TEEM'16

FOURTH INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ECOSYSTEMS FOR ENHANCING MULTICULTURALITY

SALAMANCA 2-4 NOVEMBER 2016

VNiVERSiDAD D SALAMANCA
CAMPUS OF INTERNATIONAL EXCELLENCE

iuce

GRIAL Research Group in InterAction and eLearning

ICPS Published by ACM

Proceedings
TEEM'16

Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality

Salamanca, Spain
November 2nd – 4th, 2016

Editor:

Francisco José García-Peñalvo
University of Salamanca

TEEM'16 is organized by the Research GRoup in InterAction and eLearning (GRIAL) and Research Institute for Educational Sciences (IUCE) at the University of Salamanca.

Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality
(TEEM'16)
GRIAL Research Group
Research Institute for Educational Sciences (IUCE)
Paseo de Canalejas 169
37008 Salamanca, Spain

The Association for Computing Machinery (ACM)
2 Penn Plaza, Suite 701
New York New York 10121-0701

ISBN: 978-1-4503-4747-1

Designed by:

Felicidad García Sánchez
Research GRoup in InterAction and eLearning (GRIAL)

Digital Curation and Costs: Approaches and Perceptions

Luís Corujo
Universidade de Lisboa,
Faculdade de Letras
Alameda da Universidade
1600-214 Lisboa, Portugal
+351 217 920 000
luiscorujo@campus.ul.pt

Carlos Guardado da Silva
Universidade de Lisboa,
Faculdade de Letras
Alameda da Universidade
1600-214 Lisboa, Portugal
+351 217 920 000
carlosguardado@campus.ul.pt

Jorge Revez
Universidade de Lisboa,
Faculdade de Letras
Alameda da Universidade
1600-214 Lisboa, Portugal
+351 217 920 000
jrevez@campus.ul.pt

ABSTRACT
The production of large volumes of scientific information, considering its cost, requires approaches that ensure its maintenance, reuse and recovery. These concerns prompted the emergence of digital curation. We intend to discuss the relevant thinking concerning the costs of digital curation. This means addressing the definition of the concept and the issue of costs, based on the studies related to cost models. A literature review was conducted using B-On and RCAAP as research sources, exploring the perceptions of the authors regarding the digital curation and its costs. The views expressed were organized around a scheme based on the Digital Curation Centre (DCC) lifecycle and the reference model Open Archival Information System (OAIS). It is proposed a systematization of digital curation issues bridging the DCC life cycle view of the digital object curation to the OAIS reference model approach, using a cross view seized by cost models and plan/data management policies.

CCS Concepts
• Information systems~Digital libraries and archive

Keywords
Digital Curation; Cost; Cost Model; Data lifecycle.

1. INTRODUCTION
The Open Science movement strengthened the perception of the importance of data, as it is the evidence of scientific knowledge and the basis for its development [36]. Scientific data allow us to solve problems in ecology and climate, health, national security or nanotechnology [39], which confirms it to be a defining intellectual asset that can be not only peer-reviewed, but also subject to quality assessment and reuse [26]. Despite the fact that scientists argue that published data belongs to the scientific community, a large number of publishers claim their rights on those data, disallowing their reuse without their permission [37]. Within the framework of an Open Science, it is crucial to ensure that these data are made available by means of a free and open access, without any impairment, in order to benefit society. A larger amount of open data leads to a higher level of transparency and reproducibility, which results in a greater efficiency in the scientific process [36]. This leads to an increase in an already data intensive scientific production, as it manipulates a large amount of data from every scientific domain using new methodologies for data processing (Big data). Additionally, scientific endeavour is carried out by large-scale cooperation projects, sometimes multidisciplinary, with tremendous budgets supported by consortia of public, private, and even international institutions, denominated by Weinberg [54] as Big Science.

This Big data represents not only a greater number of recordings, but also a diversity in their nature: besides written reports, we witness the emergence of digital non-textual datasets, i.e., raw data that needs to be polished into increasingly more refined data [9]. This investment in resources and time is endangered by the volume, complexity, dynamic, provenance, storage, and by the fact that custody is shared amongst more than one producer. These vulnerabilities of digital data urge us to undertake approaches that ensure their maintenance, reuse and valorisation, particularly due to the high costs that impair their reproducibility. Digital curation1 champions an answer to these problems.

For many authors, digital preservation is an aspect of digital curation, although it is not our intention to explain its evolution or to distinguish it from digital preservation. To the Digital Curation Centre (DCC), it includes the maintenance, preservation and valorisation of scientific data throughout their useful life, which has been theorized in a life-cycle that considers every stage, starting with data conception and ending with reuse/transformation. Abbot [1] argues that it will imply the

[1] The term *curation* appears in Roman law embodied in the *curator*, a person responsible for people and the protector of heritage during the enforcement of liabilities. Later, the catholic "cura" appears as the keeper of a parish. This designation was salvaged by social communication, scientific research and arts [42]. Lee & Tibbo [32] point out the use of the term "data curation" in the 1980s-90s regarding the management of scientific data, while in 2011, a seminar on Archives, Libraries and Digital Science displays the term "digital curation". However, they locate the origin of this field in a report authored by Waters & Garret [53]. The "Digital Data Curation Task Force Report", published in 2002, debating the curation of scientific data, preceded the creation of the Digital Curation Centre (DCC) that strived to disseminate this discipline throughout projects, tools, training and support units. In 2006, Beagrie explained the evolution of this term in the opening article of the *International Journal of Data Curation*.

