Developing Costing Models for Emulation Based Access in Scientific Libraries

Euan Cochrane
Yale University Library
Digital Preservation Department
New Haven, Connecticut, USA
euan.cochrane@yale.edu

Dirk von Suchodoletz, Klaus Rechert
Faculty of Engineering
Albert-Ludwigs University Freiburg
Freiburg i. B., Germany
{givenname.surname}@rz.uni-freiburg.de

ABSTRACT

Digital preservation practitioners are beginning to answer questions about the costs related to the long-term availability of digital information. In order to successfully model these costs archival systems and workflows need to be fully understood and their costs identified. This can become exceedingly difficult for complex access and preservation strategies like emulation. If emulation is to be considered in a strategy mix its cost components need to be gathered and understood so that institutions can develop informed preservation plans and decide which strategy to follow. The digital preservation community now has a systematic understanding of storage and repository administration costs, but emulation and surrounding services are still an emerging topic for memory institutions. While costs to produce bit-preservable representations of digital artefacts are relatively well known there is an array of rather unpredictable cost factors that need to be further researched. Many of these unexplored costs factors vary depending on the kind of digital objects and the objectives of the stakeholders involved in the activities.

General Terms

Case Studies and Best Practice

Keywords

Cost Model, Emulation, EaaS, Digital Object Access, Archival Workflow, Service

1. INTRODUCTION

Libraries, archives and museums already hold a substantial quantity of digital artefacts and receive an increasing number of digital-born objects with more and more complex structures. These objects require different handling from traditional analogue and static material. Complex digital artefacts must undergo new treatment with regard to methods and workflows to render them accessible to future users which requires memory institutions to implement or acquire from third-party novel types of services.

From a cost perspective digital preservation can be modeled as any other economic activity, i.e. as a usage-based service, or alternatively, the costs of digital preservation services can be designed following the approach used for insurance services. Incentives exist for funding digital preservation services when the benefits outweigh the costs of participating. The advantages of preserving digital artefacts extend from the fact that the discoveries of the future rely on the work of the past. Additionally, for research data, the maintenance of a complete and accurate scholarly record is essential for continued progress in research and learning [6].[1] Cost and business models for emulation services can be derived from a variety of different perspectives. Associated costs can vary heavily depending on object classes and levels of inter-institutional cooperation. Preservation planning and different levels of acceptable risk also influences costs as well as future stakeholders' expectations [23, 9].

Costs have a significant influence on the choice of a preservation strategy, but are inherently hard to quantify. Ultimately, the Total Cost of Ownership (TCO) can be the guiding figure for deciding whether or not a preservation strategy meets the needs of an institution within the constraints of its budget [10]. In addition, there is growing demand for understanding the costs of emulation services within memory institutions and further afield.[2] Institutions looking to implement emulation solutions are currently ill-equipped to do so, partly because there is little information available to provide their funding bodies on how much it might cost to do so. The TCO is also very useful for informing acquisition decisions for collecting institutions. Something that may appear to be a good-value acquisition that is well within the budget of an institution may turn out to be a cost-drain on the organisation once the total cost of ownership is taken into account. For these reasons, and in order to choose appropriate long-term preservation strate-

[1] E.g. to fulfill the requirement for reproducible code in computational science `http://www.recomputation.org/blog/2013/04/12/the-recomputation-manifesto/`.
[2] See various news articles on use of emulation to rescue old hardware, e.g. [15, 11].

gies, and assess preservation plans, proper cost models for using emulation solutions need to be available.

This paper focuses on costs which are directly and indirectly related to institutional emulation strategies. It takes the institutional perspective of a library or archive and ignores traditional repository and bit-level storage costs as they are already several noteworthy articles available on that topic [2, 10, 4, 23].

2. RELATED WORK

A consensus exist that the cost of preservation action must not exceed the estimated value of the digital object [8]. Nevertheless, it might be not entirely clear how to evaluate values of digital objects in different domains [12, 9, 19]. Economic models can be distinguished from cost models and business models, each of which is useful and may be essential for understanding an economic process, but neither of which can be used reliably except in the context of a broader economic model [9].

