend resources on curating digital assets, but receive no direct return for use of those assets by others, they may need incentives to act
- Unwillingness to provide an exit strategy, handoff mechanism or even adequate resource extraction capability (applicable at the failure of some social networking sites).
- Publisher takeover or merger with subsequent review of title coverage leading to journal closure.

Resource issues, e.g.
- getting sufficient initial investment, or sufficient continuation resources, or
- getting resource timing right, AKA "cash flow"
- Inability to gain sufficient audience share to attract advertising revenue to sustain the service (a factor in the closure of some social networking sites)
- Unwillingness to provide sufficient initial staffing support to build a service strong enough to get a critical mass of content (a factor reported in the low success of some Institutional Repositories)
- all resource allocation mechanisms are subject to risk over time, including not only funding and staffing, but also volunteer effort and so on.
Being able to make a strong value proposition that is persuasive to relevant decision makers will be critical in addressing these risks.

### 6.3. Risk treatments and remedies

It is sensible to do a risk assessment when deciding how to respond to risks. See the IRM informal standard referred to above for approaches to carrying out risk assessments.

There are at least 4 things you can do about risk. They include:

- **Risk avoidance**: stop the activity with the risk (e.g. only preserve public domain material)
- **Risk reduction/mitigation**: lower probability (e.g. improve procedures) or lower impact (e.g. get multiple resource streams)
- **Risk transfer**: pass risk to another party (e.g. insurance)
- **Risk acceptance**: recognize the risk but decide to live with it.

(Note, terminology here varies widely, but the intent is similar in most examples.)

#### Risk avoidance

Some organizations are extremely risk-averse, and are unwilling to expose themselves to any significant risks. As this section has mentioned, risk is unavoidable in organizations as in life; however, assessing the possible impacts and the likelihood of them occurring (a risk assessment process) is sensible.

An example of a class of works that carries in-built risk would be orphan works. These are works whose rights-owners cannot be traced even after reasonable effort. Legally, making these available or even carrying out some preservation actions could be viewed as copyright violations, so this does become an economic risk. Risk avoidance in this situation would be implementing a strict policy of curating only resources whose rights-ownership status clearly allows this. (The reality of this risk is shown by legal action taken in 2011 against HathiTrust and 5 universities by various authors’ organizations.)

The example above also suggests that curating only public domain works, or even deciding that making the resource a dark archive (not making available to the public at all) could reduce risk even further.

As should be clear, risk avoidance tends to result in reduction in capability and/or value of the resource, which must be balanced against the reduction in risk.

#### Risk reduction or mitigation

Once the more serious economic risks are identified, and a decision has been taken that the risks cannot be avoided, then if possible steps should be taken to reduce or mitigate those risks.

There are two approaches here, preferably to be used together but if necessary in isolation. These are to reduce the likelihood of the risk event occurring, and to reduce the effect or impact if it does occur.
• For example, if the risk were loss of trust by depositors or users, then improving procedures or undergoing some quality assurance process (such as the Data Seal of Approval\textsuperscript{xv}, for example) would reduce the likelihood of this happening.

• Value perception risks might be mitigated from two directions: by ensuring that the cost base is as low as possible (decreasing the divisor in value for money discussions), and ensuring that outputs and outcomes have value and are seen to have value from the points of view of both the user community and decision-makers.

• Where there is doubt about future value, then it might be prudent to carry out an option strategy: make an interim selection but spend little resource in doing so, expecting to spend more resource on making available if demand materializes. For example, movie studios now retain more out-takes and material previously discarded, which is packaged with DVDs to provide further sales opportunities.

• Demand risks could be tackled not only by ensuring the value proposition is clear and known, but also by engaging with potential users. It is also important to keep evidence of demand in ways that can be queried later.

• Rights risks can be tackled by a combination of appropriate advice, soundly-based policies, licenses and disclaimers, and prompt action when appropriate. They might also be addressed by incentives to the rights-owners.

• Political/policy risks are difficult to anticipate, and too often are driven by dogma rather than evidence. They might be addressed indirectly through the value and demand risks mentioned above. Note, these three risks suggest that a sustainable service needs to get its value message across, and this implies something akin to a public relations role as a key element.

• Developing multiple funding streams might mitigate resource risks; this helps reduce the impact if one stream were to dry up. It is possible that multiple funding streams might also increase funder confidence and reduce the chance of ending support (that was certainly suggested as one reason for the survival of the Archaeology Data Service when AHDS lost its funding, given the ADS’ other funding streams, e.g. from developers).

**Risk transfer**

The common approach to risk transfer is to take out an insurance policy (transferring the risk to the insurance company in exchange for a premium; the insurance company aggregates many risks and aims to be able to cover that proportion of them that do occur). There are insurance policies that cover economic risks, but we do not know of any that might be specifically applicable to archives and data services, other than legal protection.

An important form of risk transfer is to simply transfer an activity to another organization. So the arrangements that publishers have with Portico\textsuperscript{xvi} or CLOCKSS\textsuperscript{xvii} are essentially risk transfers, because they are transferring the responsibility (and therefore the attendant risks) of digital curation to these organizations. This is an example where, although risks were misaligned between the
rights owners and the preservation services, rights owners have an incentive to enable
trustworthy preservation due to demand from their large academic library customers.

**Risk acceptance**

Let’s emphasize again that the last of these, risk acceptance, is essential if an archive
or data service is even start. A sensible approach would certainly put in place some
risk transfer and/or mitigation. Among other things, best practice should be adopted.
And part of the point of this reference model on economic sustainability is that
attention should be paid to expressing the value and impact of the service, and making
sure this message gets through to the key decision-makers.

Another example of risk acceptance could be to decide that the service will archive
(and perhaps make available) orphan works (described above). It would be sensible to
include some risk reduction or risk transfer such as developing (and implementing) a
takedown policy, and perhaps reserving some resources or taking out insurance. To an
extent then, risk acceptance is orthogonal to risk treatment and risk transfer; where we
accept a risk, we generally try to do something to reduce, mitigate or transfer it, and
only rarely accept it without further qualification.

[Report recommendations table vs risk TBA…]

Identifying key risks in the economic environment, and designing appropriate
remedies to address them, is the essential purpose of an effective sustainability
strategy. But identifying these risks requires a thorough understanding of the
environment in which digital curation activities operate. In particular, planners need
to understand the properties of the key entities present in virtually all digital curation
activities – Digital Assets, the Curation Process, and Stakeholders – and use these
properties as starting points for identifying the most significant risks associated with
their curation activities.

**7. Key entities**

The economic environment surrounding any digital curation activity is complex,
involving a host of factors interacting to shape the circumstances in which a particular
activity must operate. Understanding this environment, and its implications for
achieving sustainability, can be challenging. Fortunately, it is possible to distil the
economic environment into a few key elements that are present in all digital curation
contexts. A thorough understanding of these elements and their properties is a first
and necessary step toward building a successful sustainability strategy.