TEEM'16, November 02 - 04, 2016, Salamanca, Spain

ACM 978-1-4503-4747-1/16/11...$15.00
DOI: http://dx.doi.org/10.1145/3012430.3012529

management of large amounts of data that need to be reused, while ensuring its recovery and interpretability, by a wide range of professionals interacting with its life-cycle: from the content producers to the investors, policy-makers and repository managers. It is, therefore, an expensive effort that demands significant investments in time and skills, which can be a problem for smaller institutions since most advantages in digital curation can only be observed in the long-term and those investments may need time to pay off. Higgins [27] argues that it stems from technical evolution, from the understanding and maturation of the organizational activity, and from the workflow that highlights access and reuse of contents during their life-cycle, knowing that long-term management of digital contents gradually shifted from a passive preservation to an active curation. We consider digital curation a part of an overarching view of data life-cycle that depends on creation's context and implying different agents and interdisciplinary activities. It calls for planning and management and business policies, including sustainability planning and risk management, which implies transparency and identification of investment partners, costs/investments, tangible and intangible benefits, and mitigation strategies. This approach becomes possible only when digital curation is carried out by an institution endorsed with resources and competences.

Like Big Science, curation requires financial availability and a long-term budget planning in order to support the necessary expenditures with the preservation of digital assets, a calculation that is one of the weakest links of their management. Expenditures vary according to technological and socio-political changes that increase the difficulty of optimizing the available budget [23]. Its early phases cause a lack of consensus and engagement in contexts other than its practice community. The first attempts date back to the 1990s and, although explicit costs were not calculated, an effort was observed in assessing and defining the costs of several variables throughout the life-cycle stages. These studies preceded what later in the 2000s would become the reference model Open Archival Information System (OAIS) [10] and the DCC life-cycle, which in turn were the base for the costs models that have been developed and thoroughly discussed by Mundet & Carrera [12] Ferreira et al. [24] and Kesjser et al [29]. More recently, a Portuguese model for a shared structure of digital continuity was developed [44].

Scientific research on the several aspects and approaches regarding Big Data and Big Science have benefitted from financial support that have helped to control the experimentation costs. However, supporting agencies have begun to urge research bodies to preserve and grant long-term access to those contents, often formed by large amounts of data [40]. Therefore, it is imperative that institutions foresee and include in their research budget the needs of digital curation [52]. Although we assume that digital preservation is fairly expensive [44], Mundet & Carrera [12] argue that we are still far from pinpointing its exact costs. This becomes a problem when trying to obtain financing and convincing decision-makers and managers that are used to track and control costs. Despite a growing interest, they consider that the costs predictive models are still underdeveloped and premature because they don't reflect reality, which leads us to believe that we are still living in the infancy of digital curation. The authors also perceived the existence of a general awareness that costs can't be detached from the environment where contents are produced, since they depend on third-party interests, organizational structure and cultural context. For this reason, our goal is to understand the meaning of digital curation and the approaches regarding its costs as it triggers different reactions amongst producers, users and managers of digital data, especially researchers. This task implies to define digital curation and the costs topic based on studies concerning costs models.

2. RESEARCH METHODOLOGY

On the 5th January 2016, a research at the Online Library of Knowledge (B-ON) and at the Portuguese Repository of Open Access (RCAAP) was carried out in order to collect bibliographic references containing the terms "Digital curation" or "Curadoria Digital" related to words with "cost" or "custo" in their root. The query used the Boolean operators OR and AND paired with a truncation symbol at the end. In the RCAAP research, parentheses were not used to separate the sets of terms because this connector does not assume research conditions in which they are contained. The results showed 11 and 36 references, respectively, and a duplication of results contained in two different repositories was noted in the latter. From these results, 9 and 30 studies were used, respectively.