Early cost models for preserving digital information projected traditional library operation into the digital realm, estimating the efforts required to run repositories and access systems for documents like electronic volumes [22]. The model assumed that all equipment and software costs were capitalized over a life of five years and then replaced for obsolescence. The same cycle was projected for media refresh because of technological change, copying the objects from one bit-level storage to a new one. Equipment maintenance and operations costs were calculated as a proportion of the original purchase price. The personnel costs generated by management and systems engineering services were estimated as a proportion of the salary of a full-time employee including inflation.

The LIFE[2] report discusses possible preservation costing aims and approaches. To cost digital preservation activity two ways have been identified: A top-down audit of all preservation and repository activity; and a bottom-up lifecycle costing of activities relating to a particular content stream [3].

The JISC commissioned the development of application-neutral cost models for digital research data including consideration of different data collection levels and their requirements, the need for relevant documentation and metadata [6, 7]. One of the core goals of "Keep research data safe" was to identify potential sources of cost information. Recommendations hint that institutions repositories should take advantage of economies of scale, using multi-institutional collaboration and outsourcing as appropriate. Typically, once core capacity is in place additional content can be added at increasing levels of efficiency and lower cost.[3] The EU-sponsored 4C project[4] tries to boost uptake of the tools and methods that have been developed. The main objective of 4C is not to develop just another cost model but to ensure that where existing work is relevant, stakeholders realise and understand how to employ those resources.

[3]See `http://www.beagrie.com/KRDS_Factsheet_0711.pdf`

[4]See *Collaboration to Clarify the costs of Curation* self-description `http://www.4cproject.eu/about-us`

Successful digital preservation requires long-term planning. There is growing demand for "paid-up" cost models for digital preservation services[5] in order to be able to include provision for funding the long-term preservation of digital content produced by projects, within the projects' proposals. Paid-up cost models are also very attractive for institutions who seek to understand the TCO when making acquisition decisions or when deciding whether to accept donations.

To determine upper limits of acceptable costs it can be useful to change perspective: Billing models and use patterns of existing (non-digital) centrally managed repositories are relevant indicators of what content owners can afford to pay for managed storage services – independent of costs and benefits associated with retrieval [10, 1].

3. EMULATION USAGE SCENARIOS

The concept of emulation of legacy platforms has been included in digital preservation discussion for quite a while [20]. Nevertheless, compared to well established tools and workflows for traditional media, the tools and services for emulation like the KEEP emulation framework [16], and Tesella's Safe Deposit Box that was derived from it, or services like bwFLA Emulation-as-a-Service [21, 18, 14], are comparably new and there is not yet a great deal of experience of deploying these tools in memory institutions.

Within institutions working with digital artifacts there at least three primary use-cases for emulation. Emulation solutions can be applied for:

1. Appraising and/or selecting content in difficult-to-access formats or of dynamic, interactive content
2. Normalizing or migrating content between file formats
3. Accessing content and interacting with it

Each of these roles may present quite different usage patterns and therefore may require different cost models to support them. Below each of these scenarios are explored first and that exploration is then followed with an evaluation of the possible cost models that best support them.

3.1 Emulation in appraisal and selection

Emulation is of use when appraising and/or selecting content as it can give users the ability to investigate content within disk images, or within sets of older digital files and open them in software from the era in which they were created. This can give appraisers and selectors a much richer feel for the content they have presented to them and can help provide a much greater level of context than they might otherwise have had available. Emulation also allows all of this work to be undertaken within closed-environments that can be configured to not save any changes that may have been made (inadvertently or otherwise) during the process.

This appraisal/selection use case requires the organisation using the emulation solutions to have access to a limited

[5]See CNI/CDL model `https://wiki.ucop.edu/display/Curation/Cost+Modeling/Princeton` and `http://dspace.princeton.edu/jspui/handle/88435/dsp01w6634361k`

set of generic emulatable environments which have multiple software applications installed on them. They might, for instance, require one or two environments for each major operating system with different sets of software installed on each environment. For costing purposes it is useful to note that this scenario involves a limited number of emulated environments used by a limited number of users on a regular basis.