**7.1. Digital Assets**

Digital assets are the *raison d’être* of a digital curation activity. They can take a wide
range of forms: research data sets, e-prints, executable software, web sites, and so on.
To qualify as a digital asset, two criteria must be met: *first (and obviously), the object
in question must be digital, and second, it must be judged to have a value that will
persist over some period of time*. The second criterion is crucial, and must be
considered carefully by digital curators in the context of any curation activity: is the
digital object in question truly a digital asset? The question is not trivial: *not all
digital objects are digital assets!*
Each class of digital asset exhibits a variety of properties that to a greater or lesser extent has an appreciable impact on the nature of the curation activity itself. This can be readily seen from the perspective of the technical aspects of curation: the techniques and workflows needed to curate a collection of research data sets are likely to be very different from those needed to curate a collection of executable software. In the same way, the properties of digital assets impact the strategies needed to support the curation activity from an economic standpoint. As in the technical sphere, different kinds of digital assets will have special properties that impact sustainability in unique ways. However, digital assets also share certain core properties that must be taken into consideration when organizing a curation activity.

- **Digital assets are durable yet depreciable:** If kept in proper condition, digital assets can continue to release value for scholarship, private enterprise, education, and entertainment over extended periods of time. However, the clause “if kept in proper condition” is crucial. In order to maintain digital assets’ value over time, those responsible for their stewardship will need to expend resources on an ongoing basis to support their curation.

Digital assets are depreciable in the sense that if not properly maintained, they will tend to succumb to technological obsolescence or bit rot, and in this way, their ability to release value to users will be reduced and eventually eliminated.

The implication of this property for economic sustainability is clear: in assessing the costs of a proposed curation activity, planners must look beyond the current costs of gathering digital assets into a collection and consider something akin to “total cost of ownership”. What will be the costs of maintaining the collection in a usable condition for an extended period of time? And what mechanisms can be put in place to support a sustained flow of resources to the curation activity? *From the outset, curation planners must discard notions of one-time chunks of funding, and think instead in terms of ongoing flows of funding.* [See BRTF Report, p. 25-26; also see Chapter 4 of the reference model regarding the economic lifecycle]

- **Digital assets can be curated by one, but used by many:** A digital asset residing on a public server is, at least in principle, accessible to any Web-enabled user. In this sense, the user base for a curated digital asset is potentially far larger than that of a “physical” item – e.g., a print book. Moreover, a digital asset can be used simultaneously by many users, also in contrast to a physical item, where use by one generally precludes use by another (again, consider a print book). Expressed in economic terms, the use of digital materials is *non-rival in consumption.*

The important implication of this property for economic sustainability is that although the potential user community – in other words, those that benefit from curation – is potentially quite large, the incentive to contribute resources to the curation activity by any one user is generally weak. Once a digital asset is curated by one organization, it is at least in theory available for use by all; there is no need for each organization to curate a local copy of the digital asset, as long as a single curated copy is available on the Web. In contrast, there is a far stronger need to curate local copies of print materials, since a particular print copy can only serve a limited pool of primarily local users. Since one organization can curate a digital asset on behalf of all, it tends to be
in the interest of any one organization to have someone else incur the trouble and expense of curation, while still enjoying free access to the digital assets. This is what economists call the free-rider problem, which can make it a challenge to collect sufficient resources among beneficiaries to sustain the curation activity. To overcome this problem, it may be necessary to exclude some beneficiaries from access to the digital assets if they do not contribute toward curation. If the planners for a curation activity are not prepared to exclude non-contributors, or for some reason it is impossible to do so, this may create significant problems for long-term sustainability; the incentives for contribution to the activity will likely be weak. [See BRTF Report, p. 26-28]

7.2. Process

Long-term accessibility to digital assets is achieved through the curation process, which is the set of activities involved in maintaining digital assets in a usable form for an extended period of time. It is important to distinguish between the process of digital curation, and the outputs of digital curation. In particular, the value of digital curation is experienced through the value-creating capacity of the curated digital assets themselves. Put another way, we generally do not value curation for the sake of curation; we value it because of the uses to which we can put curated digital assets. This establishes an important property of the curation process that is crucial when considering economic sustainability.

- The value of the curation process (and the associated digital information service in which it is embedded) is derived from the value-from-use of the curated digital assets. A curation activity that cannot make a compelling case for the value of the curated digital assets that it manages will find it difficult to attract the funding and other resources needed to sustain itself over time. This ties in with the distinction noted above between digital objects and digital assets. A digital asset is a digital object with a perceived future value. The value of the curation process, therefore, derives from its ability to deliver the value of the digital asset. In economic terms the demand for the curation process is a derived demand. In other words, the demand for a curation service is derived from the demand for curated digital assets. [See BRTF Report, p. 24-25]

Naturally, many of the activities associated with the curation process are technical in nature, consisting of curation techniques and workflows invoked on the digital assets to ensure they persist in a suitable condition for use. For the purposes of economic analysis, however, it is not the technical aspects of the curation process that are key; instead, it is the decision-making process that overarches the day-to-day management of curated digital assets. The decision-making processes associated with digital curation also exhibit several properties relevant to consider in regard to economic sustainability.

- Curation decision-making is path-dependent: Digital materials pass through a sequence of stages, also called a lifecycle, with endpoints of creation and disposal. Decisions made at one stage of the lifecycle often shape the choices available to decision-makers at later stages. This point is illustrated in its starkest terms by the initial choice of whether or not to curate; if the decision is “no”, it is often impossible to revisit and change this decision at a future
date, because by then the digital asset may be unavailable or have deteriorated to such an extent that effective curation becomes prohibitively expensive or even infeasible. [See BRTF Report, p. 28-30]

7.3. Stakeholders and the Stakeholder Ecosystem

The network of stakeholders surrounding a particular curation activity can be complex and difficult to characterize. Moreover, these stakeholders can represent an equally wide range of interests in regard to the long-term future of the assets in question. The organization of these stakeholders, and in particular, the distribution of curation roles across them, heavily impacts the prospects for achieving sustainability, and by extension, the shape of the sustainability strategy best suited for those circumstances. In order to understand the basic contours of the stakeholder interests associated with a set of digital assets, it is useful to employ an organizing device called a stakeholder ecosystem. A stakeholder eco-system designates the key stakeholder roles in lifecycle digital management, and articulates important relationships between them as they relate to sustainability. The stakeholder ecosystem takes the form of a high-level abstraction that helps organize stakeholders into broad categories of interest. Figure 4 provides a view of the types of stakeholders associated with a digital curation activity. It is important to recognize that the stakeholders illustrated in the Figure are roles, and not necessarily distinct entities. A single organization (or individual) can fulfill multiple roles simultaneously. As we will see, this can have important implications in terms of creating conditions that either encourage or discourage long-term sustainability.

Figure 4: Types of Stakeholders

The stakeholder taxonomy depicted in Figure 4 is divided into three areas:
• **Supply-side**: stakeholder roles pertaining to those who create and/or own the digital asset.

  - Creators: entities responsible for creating digital assets (e.g., scientist who creates a data set; artist who creates a digital work of art; company that develops a new computer game)
  
  - Rights Holders: entities who currently own key intellectual property rights related to digital assets, such as the right to provide access, the right to preserve, etc. This will often be the same entity that created the digital asset (i.e. the same stakeholder has both the Creator and Rights Holder roles), but not always. For example, a publishing company may hold the rights to an e-journal article authored by someone else; a social media site may own the collective contributions of its members, etc.

• **Demand-side**: stakeholder roles pertaining to those who benefit from availability of the digital asset.

  - Current beneficiaries: those who currently have well-defined uses for the digital asset, and derive value from its ongoing availability.
  
  - Future beneficiaries: those who could be expected to benefit from the digital asset in the future, or those whose uses of the digital asset are as yet unknown/undefined. The BRTF report noted that the interests of future beneficiaries are often left out of decision-making on digital assets.