3. DIGITAL CURATION AND ITS COSTS ACCORDING TO STUDIES

In one of the studies extracted from RCAAP and B-ON, Constantopoulos et al. [11] formulates digital curation as a new interdisciplinary practice across various fields that seeks to define guidelines for the coherent management of information and to ensure its future use as the respective contexts evolve, in articulation with life-cycle models. It includes different participants, such as researchers, digital curators and documentalists, information managers, scholars, educators, exhibitions' curators, and the overall public. Dallas & Doom [14], Donnelly et al. [18], Sayão & Sales [49] and Santos [47] consider digital curation to be an evolving concept that comprises the active management and preservation of scientific data throughout their life-cycle, in articulation with the need to ensure the proper epistemic equivalent and the future reuse by interested communities. Consequently, it stands as an umbrella term that encompasses production, organization, storage, selection, valorisation, treatment and long-term preservation of digital assets, diminishing the risk of obsolescence. We are facing a new area characterized by interdisciplinary research and practices and involving a wide range of professionals. Ayris [3] and Whyte & Pryor [55] locate digital curation and preservation within the access, share and reuse of digital data, as promoters of data policies. Bernardou et al. [7] argue that specialists in humanities are the curators of academic data par excellence since they perform and value information research activities, as well as research concerning the curation of information objects and academic production. Ferreira et al. [25] and Bachell & Barr [4] mention actions concerning the production of context meta-information as a means to guarantee understandability, track record and reusability of data in the long-term. Queiroz [41] considers it an activity that aims to obtain best practices regarding the processing of digital objects, depending on production and technological contexts affected by human, socioeconomic, financial variables, amongst others. Bicarregui et al. [8] approach to curation underlies management's planning and data preservation for the Big Science for on-going use, and Wilson & Jeffreys [56] argue that specialized repositories possess an acknowledged experience in the curation skills of their human resources, while Dillon [17] and Strasser et al. [50] highlight the need for researchers to receive training in data management and digital curation in order to acquire good practices. From an applied linguistics and journalism perspective, Barros [6] states that digital curation is related to contents' selection, organization, understandability and evolution, representing a new practice that media professionals and volunteering users have to partake in, and that can only be understood when considering its cultural framework. Subsequently, the curator is essential to select content using gatekeeping and gatewatching, which relate to the sources of social communication. Faria & Ferreira [23] associate digital curation with the development of repositories to ensure preservation and present and future access, including selection, appraisal, preservation and access activities, data management and storage, and common and administrative services. Edmond & Garnet [21] mention that this practice requires different tools and strategies in order to guarantee data accessibility. Saraiva & Quaresma [48], Machado et al. [34] and Poole [39] explore curation from the perspective of the life-cycle, principles and actions necessary to ensure the sustainability and validity of scientific data so that in the future they can be accessed, shared, reused, aggregated or transformed. This requires human infrastructures, such as cyber-infrastructures, research communities, cooperation, planning, policies, standards and good practices. Therefore, human and social factors must be considered within the institutional contexts as they can support communication, coordination and cooperation regarding financing, revision, long-term preservation, scientific practice, resources and infrastructure. The authors also raise compelling questions: sustainability, cost, policy and planning that result from coordination and cooperation, theoretical and practical training, researchers' daily practices when interacting with data, and awareness to this issue.

3.1 Organizing results

During the attempts to systematize the concerns expressed in the bibliography about digital curation, it was decided to adopt a proposal stemming from the view of the DCC life-cycle and the OAIS reference model [10]. The reason is due to the fact that, while the former discusses the issues from a perspective of the digital curation's object, the latter is more specifically concerned with the system that supports the digital curation of those objects. Additionally, other elements common to both models are present, such as the Costs Models, the Model's Usability, the Model to compute cost/benefit, the Value/profitability, the Detailed Financial information of the Model/transparency and Costs. These elements are related to aspects such as Infrastructure and Sustainability that should be considered in the Management of Scientific Data Plan/Policy, without disregarding Legal Issues that influence a Decision-making procedure governed by a Preservation Management and Access Policy. This proposal is unprecedented as it introduces a perspective that is not only limited to digital objects and their life cycle, neither to systems with curation missions. Instead, it assumes the need to merge both standpoints in a cross-view of the Costs Models and the Data Management Plans/Policies (Fig. 1).

In terms of:

a) **Data production**: Bicarregui et al. [8] and Santos [47] discuss **scientific research costs** in the Big Science – since it deals with a data volume that can hardly be depleted in published papers – by introducing management problems that qualitatively differ from other disciplines and that, as a result, must forcibly be large-scale projects familiar with costs estimates.
b) **Appraisal and Selection:** it needs to consider the materials' cost, type, condition, quantity, accessibility, singularity and possibilities of future use [39].