3.2 Emulation for Content Migration or Normalisation

Often the only software that can open a file (or present its contents with full integrity) is the software that created the file or was originally used to open it. This original software can often also save the content of the file into new files with different formats, and even when that is not an option it is normally possible to use operating-system level utilities, such as print-to-file applications, to save content in different, more accessible, formats. This approach can be useful when a memory organisation has a set of files that cannot be opened in modern software but for which the original software is available. Under this "migration by emulation scenario" content files are opened in original software running on emulated hardware, and the content is saved into a different format that is still accessible in modern software. For costing purposes it is useful to note that this scenario can be broken down into two distinct subsets with different usage patterns:

- **Just-in-case usage** Used for normalising[6] content at point of ingest. This scenario requires on-going access to emulated environments. These environments contain specific applications for each format that the organisation wants to normalise away from. In this scenario usage is unpredictable, and the emulated environments need to be available at all times just in case a file is acquired that requires normalisation. In this scenario the emulated environments are normally used to process only a small number of files at a time.
- **Just-in-time usage** Used for migrating content when software is completely inaccessible. This usage requires access to emulated environments on demand, when needed. The need for the use of emulated environments for just-in-time usage is usually identified well in advance of the actual use of the environments, and normally does not require emulated environments to be available at all times. Usage of emulated environments in this scenario is predictable and they are normally used to process a large number of files at a time.

3.3 Emulation for access

The most common scenario is to use emulated software to access content in old digital files or to interact with dynamic content. This scenario requires an original environment that includes an operating system and application software to be made available via an emulator. That environment is then deployed to access content stored in one or more digital files held by the organisation. This scenario also has multiple usage patterns, including:

- **On-demand use for specific access requests** In this scenario emulated environments are configured and made available via an emulator and/or emulation service on demand. This use-pattern requires software and emulators to be available but does not necessarily require emulatable environments to be pre-configured for immediate provision.
- **Comprehensive use for all content falling into predefined categories** In this scenario emulated environments are deployed to provide access to all artefacts that fall into a certain category (e.g. when the original interaction software is unusable on modern computers). This use-pattern requires pre-configured environments to be available immediately on request, and emulation access services that can scale to meet user-demand.
- **Mixed usage depending on user-community attributes** In this scenario usage would otherwise be the same as in the blanket-use scenario but it is artificially restricted for some purpose leading to low usage. For example access to content may be restricted to reading rooms in the content-controlling institution. This use-pattern would still require pre-configured environments to be available immediately on request but would not require extensive emulation infrastructure that could scale to meet demand.

These use cases identify a number of factors that help to clarify the best approach to provide the necessary emulation solutions:

- Frequency of use of emulation solutions
- Scale of use of emulation solutions
- Uniqueness of needed emulation solutions
- Timeliness required of emulation solutions
- Regularity of usage of emulation solutions
- Data Security requirements

When developing their own cost models organisations need to identify the use cases that are relevant to their institutions and identify the above factors in order to decide how to model, plan for and recoup the costs of providing these solutions.

4. DIFFERENTIATION

There are a number of components that contribute to the cost of using emulation. These cost components differ depending on how emulation is used and in what workflows it is used. Some workflows would supplement existing ones, others are novel. For example, depending on the type of delivery to be used for digital artifacts the artefacts may have to be copied from their original medium in pre-ingest to the

[6]Normalisation usually means unifying various formats of a certain domain like office documents into a single format, which serves as a standard within the receiving institution.

bit-level storage of the memory institution because of media decay and technological obsolescence [13] independent of the chosen long-term access strategy. Studies of media migration were e.g. done by KEEP.[7] The challenges and related processes are well understood and thus not part of the following considerations.

4.1 Emulation cost components

In order to begin developing cost models for providing emulation solutions it is first necessary to identify the source components that contribute to the TCO for the solutions (list of key cost variables and units, [6]). Once these costs have been identified it will then be possible to group the costs into the products and services that may make up the emulation solutions implement in organisations.

There are numerous cost factors that need to be considered when modeling long-term costs for providing emulation solutions. These include:

Costs related to hardware emulation software

- Emulator development, testing and maintenance costs
- Cost to access original hardware to validate emulation accuracy
- Emulator support costs
- Emulator use costs

Costs related to enabling non-expert access to emulators, e.g. via bwFLA Emulation-as-a-Service (EaaS)

- Remote EaaS software development costs
- Remote EaaS software support costs
- Local EaaS software development costs
- Local EaaS software support costs
- Cost to provide EaaS services

Costs related to intellectual property

- Operating system licensing costs
- Software application licensing costs
- Emulator patent-related costs
- Emulator licensing costs
- License management costs
- Software documentation and manuals copyright costs

Costs related to emulator and environment management

- Cost to configure and maintain environments for ad-hoc immediate usage
- Cost to document environments and provide unique identifiers/handles.