• **Lifecycle management**: stakeholder roles pertaining to those responsible for curating the digital asset, and for ensuring its ongoing inclusion in the cultural/scholarly record.

  - Managing Agencies: entities responsible for managing, curating, preserving or providing access to digital assets. Essentially, this is the entity that actually has custody of the material and ensures that it continues to exist and is accessible for use.
  
  - Representatives of the Public Interest: entities who advocate for the digital asset’s inclusion as part of society’s ongoing cultural or scholarly record. This reference model focuses on digital assets in which there is a long-term public interest. Generally speaking, these are materials that potentially would be included as part of society’s cultural, scholarly and scientific record (acknowledging that these terms are only vaguely defined). Typically cultural heritage institutions (libraries, archives, museums) play a prominent role in advocating or actively intervening to ensure certain materials are in fact curated on behalf of the long-term public interest. Various philanthropic organizations, funding agencies and informal volunteer networks would fall into this category as well.
  
  - Resource providers: this term is used broadly for those who provide resources (i.e., funds, in-kind transfers, etc) to support the curation process. Provision of resources can occur in a variety of ways: for example, through payment of a price-for-services (e.g. where the Managing Agency is a third-party digital archiving service and the Resource Provider is a customer of this service); donation (e.g. where the Resource Provider is a corporate sponsor of a digital collection or of a curation organization); grant award; advertising (e.g. where the Resource Provider pays the
Managing Agency for the placement of advertisements in association with the latter’s access services); and so on.

As Figure 4 illustrates, many different categories of stakeholders are associated with a digital curation activity. But this in and of itself does not create an ecosystem. The stakeholder ecosystem is created by the pattern of relationships between the stakeholder roles that holds for a particular activity. This pattern of relationships can be quite complex, and is illustrated by Figure 5.

Figure 5: Stakeholder relationships

Relationships in the ecosystem define how the various stakeholder categories interact with one another. For example, in many digital curation contexts, Representatives of the Public Interest such as libraries and archives act on behalf of Future Beneficiaries by making sure that the latter’s interests are represented in today’s curation decision-making. Similarly, the Rights Holder and Managing Agency often have a relationship in that the former must grant an exclusive or non-exclusive right to curate to the latter. Investors also have a relationship with the Managing Agency, in that they supply resources sufficient for the Managing Agency to carry out its curation activities.

The pattern of relationships depicted in Figure 5 is not necessarily complete – many other relationships between the various stakeholders could be imagined. However, it does suggest the potential complexity that might prevail in the pattern of relationships associated with a particular digital curation activity. It also points up the importance of orchestrating the interaction of all stakeholders in a sustainable digital curation activity. A sustainability strategy that neglects to account for the interests and participation of any of these stakeholder categories is likely to founder at some point in the economic lifecycle.

It is important to emphasize again that all of the stakeholders represented in the ecosystem (i.e., in Figure 4 and Figure 5) are roles, and not necessarily distinct
entities. So a single person or organization could simultaneously fill multiple roles. Many of the important issues regarding incentives to preserve, allocation of long-term curation responsibilities, etc. hinge on whether various roles are combined within a single organization, or separated across distinct entities. The more that stakeholder roles are dispersed across distinct entities, the more negotiation and consensus must be achieved to reach sustainability goals, and the greater the probability that individual interests will conflict. This leads to an important property of stakeholders that has enormous implications for the prospects of achieving long-term economic sustainability:

- The distribution of curation roles across the network of stakeholders is critical for understanding both the factors encouraging sustainability in a particular digital curation context, as well as the risks that might prevent sustainability from being achieved.

The ecosystem diagram illustrates the major categories of stakeholder interest in curated digital assets. The distribution of stakeholder roles across distinct individuals or organizations is a key factor in influencing the nature of the economic risks associated with a given curation activity, and the kinds of economic remedies that are appropriate for addressing them. For example:

- The incentives to curate are strengthened when the Rights Holder for a particular digital asset is also a Current or Future Beneficiary. They are weakened when this is not the case.

- Allocation of responsibility for undertaking curation is clearer when the Rights Holder is the same as the Managing Agency. When this is not the case, it is less clear who should undertake the curation process, especially when ownership (i.e., the Rights Holder role) is split among many entities.

- The interests of Future Beneficiaries are often not represented in the curation decision-making process. Do Representatives of the Long-Term Public Interest adequately represent these interests and are they being taken into account in today’s curation decision-making?

- When Resource Providers are distinct entities from Current and Future Beneficiaries, it is vital that a compelling value proposition is made for providing resources to support long-term curation.

In summary, the stakeholder ecosystem is derived from the layering of three components one on top of the other: first, the set of stakeholder categories depicted in Figure 4; second, the pattern of relationships between stakeholders depicted in Figure 5; and third, the distribution of stakeholder roles discussed above. The contours of the stakeholder ecosystem associated with a given digital curation activity are fundamental in determining the potential “traps” that might impede the activity from achieving long-term economic sustainability. By extension, then, the design of an effective sustainability strategy to mitigate these traps must rest on a thorough understanding of the underlying stakeholder ecosystem.

The impact of the stakeholder ecosystem on the economic risks faced by a digital curation activity can be illustrated with an example. ArXiv is a repository for e-prints in physics, mathematics, and other sciences. Access to the contents of arXiv is free and open to all. Over the years, arXiv has become a key mode of scholarly communication between scientists, who rely on it both as a means of exposing their work to colleagues around the world and eliciting comment and discussion. Cornell
University currently serves as the Managing Agency for arXiv, and prior to 2010, was the sole Resource Provider as well. Although faculty and students at Cornell certainly benefit from use of arXiv, the vast majority of those who obtain value from arXiv – that is, the Current Beneficiaries – are unaffiliated with Cornell University.

A key feature of arXiv’s stakeholder ecosystem is that the Resource Provider is distinct from Current Beneficiaries. Those who benefit from using arXiv are not obligated to contribute toward its costs, and indeed have a weak incentive to do so, since they are not excluded from the benefits even if they contribute nothing. This creates a free-rider problem, since those who benefit from arXiv are free to do so without contributing resources to its ongoing operations. This situation was sustainable only as long as Cornell University was willing to fund arXiv’s approximately $400,000 annual budget on behalf of the general scientific community. This in turn introduces an economic risk into the curation activity: the Current Beneficiaries have little incentive to provide resources to support a curation activity undertaken in their interest, which calls into question the ability of the curation activity to marshal sufficient resources on an ongoing basis to continue its operations. One way to mitigate this risk is to induce or persuade Beneficiaries to contribute to arXiv – that is, to combine the roles of Resource Provider and Beneficiary in the same entities.

In 2010, Cornell made its first step toward doing just that, by issuing a call to other institutions whose affiliated faculty and students benefited from arXiv to make voluntary donations to support arXiv’s operations, with the requested donations scaled to each institution’s level of use (as measured by number of downloads). In describing this new funding strategy, Cornell noted that “[s]cholars worldwide depend upon the stable operation and continued development of arXiv. Sustainability is best assured by aligning revenue sources with the constituents that realize value from arXiv, and by reducing dependence upon Cornell University Library’s budget. We have decided to pursue a collaborative business model that will engage the institutions that benefit from arXiv”. In short, Cornell has realized the potential pitfalls of a stakeholder ecosystem where beneficiaries are completely distinct from Resource Providers. The donation program is a first step toward re-aligning the ecosystem such that the roles of Resource Provider and Beneficiary become merged within the same entities. It is significant to note, however, that Cornell considers the donation program to be only an interim solution until a long-term business plan can be worked out. An obvious weakness of the interim plan is that it is voluntary, with no “repercussions” if institutions choose not to participate. A long-term strategy might consider restricting arXiv access only to those institutions that contribute toward its upkeep. However, given its long history as a freely available resource, such a strategy is bound to encounter resistance.