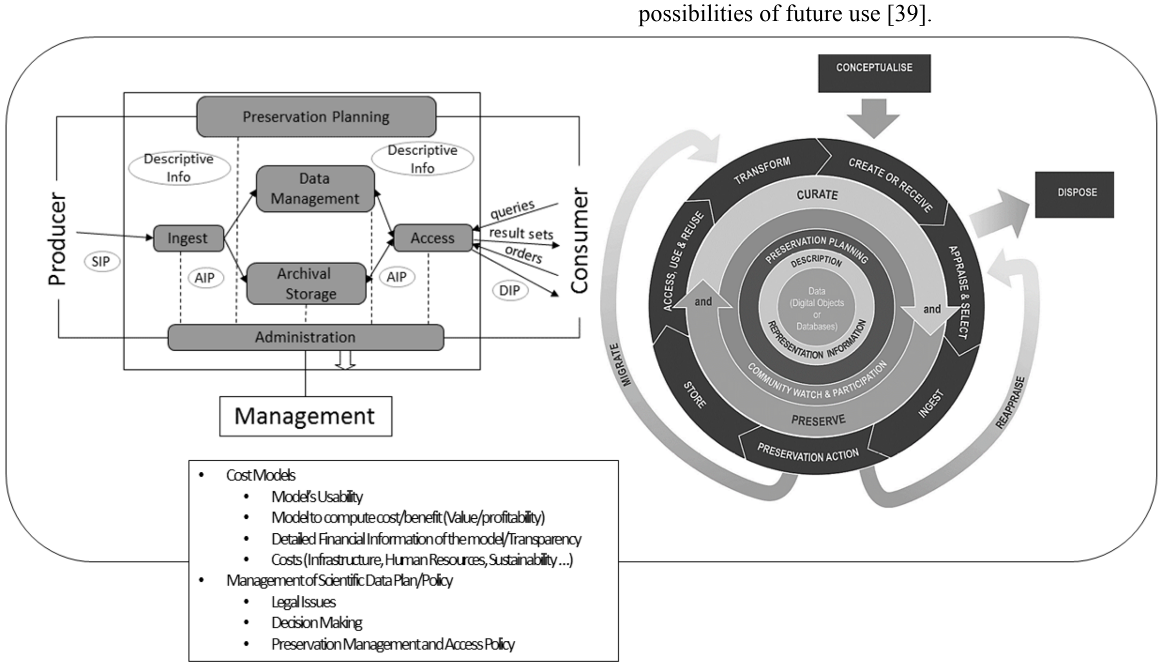

Figure 1. Proposal for the standardization of problems regarding the costs of digital curation. In other to better display the figure, the links between OAIS and the DCC Life-cycle. DCC are not shown. Producer <-> Create or Receive; Ingest <-> Ingest; Archival Storage <-> Store; Preservation Planning <-> Preservation Planning; Access <-> Access, Use & Reuse.

c) **Ingestion:** they cannot be overlooked, and quality control plays an important role at this stage [20].
d) **Preservation:** Avris [3] mentions the need of a generic model to detect key-elements in preservation activities and factors that boosted expenditures, and to identify and reduce bursts in cost as well as the frequency of those interventions. Kejser et al. [30] argue that these represent continuous costs that depend on the range of services provided by an institution and that are hard to isolate from other costs associated with the life-cycle, such as production and access/dissemination. Since the fact that, as expressed by Bicarregui et al. [8], most curation costs stem from preservation, the proper management of this activity would erase a significant set of practical problems concerning the release of data. **Migration and emulation strategies** [28] are included in this stage.
e) **Storage:** Wright et al. [57] and Santos [47] alert for the fact that the constant drop in the price of storage devices promotes the use of a larger capacity, although there are growing costs with energy, area, cooling and management. Whyte & Pryor [55] include these expenditures in a repository's reuse costs. Subsequently, Rice et al. [43] indicate that storage should obey open standards, be scalable, and foresee the use of hardware from different sellers by ensuring access mechanisms of flexible data. Rosenthal & Vargas [45], Suchodoletz et al. [51] and Saraiva & Quaresma [48] compare local storage with **cloud computing** services for a shared preservation, however, while the former conclude that cloud storage is not competitive in the long-term, the latter believe that there is a reduction in storage costs and an enhancing of interoperability with other systems and services.
f) **Access, use and reuse:** Rushbridge & Ross [46], Queiroz [41] and Machado [33] discuss the increase in **journals' subscription fees** that stimulated the emergence of open access. This is related to the philosophy of open access to scientific data (Open science) about which Whyte & Pryor [55] and Saraiva & Quaresma [48] defend savings in acquisition, access and managements, as well as less barriers when users external to a scientific community attempt to participate and actively cooperate. Bicarregui et al. [8] assert the importance of open data towards preservation, although they also express the need to select which data should be preserved. Evans & Moore [22] and Poole [39] highlight the role that **data reuse** and sharing in different forms and formats from the original play in reducing production costs of new data, an especially important argument during a period of economic austerity. Sayão & Sales [49] reinforce the idea that data value is related to the reproducibility of the search that returned them in order to analyse the access' cost-benefit and the ability to reuse data. Dürr et al. [20] present a system conceived to ensure the reuse of primary data and to produce secondary data, in a similar manner to the appeals of Minor et al. [35] to guarantee the protection of critical data collections beyond the lifetime of projects and efforts that generate them. The use of secondary data has consequences: reduction in data harvesting and duplication costs; distribution of direct and indirect harvesting costs, and new unforeseen benefits when harvesting (ex. data mining). However, these also bear costs regarding storage and data preparation for curation [55]. Rice et al. [43] argue that it is the duty of research institutions to develop infrastructures that support data variety and their reuse, especially when considering that, according to Wilson & Jeffreys [56] it is the case of research data supported by a public budget. For this reason, Delasalle [16] points out that one of the challenges is to overcome skepticism amongst researchers. In Donnelly & North's [19] opinion, research councils and information service providers must aim their policies and strategies to the needs of the research communities and to the optimization of information use and exchange. Bicarregui et al. [8] mention the choice of a project's products as representing, for the scientist, a commitment comprising the amount of time available to invest in understanding those data, the amount of support he is given by his colleagues and the data owners, and the delicateness of the issue he aims to research. Machado [33] reports the lower costs of new research conducted upon **shared scientific data**, a value identified by Whyte and Pryor [55] as a saving in researchs' indirect costs, enabled by the reuse of congregated resources (avoiding harvesting exhaustion) and by an "exchange of available presents" as a way of accelerating derived scientific research, leading to an increase in efficiency. Poole's [39] opinion is that any model that asserts the economic efficiency and effectiveness of data sharing does not mirror the complexity of human relationships embedded in its production. Cooperation benefits include the access to a broad range of experiences, costs and sharing of resources, access to new tools, development of standards and good practices and, finally, an increase in awareness of the problem, having Akers & Green [2] argued that academic libraries must play a part in promoting and preserving scientific data. Whyte & Pryor [55] follow the OCDE Guidelines to defend the lowest **access cost** possible, which preferably should not to exceed the minor cost of dissemination, also stating that a reduction in access costs to previous scientific knowledge depends on public financing researchers are granted.
g) **Dissemination:** Bicarregui. et al [8] stand against a wide data dissemination because it is not free, which instead implies significant costs identified by Whyte & Pryor [55] as relating to data understandability (preparation and documentation using acknowledged meta-information schemes - revision costs) to conclude that sharing should increase the cost, instead of decreasing it. Dürr et al. [20] recommend research on displaying tools and on recovery of data sets, whereas Wilson & Jeffreys [56] advocate costs in disseminating data management services.
h) **Administration:** The positing of **monitoring**, as expressed by Kejser, et al. [30] as the vigilance of the target community (costs inversely vary according to the influence of the curating body in producing and using formats) and technology (costs depend on development and complexity). Therefore, Poole [39] states the need of a balance so that pro-active vigilance costs do not become disproportionate.
i) **DCC Life-cycle Model & OAIS Reference Model**: There are a large number of studies that comment these models by occasion of projects devoted to data preservation and management cost models. They state that their mere reading and implementation is a fair piece of advice in Data Management and Preservation, unless we consider validation, audit and costs modelling according to these specifications.
j) **Cost models**: Some studies report projects such as LIFE [5], ESPIDA, CMDP, and 4C. Currall et al. [13] argue that understanding and communicating the cost and value of curation activities is key to guarantee the long-term survival of digital assets, although there are some problems when