Costs related to documentation and user-support

- Documentation library creation and maintenance
- Cost to provide remote access to
- Cost to digitize documentation
- End user support for obsolete software
- Cost to provide seamless ”on-line” support within emulation solutions

For the purposes of this paper these costs include all staff costs and hardware costs with the exception of costs related to obsolete hardware needed to compare emulators against for quality assurance.

Regardless of the institutional context there are many emulation-related activities that would benefit from collaborative approaches provided as services in order to reduce the costs for each institution. There are many emulation cost components that could be shared across the community including:

- Development and maintenance of emulators
- Development and maintenance of emulation access services
- License management
- Configuration, management and preservation of installed software environments
- A software, file format and hardware documentation library
- Provision of the ability to run emulators at scale

Nevertheless, several non-shareable costs factors remain:

- Licensing
- Running local hardware
- Running emulators at scale
- End-user support at scale

Having identified the various components of cost that contribute to the TCO for emulation solutions it now possible to begin outlining the different ways these costs can be packaged into products and services which can be sold to internal stakeholders and/or clients.

Most emulation solutions and respective costs can be packaged and costed as fixed-cost products or variable-cost services. Table 1 gives examples of emulation related products and equivalent services:

[7] See `http://www.keep-project.eu/ezpub2/index.php?/eng/content/download/19824/99318/file/KEEP_WP1_D1.2a_v4.0.pdf`

Fixed cost "products"	Variable-cost services
Normalisation/migration environment	Normalisation/migration of "x" files
Emulatable environment	"x" hours of access to an emulated environment
Emulation software (emulators)	Emulation as a Service
Emulation experts	Emulation support
Software documentation Library	Access to a software documentation library
Software Licence	"x" hours of access to software
Local EaaS implementation/Emulation workbench	Remote access to Emulation as a Service

Table 1: Emulation products and equivalent services

5. POSSIBLE COST MODELS

Having identified the cost components that contribute to the cost of providing emulation products and services, possible products and services that might be used for providing emulation solutions, and scenarios that emulation solutions might be used within it is now possible to outline possible emulation solutions that might be used within organisations and to develop the cost models to support those solutions. Four models relating to four generalised example scenarios are outlined below. These models assume outsourcing the provision of the emulation services and/or acquiring the full solutions from a third-party provider. Costs for doing all of the work in-house would likely differ greatly depending on context, particularly in regards to managing software licensing fees. For example, costs for just running the hardware to support a remote access to emulation service (EaaS) are currently being determined but are definitely much lower than the overall costs included in these example models. The difference in cost is due to the number of factors related to providing these emulation services as a third-party provider, including (but not limited to):

- Administrative costs
- Legal costs
- Marketing/sales costs
- Human resource costs
- Emulator development costs
- Service development costs.

By assuming the provision of these services by a third-party this simplifies the models and helps to enable readers to understand how such services might be accounted for in their organisations. For example, trying to account for all of the cost components that might go into migrating one digital object from one file format to another can otherwise be quite challenging if this was being done "manually" within an organization. By assuming the provision of such functionality as packaged services the reader is better able to understand how realistic these might be for their organization to implement.

Model 1: small organisation using emulation for appraisal, selection and infrequent access

Considerations: Small budget, no in-house support
Requirements: Access to "x" emulation environments provided via an intuitive access system for appraisal and sentencing, infrequent access to a diverse set of remotely provided emulated environments for use in interacting with content for end-user access purposes, no automated migration of objects using the service offered
Appropriate Solution: Small comprehensive set of emulation products for appraisal and selection EaaS provided remotely (or locally depending on security considerations) for access purposes
Rationale: In this scenario the organisation requires a comprehensive set of tools to aid in appraisal and sentencing but these tools would be static and could be acquired as products. The organisation has an unpredictable need for emulation tools for accessing its content so would be best to use a service to provide these, especially given the lack of in-house expertise.