This example demonstrates that knowledge of the stakeholder ecosystem underlying a particular digital curation activity helps identify potential economic risks impacting long-term sustainability, as well as possible remedies to overcome them.

8. Sustainability Concepts, Glossary, Definitions (CR)

Likely a late contribution (terminology may change) but can start now.
9. Tools (BL)

Early stage.

10. Examples and case studies

The Blue Ribbon Task Force looked at four different sustainability contexts: Scholarly discourse, Research Data, Commercially-owned cultural content, and Collectively produced web content. These are useful examples to revisit in the light of the developing reference model.

Each analysis will address 7 important questions:

- Who benefits from use?
- Who selects what is kept?
- Who owns the resource?
- Who preserves (or manages) the resource?
- Who pays?
- What are key attributes specific to this context?
- What are the key risks?

10.1. Sustainability in context: commercially-owned cultural content

The first example is commercially-owned cultural content.

This term addresses most of what we think of as traditional cultural content, updated for the modern era. Books, Magazines, Movies, recorded music, broadcasts are all obvious areas affected by the digital revolution. Fine art, theatre, dance and similar areas are probably less affected; the visceral, real life experience remains important in those areas. It is already clear that traditional business models for recorded music and movies are affected by the move to the Internet, and the reaction by the rights-owners has been damaging to those who need to curate and preserve this material.

Note: this is an extremely broad area, and this analysis must perforce be somewhat superficial. There is a great opportunity for further study here.

Who benefits from commercially-owned cultural content?

The first and obvious answer is that we, the public, benefit from this material. It is part of our life and our society would be the poorer without it. This content is (or becomes) our “digital cultural heritage”.

However, more to the point of this question, the rights owners (in the scenario, commercial rights owners) aim to benefit from the exploitation (use) of the content. This benefit lasts for the economic life of the content (almost a tautology), limited as that theoretically is by the term of intellectual property rights such as copyright.
In the past the economic life was often comparatively short for much material (e.g. magazines in paper form). The copyright might last for another hundred years or so, but rights owners effectively had one shot at revenue, when the issue hit the newsstands. A small amount might be made from back issues, but there was essentially no continuing revenue. Today, the economic life of content is growing ever longer, with new possibilities for re-purposing, re-mixing and re-releasing content.

In reality of course, there has always been a form of re-purposing, re-mixing and re-releasing going on in the cultural heritage area. Authors, artists, musicians take today’s content (or yesterday’s) and turn it into new content, sometimes transformed beyond all recognition. The nature of these opportunities has changed, not the general principle.

A side effect of the move of cultural heritage content from the analogue (for which most copyright laws were designed) to the digital world is that perfect copies can trivially be made; I can have a copy without taking yours. Digital content is, in economists’ terms “non-rival in consumption”. Copyright laws are being extended wider to try to preserve the monopoly for rights-owners, while a proportion of the public rebels and finds ever more ingenious ways to consume the content without rewarding the rights-owners (given the label of “copyright pirates”).

Who selects what is kept?

In the first instance the commercial owners of cultural content are often sponsors for the creation of the content (e.g. providing advances to writers, financing movies, employing journalists or other writers, photographers etc), and there are in-built selection processes in that sponsorship (often known as the editorial process). Publication often used to be on a “fire and forget” basis, as noted above; there was often little continuing revenue. The extended economic life of cultural content is partly due to low marginal continuing costs for holding the content. It only needs a small proportion of the content to make a significant amount of money for the overall costs to be justified (this is the “long tail” argument”). The extended economic life is also due to the change in practice from selling the content to leasing it, retaining all rights and binding the user with a contract. Provided the cost of de-selection remains higher than the cost of continuing to make available, content is likely to stay available. However, it is at the whim of the content owner, and is susceptible to accidents of technology change.

Commercial content owners cannot easily distinguish the kinds of actions required for preservation from those for piracy. In particular, actions taken for preservation that include making the content available by the preserver, are close in effect to those of pirates. Commercial owners therefore fear preservation by others, in case it erodes their market. Even where compulsory legal rights such as legal deposit for non-print exist, commercial content owners have lobbied for access to be highly restricted (e.g. to the premises of the libraries concerned). By default, the content owner defines who can preserve their content and what can be preserved.

Cultural heritage institutions are generally charged by society with the preservation of the cultural content they hold. This role is badly affected by the change from cultural content in artifacts they own to cultural content in products they license. However, institutions are battling to carve out roles that will allow them to preserve digital material, e.g. through partnerships such as HathiTrust. These institutions select content based on their long-term expertise in assessing future value for their particular
community. This localization of selection is their weakness (in being divided against the content owner) but also perhaps their strength (in that individually their threat is small).

**Who owns the resource?**

Almost by definition, these resources are owned by commercial entities. However, rights may also be highly fragmented. See for example Lessig\(^\text{xx}\) (2010) where the lead story element reports the difficulty of renegotiating expired rights to make remastered DVDs from old documentary films. Multimedia resources in particular may have very complex webs of rights with various conditions on them.

Sometimes a commercial entity will have all the rights to a resource, e.g. where an employee created the resource, or rights have been assigned. This may have some advantages for curation, as at least the path to agreement will be clear.

All too often, however, the ownership rights to a particular digital (or indeed, physical) object may be unknown. There are many books, movies, images and datasets whose ownership is not recorded. If ownership cannot be established after due diligence, these are referred to as orphan works. There have been proposals to change copyright law to put procedures in place for dealing with orphan works, but there are few success stories yet. To the contrary, the Authors’ Guild and other organizations are suing\(^\text{xxi}\) HathiTrust and others, partly over proposals for making some orphan works available, even under controlled conditions.

**Who Preserves (or Manages) the resource?**

In the context of this scenario, the commercial owner acquires the resource and does what is needed to make it available and earn a return from it. The longer economic life mentioned above provides incentives for them to keep such resources available in some cases (and in turn, keeping the resource available extends its economic life).

In other circumstances, where the resource may be in the public interest but is not likely to earn a current or short-term return, external incentives would encourage them to do so (e.g. tax credits for preservation if available).

Commercial organizations are rarely interested in preservation over the long term in the public interest; they are interested in keeping the resource available over the short to medium term in their private interest. Long-term preservation in the public interest usually falls to stewardship organizations; there are many kinds of these, ranging from the obvious library, archive and museum examples to various kinds of Foundations and Trusts. In the digital world, they can usually only carry out their preservation actions on the basis of permissions granted by rights owners, or on the basis of material entering the public domain.

It is worth mentioning one other group, not precisely definable. In some areas (e.g. computer games, for instance) there are amorphous, distributed networks of enthusiasts who are undertaking preservation actions, sometimes against the expressed wishes of the rights owners. Society may have cause to be grateful for groups such as these in the future.
Who Pays?