trying to clearly express their value to every interested party, particularly potential funders. Davies et al. [15] conclude by stating that costs can appear in various stages of the life-cycle, and be recurring and diverse in occurrence. Wright et al. [57] introduce an approach to risk that combines cost dimensions, uncertainty and benefit. It considers value and costs models to be merely part of a broader economic modelling, which in turn is a part of achieving digital preservation and sustainable access. Ayris [3] enumerates some recommendations on how to compute costs according to inflation and suggest that costs should be considered external to the life-cycle. Bicarregui et al. [8] mention that these projects are not consensual, which may even not be possible due to the variety of preservation contexts. As far as data that have not been accessed are concerned, there is even less support in costs modelling. Kilbride & Norris [31] state that costs modelling, in order not to become and end in itself, needs to include concepts close to "risk", "value", "quality" and sustainability". They also affirm that research on curation modelling costs has been quite active and tends to prize preservation costs due to the difficulty in harmonizing preservation benefits with the complexity of the task. Evans & Moore [22] provides an example of how a seemingly simple decision, such as a file's format, has serious consequences in preservation and reuse. They conclude saying that, despite the fact that costs models claim to be generic, they tend to be specific to certain institutions while, at the same time, there are bodies that protect confidential information about costs. Other identified flaws were poor usability and a lack of consensus on how to structure costs data. Faria & Ferreira [23] believe that this field has greatly evolved due to a better understanding of costs and that these models ensure that strategic decision-making processes are more efficient and accurate, despite the absence of a model that can be executed in every instance.

k) **Model's Usability**: Kejser et al. [30] argue that costs models are not precise enough to estimate future costs due to challenges created by handling the predicting element, which conditions several expressions of the model. Similarly, Kilbride & Norris [31] indicate the rare assimilation of tools and methods developed with these models.

l) **Model to compute cost/benefit**: Currall et al. [13] point the need of possessing better information on costs and benefits so to stimulate investment managers to make solid-decisions. There are several questions that are raised by this: what benefits can it bring for external clients of the organization? Will the institutions' procedures improve? Will it help developing business and expand knowledge? What's the impact on finances? And even what if the penalty fine's amount is inferior to the cost of required actions? Bicarregui et al. [8] state that costs are fairly precise and can be materialized, notwithstanding the difficulty in estimates beyond a certain magnitude. On the other hand, valorisation is often imprecise as it includes educational benefits and assistance, which are real, although only measurable by a formal cost-benefit analysis. Kilbride & Norris [31] consider that it is necessary to provide clear cost-benefit models, including their conceptual description and standardised vocabulary.

m) **Value/profitability**: Currall et al. [13] describe it as the income/revenue creation by selling assets, licencing goods and/or rights, teaching and research, contracts, grants, taxes, donations, reduction of costs in work, time, space and direct expenditures. Evans & Moore [22] conclude that the assessment of data's value and impact has become increasingly important in a climate of recession and decrease in economic activity.

n) **Detailed Financial Information of the model/transparency:** its importance is asserted by Kilbride & Norris [31] when detailing if current cost-benefit models meet the needs of users interested in calculating and comparing financial information. Ayris [3] explains how fixed-term financing models (research scholarships or contracts) are unsuited to satisfy the needs of long-term access and preservation.