Component	Cost/Unit
Number of environments for Selection/Appraisal	15
Cost per environment	$500
Cost of emulation workbench tool	$500
Total cost of Selection/Appraisal emulation products	$8,000
Number of hours of emulation instances in EaaS per year	$ 520
Average cost per hour	$3
Total Cost for EaaS per year	$1,560
Emulation support services per year (including documentation access and end-user support)	$750
Total cost for emulation solution over 5 years	$19,550

Table 2: Example Cost Model 1

Model 2: Medium sized organisation using emulation for appraisal and selection, a medium level of access, and irregular content migration

Considerations: Medium budget, little in-house support
Requirements: Access to "x" emulation environments provided via an intuitive access system for appraising and sentencing content, access to a limited set of migration-by-emulation environments and services on an irregular basis and access to a large number of environments for accessing its content that would be used for around 5000 hours a year by users
Appropriate Solution: Comprehensive set of emulation products for appraisal and selection, EaaS provided remotely (or locally depending on security considerations) for access purposes, use of migration by emulation services for 1000 files per year
Rationale: In this scenario the organisation requires a comprehensive set of tools to aid in appraisal and sentencing but these tools could be static and could be acquired as products. The organisation has a medium level of need for emulation

tools for accessing its content so would likely still be best off using a service to provide these. The organisation has a limited need for migrating digital artifacts using emulation each year so would likely be best off using a service for these (table 3).

Component	Cost/Unit
Number of environments for selection/appraisal	15
Cost per environment	$500
Cost of emulation workbench tool	$500
Total cost of selection/appraisal emulation products	$8,000
Number of files migrated using emulation each year	1,000
Cost to migrate each file	$0.10
Total migration cost, per year	$100
Number of hours of emulation instances in EaaS per year	5000
Average cost per hour	$3
Total Cost for EaaS per year	$15,000
Total cost for emulation solution over 5 years	$83,500

Table 3: Example Cost Model 2

Model 3: Large organisation using emulation for appraisal and selection and for comprehensive use for content normalisation upon reception of the content

Considerations: Large budget, available in-house support
Requirements: Access to "x" emulation environments provided via an intuitive access system for appraisal and sentencing, access to a comprehensive set of migration-by emulation environments/services for migrating 150,000 files per year
Appropriate Solution: Comprehensive set of emulation products for appraisal and selection, use of migration by emulation services for 150,000 files per year
Rationale: In this scenario the organisation requires a comprehensive set of tools to aid in appraisal and sentencing but these tools could be static and could be acquired as products. The organisation has an extensive need for migrating digital artifacts using emulation each year. Depending on the variability of the environments needed for undertaking this emulation it might make sense to undertake this using in-house supported tools. If there is extensive variability in needed-environments a services approach might be more appropriate (table 4).

Model 4: Large organisation using emulation for appraisal and selection, as well as for comprehensive access

Considerations: Decent budget, available in-house support
Requirements: Access to "x" emulation environments provided via an intuitive access system for appraisal and sentencing, access to a comprehensive set of emulation tools for accessing digital artifacts
Appropriate Solution: Comprehensive set of emulation products for appraisal and selection, and access to a large number of environments for accessing its content that would be used for around 100,000 hours a year by users
Rationale: In this scenario the organisation requires a comprehensive set of tools to aid in appraisal and sentencing but these tools could be static and could be acquired as products. The organisation has an extensive need providing comprehensive access to its objects using emulation tools. Depending on the variability of the environments needed for undertaking this emulation it might make sense to undertake this using in-house supported tools. If there is extensive variability in needed-environments a services approach might be more appropriate (table 5).

Component	Cost/Unit
Number of environments for selection/appraisal	15
Cost per environment	$500
Cost of emulation workbench tool	$500
Total cost of selection/appraisal emulation products	$8,000
Number of files migrated using emulation each year	150,000
Cost to migrate each file	$0.10
Total migration cost, per year	$15,000
Total cost for emulation solution over 5 years	$83,000

Table 4: Example Cost Model 3

Component	Cost/Unit
Number of environments for selection/appraisal	15
Cost per environment	$500
Cost of emulation workbench tool	$500
Total cost of selection/appraisal emulation products	$8,000
Number of hours of emulation instances in EaaS per year	100,000
Average cost per hour	$3
Total Cost for EaaS per year	$300,000
Total cost for emulation solution over 5 years	$1,508,000

Table 5: Example Cost Model 4

5.1 Applying example cost models

The cost models outlined above are indicative examples at best. Actual costs for implementing emulation solutions will vary significantly and will depend greatly on the institutional context. For example, if the institution has an extensive legal team on staff then they may be better equipped to deal with the licensing issues. If an institution has emulation experts on staff then they may be able to configure and run some of the services themselves. When developing a cost model for the use of emulation in a particular real-world context an effective approach may be to:

1. Compare the institutional context to the examples out-

lined above and select the model that best fits with the context.