In the first instance, the commercial owner pays to make the content available and keep it usable for the economic life of the material. Of course businesses are not charities, so this translates into the public paying directly or indirectly for access to the content; the commercial owner is interested in getting a return on their investment.

Once the economic life of the content is over (which might come before the expiry of copyright), then stewardship organizations might take on the task on behalf of their communities. Academic libraries, for instance, will collect content of future relevance to their community. Most are still relatively new to collecting and managing digital content for the long term, but changes in this direction seem inevitable. National libraries will have similar interests, and increasingly will have Legal Deposit powers that will enable preservation activities.

There are also emerging self-organizing groups of interested stakeholders (largely after the economic life of the materials are over, but quite often before, and without any legal basis) who are willing to collect and manage content in which they are interested.

There is effectively no “second-hand” market for most digital cultural content, other than in “hand-held” form (e.g. DVDs, etc).

What are key attributes specific to this context?

Commercial owners have the preponderance of decision-making power when it comes to deciding who preserves, who pays, and who selects. There are signs they are wielding that power based on a combination of fear and short-term interests, to the detriment of the long-term public interest.

Preservation decision-making is heavily impacted by whether or not the economic life of the materials has ended. In the digital world, this usually relates to whether copyright has expired, which puts almost all digital material at risk.

Hand-offs between commercial owners and non-commercial preservation organizations are especially key in this domain.

Of all the scenarios, perhaps this is the one with the starkest contrast between materials with an economic value vs. a public value, and therefore there is a critical need to integrate the incentive to preserve for private interests, and the incentive to preserve for public interests.

Key Risks?

There are many risks here. However, essentially they boil down to a few:

- the risk that commercial owners will continue to hug the rights to their content close to their chests, preventing preservation and leading to loss of material of great future value
- the risk that stewardship organizations will take action to preserve and make available (within limited contexts) material still in private ownership, and get sued for their trouble
• the risks associated with material where rights-owners cannot be traced (“orphan rights”), or where the rights are too complex to unravel (as in some multimedia content)
• the risks that powerful commercial rights owners will successfully lobby to extend the scope of intellectual property rights, reducing the scope of the public domain, leading to loss of our cultural capital (and potentially damaging other businesses, and potentially themselves)
• the risks that those rebelling against difficult and expensive rights regimes will further stimulate and feed these lobbies, leading to ever greater penalties for those who are trying to do the right thing by society and the future
• the risks that all attempts to continue previous business models will fail in the face of this continued rebellion, leading to the collapse of major commercial businesses in the cultural field without stimulating societally-valuable alternatives.

10.2. Sustainability in context: collectively-produced web content

The second example is collectively-produced web content. This term addresses much of the area sometimes referred to as “Web 2.0”: content that is produced by members of the public acting for their own benefit or for the public good. It would include major resources such as Wikipedia, Facebook and its ilk, but also more amorphous concepts such as the “blogosphere”, Twitter and other instant-messaging sites, and even perhaps community gaming.

Note: this is an extremely broad area, and this analysis must again be superficial. There is a great opportunity for further study here.

Who benefits from collectively-produced web content?

Wikipedia is perhaps the most obvious example in this category of collectively-produced web content. It may however be rather misleading as an example, as there are significant differences between it and almost all other examples. It doesn’t just stand at one end of a continuum, it is a far outrider. There can be few reading this who have not at one time or another (or even very often) made use of Wikipedia.

Many other examples are much more complex. It’s probably true that many web 2.0 resources are produced primarily for the benefit of the creator(s). However, that benefit often accrues from features provided by an underlying platform (e.g. Facebook, Wordpress, Twitter), in combination with the social features of the systems. That is, these resources get value not from themselves, but from network effects resulting from linkages and interactions with other elements of their environment (and other environments).

A result of this is that short-term benefit is much clearer than long-term benefit. There are also well-known examples of long-term dis-benefit resulting from earlier contributions. Youthful or hasty indiscretions may reflect badly in certain situations. Issues like this have influenced questions of privacy and also control over deletion of contributions; these are treated differently in different platforms.
Strong cases have been made for the long-term value of some parts of these resources, such as blogs covering major social events such as general elections, or upheavals such as the “Arab Spring”. However, for much of the resource there is no clear-cut use case for long-term preservation, beyond vague assertions like “will be of interest to future scholars”.

One problematic characteristic is that those who benefit from long-term access are almost completely distinct from those who benefit from short-term access.

From the point of view of the platform provider, the content is a vehicle to attract usage, most often with the intent of attracting benefits in the form of one or both of advertising or premium subscription revenue.

Who selects what is kept?

This question operates at two levels: the platform provider and the content provider.

Decisions by the platform provider are likely to be made in terms of financial or other business benefits that do not relate directly to the benefits to the contributors and readers. There are many examples of community content platforms that have been closed down by their providers for business reasons. One of the most notorious is the closure of GeoCities, but there are many examples of platforms either closed or at risk. There appear to be no commonly accepted protocols for closing down such a resource, and providers generally do not provide mechanisms for wholesale transfer of all or any individual’s content to another resource. Community efforts have occurred to capture some or all of the content in several cases (including GeoCities), but the resulting content may not be easily accessible (it is based on an Options Strategy, capturing the resource in a basic way to leave the option to do more work on it later, rather than losing it entirely).

[For sheer numbers, see the Archive Team's Deathwatch: http://www.archiveteam.org/index.php?title=Deathwatch. See also TechCrunch's DeadPool, which gives a bit more background information on some http://techcrunch.com/tag/deadpool/.

The boundaries of the resource are often ill-defined; in fact, it is not clear what is the resource. The possibilities include the platform, the aggregate of contributed content, or the set of content contributed by an individual or group (who may think of their content as being entirely separate from other content on the same platform).

Time is a real factor: the resource is dynamic with additions and subtractions constantly being made. What is the authoritative version of the resource?

So the answers to “who selects?” are various:

- individual contributors select what is created and (where this facility is available) what is changed, deleted or extracted, and what comments to keep.
- some memory institutions are selecting content for preservation under various criteria. These include web archiving activities, and potentially specialist blog archiving activities. Another example is the Library of Congress agreement to archive the Twitter stream.
- where a platform becomes at risk, the platform provider will make business decisions based on balancing costs against revenues and other business benefits when deciding what to keep.
Who owns the resource?
Ownership of a resource comprising community-contributed content is extremely murky, and the IP Rights will be very complex. Rights will exist to the platform and its software, to widgets and other elements used by contributors, to designs and templates used by many contributors, to the individual content items created by contributors, and to comments left against other contributors’ content (sometimes part of a long chain of comments). Few resources make explicit reference to rights (such as declaring that both the content and comments contributed to a particular blog would be covered by a Creative Commons licenses). Quotation of potentially significant parts of other people’s posts is another common feature that can result in increasingly complex chains of ownership.

So overall the ownership is potentially diffused among all the contributors.

Most contributors never read the full terms and conditions under which they operate, but may be surprised to learn that they often grant non-exclusive, irrevocable rights to the platform providers. Sometimes these rights are qualified by words intended to make clear that they are for the purposes of the operation of the resource (and such rights are indeed essential), but sometimes the rights appear more extensive [example?].

Who preserves (or manages) the resource?
In the first instance the platform provider (or hosting site) manages the resource. This is potentially a reasonably long-term solution for content of continuing interest. However, platform providers are generally commercial organizations, and certainly have real costs to pay, so it is wise to plan for their interest to be temporary.