o) **Costs:** Studies identify fixed and variable direct costs and indirect costs. Wilson & Jeffreys [56] note that institutional services often demonstrate high fixed costs when comparing them with variable costs (the urge for constant development and human resources), while Whyte & Pryor [56] argue that it is necessary to efficiently use short resources for data collection - including research topics and instrumentation - in order to reduce indirect costs. Consequently, Kejser et al. [30] express the idea that models should represent complete economical costs, whether direct or indirect. Within these costs categories, studies discuss description costs regarding meta-information, growing costs with processing and costs derived from acquisitions and human resources.

p) **Infrastructure:** Referred in most studies, it includes platforms and informatics/scalability components. Wright et al. [57] warn that institutional "Total Cost of Ownership" models are often wrongly considered an annual value. It is crystal clear that these costs are heavy for an institution, and we also need to remember the service's volume to ensure it successfully satisfies the needs of its users. Suchodoletz et al. [51] propose an alternative use of a public cloud so that institutions may avoid the maintenance of expensive and underused servers.

q) **Sustainability:** Ayris [3] identifies a set of key-elements to obtain a long-term sustainability of digital collections and Poole [39] considers that institutions tend to dedicate a residual part of their budgets to curation and financing, which does not increase alongside the growth in data. For these reasons, plans should reveal business leadership, precise value proposals, costs minimization, exploration of diverse sources of income, and measurable engagements for accountability. Regarding **financing plans**, Ayris [3] observes a lack of incentives to implement sustainable economic models and, as a result, budgeting for digital curation is still dependent on short-term projects. Additionally, despite the funder's attempt to identify sustainable resources, they often neglect the creation of a clause to measure whether or not the agreed results are accomplished or measured [39]. In one of the studies about **management and business models**, Wilson & Jeffreys [56] conclude that business models have a better production at the level of individual components, rather than at the level of the whole infrastructure. Poole [39] argues that the development of business models, as well as cooperation through **preservation networks** are essential concerns when computing costs, and are deeply related to the issue of costs sharing, as exemplified by Suchodoletz et al [51]. Wilson & Jeffreys [56] raise a rather unexplored question about the utility of assessing the potential the public has to absorb a service, according to the capacity in planning and development of the business model. Wright et al. [57] discuss the modelling of **risk costs** related to uncertainty, unpredictability and threat of occurrence, which Ayris [3] argues can be solved by contingency plans and budgets, strategic reserves or insurances to alleviate losses. The clarification of risks allows the development of an approach that matches risk with costs and benefits [39].

r) **Management of Scientific Data Plan/Policy**: Ayris [3] advocates a careful choice when deciding the reach of preservation activities: they need to be wide enough to be able to successfully solve challenges, but not wide enough to a point where its failure would become catastrophic. He considers that there isn't a replacement to a flexible organization fully committed to preserve a material corpus. Such organization should be able to make decisions and delineate strategies to deal with every sort of unexpected problems and, if necessary, to filter problems and establish commitments. Donnelly & North [19] note that information services providers need to come closer to research communities because the existence of a single approach for the future of sciences or an unspecific data policy will neither be effective nor efficient. Wilson & Jeffreys [56] claim that researchers aren't aware of the real costs of data management solutions (which rarely are taken full advantage of), in opposition to the cases where services are centralized and costs become visible and demoralizing. For this reason, Delasalle [16] argues that data management plans enable not only to meet the expectations of researchers and their research field, but also to compute costs deriving from the project's plan. Rice et al. [43] believe that a free service provided when using data eliminates one of the main obstacles to a proper curation, while also allowing data to be kept in a suitable infrastructure, instead of any other selected due to its lower cost.

s) **Legal issues**: The lack of international consistency [5] motivates research on this topic but we can say that, at a political level, its demands should be balanced between the interested parties [39]. Here we can include intellectual property rights concerning videogames [4] and journalistic content [6].

t) **Decision-making**: Whyte & Pryor [55] identify decisions to be made by policy-makers, researchers and users' communities concerning risks, costs and benefits of a broader participation, to then proceed to the results they produced, while Delasalle [16] believes that researchers should be included in the decision-making process to provide insight on what is stored, where, and for how long.

u) **Preservation Management and Access Policy**: Studies elaborate on ideas closely related to certification, which, in Bicarregui et al. [8] opinion, should be supported by a set of methodological tools, including overviews, case studies and cost models in order to provide guidelines on how to acquire best practices in Data Management and Preservation and a solid infrastructure for those projects.

4. CONCLUSION

This study introduces an unprecedented proposal for the systematisation of problems arising from digital curation costs. Our vision is not limited to digital objects and their life-cycle, nor to systems with curation duties, since it is necessary to overlap both dimensions, resulting in an overarching perspective of the Costs Models and Data Management Plans/Policies. An analysis on author's perceptions based on this scheme allowed us to realize that the concept of digital curation has been used from a life-cycle's point of view, yet highlighting digital preservation problems that are hard to isolate from other elements of the cycle. Such approach has effects on the modelling of costs, despite the absence of a well-structured functional model and an agreement on acceptable accountable principles. Costs were always discussed but it is considered that a paradigm shift is occurring from a black-box perspective to another that identifies costs and attempts to standardise predictive models for institutional use with the aim of promoting transparency and accountability and capturing the interest of potential founders. It is interesting to note that this clarification impulse has grown with the advent of the global financial crisis. Studies reveal an overall awareness that costs cannot be isolated from their context, therefore belonging to a broader framework that includes the interests of third-parties, the organizational structure and the cultural environment. However, this approach still requires refining as models possess numerous flaws. It can also be observed that, in Brazil, international research on digital curation is closely monitored, especially under the form of academic thesis, while in Portugal it consists more on the active participation in international research teams and on the development of costs modelling in this field of study.