2. Form an initial model based on one of the selected examples.

3. Review the cost components outlined in the previous section to ensure all cost factors have been either: included in a product or service that has been accounted for, or to highlight missing cost components.

4. Add any missing cost-components to the model.

6. PRELIMINARY PRACTICAL RESULTS

A practical access experiment together with the Rhizome project[8] provided insight into dynamic costs of providing the hardware to support this service and possible usage patterns of such a service.[9]

Currently the bwFLA test and demo infrastructure uses older, written off hardware, using 12 machines, each equipped with two physical Intel Xeon CPUs (E5440) featuring four cores each running at 2.83 GHz. All instances are booted diskless (network boot) with the latest bwFLA codebase deployed. Additionally, there is an EaaS gateway running on four cores delegating request and providing a web container framework (JBoss) for the IFrame delivery. To ensure, a decent performance of individual emulation sessions, one emulation session got assigned to a physical CPU core. In total the test setup handled up to 96 parallel sessions.

The bwFLA cluster was evaluated under heavy load after the Rhizome announced access to a certain dynamic object in their collection. The publicity resulted in an overload of the system in a short period and pushed the average usage level to a higher platform. 700 sessions got evaluated, which resulted in an average session time of 15 minutes.[10] Under the assumption of baseline costs of 50 ct/hour for an 8 core machine at e.g. Amazon cloud[11] such a use case would boil down the session costs to about 2 ct/session. These are reasonable costs in such an application. These results can be used as a baseline for evaluation of migration-through-emulation scenarios, as it could be rather well predicted or measured how long a single run takes to complete. These considerations generate a fairly simple cost model for migrations.

7. CONCLUSION

The above example cost models for providing emulation solutions include reference to emulation products and services that do not currently exist or which are in different stages of development. The services, like bwFLA EaaS, still need further development to become really productive. Cost calculations and considerations for emulation strategies are only just beginning to become realistic as products and services are being made available and as memory institutions begin to consider implementing them. Preservation services can be supplied by one institution, or distributed across many. There are decreased marginal costs from sharing efforts and by sharing code-bases and developing open-source tool suites. Additionally, there are decreased marginal costs by cooperatively running a shared infrastructure.

The actual costs heavily depend on the scope of activities in ingest and access. Depending on the depth of analysis and quality assurance of the single object and expectations of future users the amount of manual labour going into it can become excessive and thus difficult to predict. The inherently long-term nature of digital preservation makes service-based cost models an attractive option as it allows for many of the costs to be passed on to those who benefit from them using a just-in-time approach rather than a just-in-case approach.

As discussed, very few of the shareable components are currently available as products or services from third parties (either for or non-profit). Furthermore, many of these shareable costs relate to activities that most organisations most-likely do not have either the money, nor the will to take on alone. These issues highlight a significant gap in the global digital preservation infrastructure that will need to be addressed if emulation based digital preservation strategies are to be successful over the long-term.

A substantial part of the cost-base of repositories consists of skilled staff and these human resources and many existing workflows and practices will not scale appropriately. There will be a need for more automation of processes and metadata generation, software tools for this, and potentially the development of greater collaboration and shared services to lower the entry and operational costs for institutions [5, 17].

8. ACKNOWLEDGMENTS

We would like to thank the Yale University Library for providing input and resources. This work was partially supported by the *bwFLA – Functional Long-Term Access* project[12] and the German Science Foundation (DFG).

9. REFERENCES

[1] Altman, M., Adams, M. O., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., and Young, C. H., 2009. Digital preservation through archival collaboration: The data preservation alliance for the social sciences. *American Archivist 72*, 1 (2009), 170–184.

[2] Ashley, K., 1999. Digital archive costs: facts and fallacies.

[3] Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., and Wheatley, P., 2008. The life2 final project report.