Platform providers have very weak incentives to act for the long term, as preservation is generally not part of their institutional mission. Although it would be helpful if platform providers made hand-off plans and post-termination agreements, these are not generally part of the landscape of entrepreneurial businesses. Indeed, attempts to sustain a resource beyond its financially viable life may mean no resources available for orderly transfer or hand-off to any other party, including the contributors themselves. At best, contributors and others can hope for a window during which transfer of material will be possible. Community actions for mass transfer are often explicitly prohibited in terms and conditions, but volunteer or self-organized groups may be determined to extract content anyway. However, these tend to be hastily-organized, post hoc activities, resulting in archives of significantly lower value than the original; they could be described as low-cost option strategies, leaving open the door for future work, that otherwise would have irrevocably closed.

It is of course possible that contributors have (some of) their own content on their own computers, anyway. However, these fragmented atoms of content do not have the power or value of the aggregated content in context.

The value proposition is too diffuse, or not compelling enough at this point for many stakeholder groups to see preservation as their problem (although experiments have been run with blog archiving, and the Library of Congress is reported to be ingesting the Twitter stream, and is collecting legal blogs http://www.loc.gov/law/find/web-archive/legal-blawgs.php). Despite its clear importance, we know of no activity to preserve Open Source hosting sites such as SourceForge.
Who pays?

The question of “who pays” tends to be quite unclear to the majority of contributors and readers of community-contributed sites. Facebook proudly claims it is “free and always will be”. Some sites (e.g. flickr, Wordpress etc) operate on a “freemium” model: free for basic use, but a subscription will buy enhanced capability, more storage, more services etc.

The “free” model works because the marginal cost of individual contributions, or even individual blogs or pages, is extremely low, whereas value accretes from the aggregation. Large numbers of writers and readers are very attractive to advertisers, and sites such as Facebook have been ruthless in changing the rules to enable them to monetize this value. We reflect this “advertiser pays” approach in the sustainability ecosystem by including them in the “current beneficiary” class of stakeholders, as they benefit indirectly from use of the resource by its primary users.

Some approaches (Open Source Software and Wikipedia being examples) are driven by a strong volunteer emphasis. Much of the work is contributed gratis, either by individuals, or by employees paid by their organizations to contribute because of perceived value. There are always real costs in such activities however, such as hosting costs that must be met each year if the activity is not to fold. The high profile example of this is Wikipedia’s annual drive to raise sufficient funds to continue operating. Some of these activities can be operating on very tight budgets, and small miscalculations can make them unable to pay their bills, and may force them to cease operations suddenly.

It is important when thinking of costs, however, to realize that costs are not just financial. Community-contributed content requires the involvement of the community to generate and sustain content. That volunteer effort is part of the “cost” of sustaining the digital assets, just as much as the money to pay hosting and other costs. Many such projects have failed because they could not attract (or retain, in the case of Myspace) sufficient interest from their intended communities.

What are key attributes specific to this context?

This is a very new form of content, and is itself highly varied; there are few useful parallels in the analogue world that can be used as a guide for preservation decision-making.

Unfortunately, there is little awareness among most stakeholders that preservation is an issue, although when it does become an issue (as in the proposed or actual closure of services), some react with anger and a strong sense of betrayal.

There is a great deal of uncertainty about most aspects of this content and its platforms. Most platforms do not succeed in attracting a critical mass for survival (hence the large numbers of casualties), and some that appear to do so may get overtaken by a more agile, more feature-rich, or perhaps just (temporarily) “cooler” competitor, as Facebook overtook Myspace.

The value proposition is too diffuse, or not compelling enough at this point for many stakeholder groups to see preservation as their problem (although experiments have been run with blog archiving, and the Library of Congress is reported to be ingesting the Twitter stream). Despite its clear importance, we know of no activity to preserve Open Source hosting sites such as SourceForge.
Decision-making power may be too diffused among thousands of contributors, making collective action nearly impossible; there are not enough decision-makers with a "big picture" view. Owners of content platforms (e.g. Yahoo as owners of Delicious, flickr etc) may become too focused on bottom-line problems to worry about preservation, or even exit strategies.

**Key risks?**

The Gartner Hype Curve (see Wikipedia article http://en.wikipedia.org/wiki/Hype_cycle) is widely referenced and sometimes derided, not least for its implication that everything eventually succeeds (clearly not the case). It does however hint at the serious risks that face new technologies:

"'Peak of Inflated Expectations' — In the next phase, a frenzy of publicity typically generates over-enthusiasm and unrealistic expectations. There may be some successful applications of a technology, but there are typically more failures."

Much of the financial case for platform providers to host this “free” content rests on successfully capturing a sufficiently large chunk of the market to bring in enough advertising revenue. There have been many startups in different niches that have attempted this, and inevitably many have failed or are likely to fail, usually most often after having attracted some community-contributed content. The circumstances of (impending) failure do not make organizing an orderly exit or handoff strategy likely, as management tend to be driven by decreasing or inadequate revenues, dwindling funding pools, and desperate attempts both to obtain more contributors and more capital, while cutting costs. This inevitably leads towards lost content. The decline of Myspace reminds us that even market leaders can be dislodged.

Given the risks they face and the community-contributed content they hold, it would be advisable for these social sites to have an explicit exit or hand-off strategy.

**10.3. Sustainability in context: research data**

The next example is research data.

It should be said at the outset that this is an extremely broad term, with very different meanings across the various scholarly disciplines, and even within the same discipline given different approaches. Nevertheless, we believe there are useful things to say.

**Who benefits from use of research data?**

The major person to benefit from good management and preservation (in the short term) of research data is the researcher herself. Managing research data well is a great help to researchers.

Once papers are written and in the publishing process, having research data accessible to reviewers can be helpful. Likewise when new project proposals are written having data accessible means greater confidence in referees.

Some research funders require a search process for pre-existing data before certain kinds of research are undertaken. Archiving data can help unnecessary experiments being run.
In some areas of observational research archiving data is a requirement, as observations cannot be made at a later date under the same conditions. This is particularly true of environmental and social sciences. Medical sciences also have strong requirements for research data to be archived. In all these cases the archived research data can be seen as another research instrument with data from the past available for re-use. Some researchers build their careers on re-analysis of existing research data.

Finally, the public at large has an interest in research data. There may be few who can make sense of it, but there are many examples where the public has made good use of data archives. It is worth remembering that many researchers enter “the public” on leaving the research area, and some of these individuals have as much skill in making sense of existing data as current researchers.

Who selects what is kept?

In the first instance the researchers themselves select the data they wish to keep, and their purpose is to further the research itself. These selections should be influenced by the data management plan (although we recognize that these are not yet common).

Researchers also select the data they wish to use as the basis for the argument written up in their publications. This selection may be influenced by reviewers of referees. These data should be kept available in a static state (if possible) in order to be accessible to readers who might want to use them to validate the publications’ conclusions.

When (or if) the data move from the researchers’ control into a data archive or data repository, the selection will be done by the relevant data archivist (or equivalent), informed by selection criteria or collection guidelines. In some cases the selection process may be influenced by (or carried out by) a peer-review panel of experts.

Who owns the resource?

This can be a seriously difficult question for much research data. The answer may depend on the nature of the data and the legal jurisdiction you live in. It may also depend on any prior agreements written into project proposals, memoranda of understanding, data management plans, etc.