5. REFERENCES

[1] Abbott, D. 2008. DCC Briefing Paper: What is digital curation? (2008).

[2] Akers, K. and Green, J. 2014. Towards a Symbiotic Relationship Between Academic Libraries and Disciplinary Data Repositories: A Dryad and University of Michigan Case Study. *International Journal of Digital Curation*. 9, 1 (2014), 119–131.

[3] Ayris, P. 2009. LIBER's Involvement in Supporting Digital Preservation in Member Libraries. *Liber Quarterly: The Journal of European Research Libraries*. 19, 1 (2009), 22–43.

[4] Bachell, A. and Barr, M. 2014. Video Game Preservation in the UK: A Survey of Records Management Practices. *International Journal of Digital Curation Volume*. 9, 2 (2014), 139–170.

[5] Ball, A., Day, M. and Patel, M. 2008. The Fifth International Conference on Preservation of Digital Objects (iPRES 2008). *International Journal of Digital Curation*. 3, 2 (2008), 89–102.

[6] Barros, N. 2014. Apropriação da curadoria na web por uma empresa de mídia tradicional: um caso de convergência entre narrativa e banco de dados. Unicamp.

[7] Benardou, A., Constantopoulos, P., Dallas, C. and Gavrilis, D. 2010. Understanding the Information Requirements of Arts and Humanities Scholarship. *International Journal of Digital Curation*. 5, 1 (2010), 18–33.

[8] Bicarregui, J., Gray, N., Henderson, R., Jones, R., Lambert, S. and Matthews, B. 2013. Data Management and Preservation Planning for Big Science. *International Journal of Digital Curation*. 8, 1 (2013), 29–41.

[9] Buckland, M. 2011. Data Management as Bibliography. *Bulletin of the American Society for Information Science & Technology*. 37, 6 (Sep. 2011), 34–37.

[10] CCSDS 2002. Reference Model for an Open Archival Information System (OAIS). CCSDS.

[11] Constantopoulos, P., Dallas, C., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., Kotidis, Y. and Papatheodorou, C. 2009. DCC&U: An Extended Digital Curation Lifecycle Model. *International Journal of Digital Curation*. 4, 1 (2009), 34–45.

[12] Cruz Mundet, J.R. and Díez Carrera, C. 2015. El cálculo de costes de la preservación digital: un análisis de modelos. *Anales de Documentación*. 18, 2 (Oct. 2015).

[13] Currall, J., Johnson, C. and McKinney, P. 2007. The world is all grown digital.... How shall a man persuade management what to do in such times? *International Journal of Digital Curation*. 2, 1 (2007), 12–28.

[14] Dallas, C. and Doorn, P. 2009. Report on the Workshop on Digital Curation in the Human Sciences at ECDL 2009. *D-Lib Magazine*. 15, 11/12 (2009).

[15] Davies, R., Ayris, P., McLeod, R., Shenton, H. and Wheatly, P. How much does it cost? The LIFE Project - Costing Models for Digital Curation and Preservation. *Liber Quarterly: The Journal of European Research Libraries*. 17, 3/4.

[16] Delasalle, J. 2013. Research Data Management at the University of Warwick: recent steps towards a joined-up approach at a UK university. *Libreas Library Ideas*. 9, 2 (2013), 97–105.

[17] Dillon, C. 2013. The Research Library as Digital Curator at Virginia Tech. *College Undergraduate Libraries*. 20, 2 (2013), 232–238.

[18] Donnelly, M., Jones, S. and Pattenden-Fail, J. 2010. DMP Online: The Digital Curation Centre's Web-based Tool for Creating, Maintaining and Exporting Data Management Plans. *International Journal of Digital Curation*. 5, 1 (2010), 187–193.

[19] Donnelly, M. and North, R. 2011. The Milieu and the MESSAGE: Talking to Researchers about Data Curation Issues in a Large and Diverse e-Science Project. *International Journal of Digital Curation*. 6, 1 (2011), 32–44.

[20] Dürr, R., Meer, K., Luxemburg, W. and Dekker, R. 2008. Dataset Preservation for the Long Term: Results of the DareLux Project. *International Journal of Digital Curation*. 3, 1 (2008), 29–43.

[21] Edmond, J. and Garnet, V. 2015. APIs and Researchers: The Emperor's New Clothes? *International Journal of Digital Curation*. 10, 1 (2015), 287–297.

[22] Evans, T. and Moore, R. 2014. The Use of PDF/A in Digital Archives: Study from Archaeology. *International Journal of Digital Curation*. 9, 2 (2014), 123–138.