[4] Baker, M., Shah, M., Rosenthal, D. S. H., Roussopoulos, M., Maniatis, P., Giuli, T., and Bungale, P., Apr. 2006. A fresh look at the reliability of long-term digital storage. *SIGOPS Oper. Syst. Rev. 40*, 4 (Apr. 2006), 221–234.

[5] Beagrie, N., 2008. Digital curation for science, digital libraries, and individuals. *International Journal of*

[8] See `http://rhizome.org/`

[9] See `http://www.openplanetsfoundation.org/blogs/2014-07-09-eaas-action-%E2%80%94-and-short-meltdown-due-friendly-ddos`

[10] This was higher than expected, due to some long running sessions, as most probably the user switched the browser tab and never closed the original EaaS session.

[11] Pricing: `http://aws.amazon.com/ec2/pricing`

[12] bwFLA – Functional Long-Term Access, `http://bw-fla.uni-freiburg.de`.

Digital Curation 1, 1 (2008), 3–16.

[6] Beagrie, N., Chruszcz, J., and Lavoie, B. F. *Keeping research data safe*. HEFCE, 2008.

[7] Beagrie, N., Lavoie, B. F., and Woollard, M. *Keeping research data safe 2*. HEFCE, 2010.

[8] Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., and Hofman, H., 2009. Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries 10*, 4 (2009), 133–157.

[9] Blue Ribbon Task Force, 2010. Sustainable economics for a digital planet: Ensuring long-term access to digital information. *Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access* (2010).

[10] Chapman, S., 2006. Counting the costs of digital preservation: is repository storage affordable? *Journal of digital information 4*, 2 (2006).

[11] Cochrane, E., von Suchodoletz, D., and Crouch, M., 2013. Database preservation using emulation – a case study. *Archifacts*, April (2013), 80–95.

[12] Currall, J., McKinney, P., and Johnson, C., 2006. Digital preservation as an albatross. In *Archiving Conference* (2006), vol. 2006, Society for Imaging Science and Technology, pp. 75–78.

[13] Hedstrom, M., 1997. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities 31*, 3 (1997), 189–202.

[14] Liebetraut, T., Rechert, K., Valizada, I., Meier, K., and von Suchodoloetz, D., 2014. Emulation-as-a-Service – The Past in the Cloud. In *7th IEEE International Conference on Cloud Computing (IEEE CLOUD)* (2014), p. to appear.

[15] Loftus, M. J., 2010. The author's desktop. *Emory Magazine 85*, 4 (2010), 22–27.

[16] Lohman, B., Kiers, B., Michel, D., and van der Hoeven, J., 2011. Emulation as a business solution: The emulation framework. In *8th International Conference on Preservation of Digital Objects (iPRES2011)* (2011), National Library Board Singapore and Nanyang Technology University, pp. 425–428.

[17] Palaiologk, A. S., Economides, A. A., Tjalsma, H. D., and Sesink, L. B., 2012. An activity-based costing model for long-term preservation and dissemination of digital research data: the case of dans. *International Journal on Digital Libraries 12*, 4 (2012), 195–214.

[18] Rechert, K., Valizada, I., von Suchodoletz, D., and Latocha, J., 2012. bwFLA – A Functional Approach to Digital Preservation. *PIK – Praxis der Informationsverarbeitung und Kommunikation 35*, 4 (2012), 259–267.

[19] Rechert, K., von Suchodoletz, D., Valizada, I., Latocha, J., Cardenas, T. J., and Kulzhabayev, A., 2014. Take Care of Your Belongings Today - Securing Accessibility to Complex Electronic Business Processes. *Electronic Markets - The International Journal on Networked Business 24*, 2 (2014), 125 – 134.

[20] Rothenberg, J., 2000. Preserving authentic digital information. *Authenticity in a digital environment* (2000), 51–68.

[21] von Suchodoletz, D., Rechert, K., and Valizada, I., 2013. Towards emulation-as-a-service – cloud services for versatile digital object access. *International Journal of Digital Curation 8* (2013), 131–142.

[22] Waters, D., and Garrett, J. *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information.* ERIC, 1996.

[23] Wright, R., Miller, A., and Addis, M., 2009. The significance of storage in the "cost of risk" of digital preservation. *International Journal of Digital Curation 4*, 3 (2009), 104–122.