Roughly speaking a fact is not copyrightable, but the expression of the fact may be. From this stems a great deal of complexity. In practice, you may never be certain who owns research data in (for example) a multi-group, multi-institution, international collaboration.

In practice, most research data is treated as being owned by the researcher or research group who generated it and have custody of it. Many researchers regard their data as being advantageous to them and do not wish to share it, holding it almost as a trade secret. There are also moves suggesting that researchers should explicitly disavow ownership of their research data after an initial period, using Creative Commons CC0 tool to put data into the public domain, so far as is possible in their legal jurisdiction. This reduces problems in re-using such data.

Data archives will usually not claim to take ownership of data, but operate under a license from the researchers to preserve the data and make it available.
Who preserves (or manages) the resource?
Again, in the first instance the researcher manages the resource. Practice here will vary widely. Some laboratory-based science requires data management according to strict protocols, using lab notebooks to record details. Much individual scholarship will be based entirely on the chosen practices of that individual. Many research groups will be unspoken amalgams of individual practice.

For data associated with a publication, in many instances the publication (e.g. a journal) will have mechanisms such as supplemental data for holding data supporting an item that does not fit within the rhetorical text. In many cases that supplemental data will be frozen as tables or images in a PDF file; not at all re-usable. It would be better for such data to be deposited in an institutional (or department, or research group, or subject) data repository in a machine-processable form.

Data with an expected longer life should not be left attached to a personal or even departmental web page, but should be moved to some kind of repository or data service with a more sustainable future.

Who pays?
The $64,000 question! Those who fund research are clearly prepared to pay for the management of the data required for the research during the project duration. Some contributions to this funding may come from other funding streams, such as institutional infrastructure funds.

A research group that has some longevity greater than a single research project (or a department of researchers in related areas) will usually need some mechanism for retaining data in a re-usable form, as it forms part of their intellectual capital.

At some point, however, data of sufficient value should be handed off elsewhere. If a relevant subject or discipline data archive or service exists then that should always be the first choice. If not an institutional data archive would be helpful. If one does not exist, chivvy your librarian or institution’s research director to get one established!

One factor limiting the spread of subject data archives (and the sustainability of those that exist) is that money for infrastructural support such as data archives is in direct competition with money for more research. This is a real limiting factor on the sustainability of data archives, and is one of the reasons behind the infamous demise of the UK Arts and Humanities Data Service.

What are key attributes specific to this context?
The late Jim Gray from Microsoft invented the term “the 4th paradigm” in 2007 to represent research based on analysis of masses of collected data (see Hey, Tansley, & Tolle, (Eds.). (2009)xxii)

The idea of using other peoples’ data is as old as research but has changed greatly in modern times. The Internet and the Web especially have made data sharing feasible in ways that were unimaginable before. However, data sharing has still not migrated into common practice amongst many researchers (particularly since many senior researchers started their careers well before current capabilities came into being).
It is worth also noting in passing that it is entirely wrong to treat research data as if they are mere extensions of familiar text objects. Research data varies dramatically in scale in at least 5 different ways: size of data object (from tiny to enormous), numbers of objects (to the billions and beyond), and rates of deposit, rates of change (yes, change), and rates of access. These and many other features make research data potentially entirely unlike data or text that have come before.

**What are the key risks?**

Risks for research data management, curation and preservation are legion.

At point of capture, there are risks of poor data management, of poor context capture (i.e. metadata etc), and of course downright fraud.

Throughout data processing (which may extend for years, of course), there are all the risks above (since new data products will be generated during processing). There is also the risk that the “computational lineage” will not be (adequately) captured (see Bose and Frew, 2005\(^{\text{xiii}}\)).

At the stage of archiving, perhaps the principal risk is that the focus of interest has moved on. The Principal Investigator is writing the grant to follow up on the follow-up project. The staff have mostly moved on. The PhD students are in the last stages of writing up and have other priorities.

These risks are compounded if the PI is inclined to horde data for some supposed advantage (the suggestion that “the coolest thing to do with your data will be thought of by someone else”, attributed to Rufus Pollock\(^{3}\), doesn’t please some people). There will likely be some uncertainty on the rights or permissions needed. There may be privacy or ethical issues. These and similar arguments will decrease the desire to archive data.

Many feel that taking short-term positions (“the data are on my web site”) is enough.

And to top this, in many disciplines and institutions, there will be very limited options for long term archiving anyway. It’s not surprising that the default option is… do nothing.

Once the data are archived, if the archive itself comes under threat. there can be limited handoff options. Given the competition for funding for infrastructure versus research mentioned above, this is always a possibility.

**10.4. Sustainability in context: scholarly communications**

The final example is the scholarly communications system.

It should be said at the outset that this is not an homogeneous term, with different approaches being tried. We characterize the system in four sectors: monographs (still mostly print-based); the formal, mainly commercial, mainly subscription-based (closed access) electronic journal sector, the formal, open access electronic journal sector (partly commercial and partly volunteer/charitably supported), and the informal, open access sector (mostly scholarly blogging but also working papers and grey literature). [Query ArXiv, REPEc etc.]

---

\(^{3}\) See http://blogs.talis.com/nodalities/2007/05/xtech_day_3_rufus_pollock_and_.php
Who benefits from the scholarly communications system?

Clearly the research community as a whole is the main beneficiary of the scholarly communications system (apart from the shareholders of the major scholarly publishing companies, of course). The system exists for, and is mostly strongly supported by researchers, who act for it in various unpaid capacities, including author, reviewer, editor and reader.

To a lesser extent, the system also supports the teaching community and students, not necessarily with the very leading edge research but with past publications.

Both business and government benefit from scholarly communications, and some members of the public also benefit. However, the most common commercial scholarly publishing sector business model tends to exclude members of the public.

Who selects what is kept?

The fundamental principle in nearly all scholarly publishing is that the community of scholars selects the content to be published through the mechanism of peer review. Almost all of the formal scholarly publishing system (whether closed or open access) is based on selection by peer review. The peer review system is not without its critics but there is no wide-spread acceptance of alternatives. Peer review is undertaken by members of the editorial board, or nominated experts in the field; the editors usually undertake selection of peer reviewers. Peer reviewers volunteer their services, as of course do most authors (at least in the electronic journals part of the system).

In a few journals the editorial staff play a large part in selection for all or part of the content. Various experiments with post-publication peer review have been made without achieving widespread success.

Open Access journals mostly use the same peer review approaches. Public Library of Science has been pioneering a light-weight version of full peer review, but most journals continue with traditional, full peer review.

The informal sector is not based on peer review, and influence of any contributions in that sector depends on many different factors, including the reputation of the author, the accumulated value of their content, and the social networks (broadly meant) that they are involved in.

Beyond the publishing stage, the mechanisms for selection for preservation are much more broad-brush. Because much of this content is encumbered by Intellectual Property Rights, selection is often a negotiation between archives and publishers. Libraries do get involved in some of these initiatives as well, and will reflect the perceived needs of the communities they represent, usually as expressed in their collection development plans.

Who owns the resource?

Theoretically when a work is written by an employee of an organization under the terms of their contract of employment, the rights to that work belong to the employer. However in the case of scholarly publishing the convention is the reverse; the author is treated as the owner of the rights and makes assignments or licenses the work as they decide. This does of course split ownership so that potentially one could be
negotiating with many individuals rather than a relatively small number of institutions involved in any one work.

However, the commercial scholarly publishing sector mostly operates on the basis of requiring assignment of all rights from the author to the publisher. This places the publisher in a very strong position to control (and deny) archiving opportunities. Unfortunately, archiving can be seen as destructive of publisher business models, and publishers have proved very reluctant to agree to archiving requests.

The Open Access publishing sector tends to offer content on a different basis: free to consume, and either pay to submit for publication, or pay to publish, or (sometimes) free to publish. Some (but not all) Open Access publishers use Creative Commons licenses, which generally will allow archiving of the content. Many will not take ownership of all rights, taking only enough to license them to publish the material. This will leave the rights with the authors (potentially very many authors per contribution); if the publisher has not made the publishing license explicit then theoretically archiving would require negotiation with all the authors.

Publishers of scholarly monographs also typically take rights sufficient to publish a first edition, with some reversion to the authors for out of print works.

For informal scholarly publishing (blogging etc), ownership will rest with the author. There is very often no explicit license, and so it is difficult to know whether archiving is permitted.

Who preserves (or manages) the resource?

In both the closed and open access cases, the publisher manages the content and makes it available. The content is rarely if ever down-loaded to an institution; content is consumed from central servers on an as-needed basis. Effectively content is leased rather than owned. Subscription-based commercial publishers employ mechanisms to detect systematic downloading of “their” content, and do not hesitate to cut off entire institutions if their algorithms detect suspect unauthorized mass downloading.

This situation is in complete contrast to the situation in the print version of scholarly publishing, where a library owns the content of the material on its shelves. Preservation in the print world happens effectively by diaspora; successful content is so widely spread that neither disaster, nor publisher failure, nor government censorship, nor malicious attack can destroy all copies. For electronic publishing, theoretically there is one copy on the publisher’s computer system, served up to the community on an as-needed basis. The risks in this case are different and potentially much higher. If the publisher is not well-managed, then a single disaster might wipe out all their content, as could a catastrophic business failure. Government (or other) censorship could cause the version of record to be changed, and malicious attack could destroy or subtly change all or parts of the content.

Libraries (as the primary purchasers of this content) have as a result been in a bind. The content is popular, and it is efficient for readers and for the libraries, but many have been reluctant to give up a simultaneous print subscription because of the risks.

In the past few years, various preservation approaches have come to the fore. A small number of national libraries have either negotiated deals or used their legal deposit powers to capture their own versions of published content. Portico has been established in America as a venture supported by some publishers and some libraries.
on a subscription basis; Portico collects the content and makes it available to its subscribers when some trigger event occurs.

The LOCKSS system (Lots of Copies Keep Stuff Safe) works on a different system. LOCKSS is highly devolved, with many member libraries, each of which operates a low cost “LOCKSS box”. Once the LOCKSS collective has struck a deal with the relevant publishers, and the publisher has installed some enabling code in their systems, then the LOCKSS box in member libraries will collect content (slowly) from the publisher web sites. The LOCKSS boxes also check up on each other, and will detect and correct content that is damaged. The member libraries continue to access this content from the publisher web site as normal, again unless a trigger event occurs. Unfortunately some of the bigger publishers are suspicious of this system, as libraries are acquiring copies of the content outwith publisher control, so some have not signed up.

There is a smaller, closed version of LOCKSS called CLOCKSS that the larger publishers are more willing to deal with, that again releases content when some trigger event occurs.

Informal publishing may in part be preserved through activities such as the Internet Archive, but in any event is similar to other community-contributed content.

Who pays?

Primarily, institutional libraries pay subscriptions to support the commercial closed-access scholarly publishing system. Some journals (particularly those sponsored by scholarly societies) also have individual subscriptions, but these make up a smaller part of revenue streams. Most journals also have a system for buying access to individual articles. Often the title and abstract will be freely accessible, but the full text (sight unseen) will require a fee of between $15 to $40, and may be time-limited to one or two days. (These prices seem set sufficiently high as to discourage rather than encourage access by this means, leading users more towards the subscription model). These publishers may also require publication fees for some kinds of content, although this may be less common than it was in the print world.

It is worth remarking (again) that authors and reviewers make their contributions to the system completely unpaid. For the larger publishers, business has been highly profitable, although there are signs of a slowing of the previously relentless growth in their profits.

Subscriptions from outside the community of authors do help sustain (and grow) this sector.

The Open Access sector operates in two ways. What is sometimes called Gold Open Access operates on the basis of paying to publish but free to consume. There is a relatively small number of commercial Open Access publishers (e.g. BioMedCentral). Several large commercial publishers make portions of their content available on an Open Access basis, sometimes at the choice of (and additional fee from) their authors, and sometimes as a matter of course after an embargo period.

Some research funders will allow publication fees as valid expenses in research proposals, and some research institutions and libraries establish pools for publication fees, but it is not known how widely either of these is taken up.
A criticism of this part of the Open Access sector has been that it takes resources from the authors rather than the consumers. To the extent that business or government contribute less as authors and more as consumers, they pay less under this model, and the academic sector as a whole pays more.

The second part of this sector is supported in other ways, and does not charge publication fees or access fees. Some of these publishers are volunteer efforts, some are house publications supported by institutions or organizations. Some might be supported by advertising (although this is rare).

The informal sector operates as describes in the community-contributed content scenario.

**What are key attributes specific to this context?**

Many of the key attributes are similar to those for commercial cultural content. The system is sustained in part because it is entwined with the reputation system for researchers, being linked to prospects for promotion (“publish or perish”), and in some countries (such as the UK) to substantial financial reward to institutions. Much of this is done not on the basis of the value of individual articles, but on “impact factors” established by a commercial concern (Thomson ISI) of journals; so publishing in a high impact journal gains more kudos than a better article published in a lower impact journal.

There are strong claims that the current system is broken in various ways, being subject to gaming, being selective, being influenced by a single commercial player, etc. We note but do not comment on this controversy.

The Open Access sector suffers from these same issues, but also tends to be on a more fragile financial footing. The Pay Once Consume Indefinitely approach is certainly a risk, requiring a constant stream of new content. The volunteer/charitable approach requires sponsors to continue to carry the (not inconsiderable) costs of the editorial process in particular, as well as hosting the content.

The informal sector suffers from all the issues of the collectively contributed content scenario. It is also generally not widely valued by many peers, and is fraught with uncertainty in various ways.

**What are the key risks?**

The commercial, closed access subscription sector remains highly profitable, but is subject to increasing criticism and potential threats from the Open Access movement. For the foreseeable future there is no likelihood of decline toward unprofitability of the larger players, but reducing growth and margins are likely and they may be subject to more severe commercial decisions, such as the closure of less profitable titles, and further market consolidation.

Smaller closed access publishers are perhaps most at risk in a time of severe financial cuts to universities, where the Big Deal arrangements reduce flexibility for libraries looking to cut their costs.

The Open Access sector, by contrast, is mostly on a much more precarious financial basis. This particularly applies to the volunteer/charitable part of the sector that does not charge publication fees. The latter is very dependent on the enthusiasm, will and
persuasive ability of a relatively small number of individuals, and fortunes can wax or wane at any time.

Open Access publishers and smaller commercial publishers should be getting involved in initiatives such as LOCKSS, as low-cost ways to provide a safe exit should their ventures fail.

Risks for informal scholarly publishing are similar to those for community-contributed content.

11. References


Draft economic sustainability reference model