[23] Faria, L. and Ferreira, M. 2015. Plataforma de colaboração para custear a curadoria digital. Évora. (Évora, 2015).

[24] Ferreira, M., Faria, L. and Silva, H. 2014. D2. 1: baseline study of stakeholder & stakeholder initiatives. 4C Project.

[25] Ferreira, M., Saraiva, R. and Rodrigues, E. 2012. *Estado da arte em preservação digital*. Universidade do Minho.

[26] Heidorn, P.B. 2011. The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*. 51, 7/8 (outubro 2011), 662–672.

[27] Higgins, S. 2011. Digital Curation: The Emergence of a New Discipline. *The International Journal of Digital Curation*. 6, 2 (2011), 78–88.

[28] Hoeven, J., Lohman, B. and Verdegem, R. 2007. Emulation for Digital Preservation in Practice: The Results. *International Journal of Digital Curation*. 2, 2 (2007), 123–132.

[29] Kejser, U., Johansen, K., Thirifays, A., Nielsen, A., Wang, D., Strodl, S., Miksa, T., Davidson, J., McCann, P., Krupp, J. and Tjalsma, H. 2014. *D3.1 - Evaluation of Cost Models and Needs & Gaps Analysis*. 4C Project.

[30] Kejser, U., Nielsen, A. and Thirifays, A. 2011. Cost Model for Digital Preservation: Cost of Digital Migration. *International Journal of Digital Curation*. 6, 1 (2011), 255–267.

[31] Kilbride, W. and Norris, S. 2014. Collaborating to Clarify the Cost of Curation. *New Review of Information Networking*. 19, 1 (2014), 44–49.

[32] Lee, C.A. and Tibbo, H. 2011. Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills. *Archivaria*. 72, Fall 2011 (2011), 123–168.

[33] Machado, D. 2015. Dados de pesquisa em repositório institucional: o caso do Edinburgh DataShare. Universidade Federal do Rio Grande do Sul.

[34] Machado, H., Soares, P. and Silva, T. 2015. Em busca duma anamnese universitária: a materialização do arquivo do ISSSL na Internet. (Évora, 2015).

[35] Minor, D., Sutton, D., Kozbial, A., Westbrook, B., Burek, M. and Smorul, M. 2010. Chronopolis Digital Preservation Network. *International Journal of Digital Curation*. 5, 1 (2010), 119–133.

[36] Molloy, J.C. 2011. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology*. 9, 12 (dezembro 2011), 1–4.

[37] Murray-Rust, P. 2008. Open Data in Science. *Serials Review*. 34, 1 (Mar. 2008), 52–64.

[38] Ogburn, J.L. 2010. The Imperative for Data Curation. *portal: Libraries and the Academy*. 10, 2 (2010), 241–246.

[39] Poole, A. 2015. How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science*. 15, 1 (2015), 101–139.

[40] Poole, A.H. 2013. Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities. *Digital Humanities Quarterly*. 007, 2 (Oct. 2013).

[41] Queiroz, B. 2013. A preservação da informação na Universidade Federal de Goiás: uma proposta de curadoria digital. Universidade Federal de Goiás.

[42] Ramos, D. 2012. Anotações para a compreensão da atividade do "curador de informação digital. *Curadoria Digital e o Campo da Comunicação*. E. Correa, ed. ECA-USP. 11–21.

[43] Rice, R., Ekmekcioglu, Ç., Haywood, J., Jones, S., Lewis, S., Macdonald, S. and Weir, T. 2013. Implementing the Research Data Management Policy: University of Edinburgh Roadmap. *International Journal of Digital Curation*. 8, 2 (2013), 194–204.

[44] Rodrigues, A., Barbedo, F., Runa, L. and Sant'Ana, M. 2015. *Continuidade digital: relatório final do projecto*. DGLAB.

[45] Rosenthal, D. and Vargas, D. 2013. Distributed Digital Preservation in the Cloud. *International Journal of Digital Curation*. 8, 1 (2013), 107–119.

[46] Rusbridge, A. and Ross, S. 2007. The UK LOCKSS Pilot Programme: A Perspective from the LOCKSS Technical Support Service. *International Journal of Digital Curation*. 2, 2 (2007), 111–122.

[47] Santos, T. 2014. Curadoria Digital: o conceito no período de 2000 a 2013. Universidade de Brasília.

[48] Saraiva, P. and Quaresma, P. 2015. Bibliotecas Universitárias: tendências, modelos e competências. (Évora, 2015).

[49] Sayão, L. and Sales, L. 2012. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. *Informação & Sociedade: Estudos (I&S)*. 22, 3 (2012), 179–191.

[50] Strasser, C., Abrams, S. and Cruse, P. 2014. DMPTool 2: Expanding Functionality for Better Data Management Planning. *International Journal of Digital Curation*. 9, 1 (2014), 324–330.

[51] 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, 1 (2013), 131–142.

[52] Walters, T. and Skinner, K. 2011. *New roles for new times: Digital curation for preservation*. Association of Research Libraries.

[53] Waters, D. and Garrett, J. 1996. Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. CPA/RLG.

[54] Weinberg, A.M. 1961. Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. *Science (New York, N.Y.)*. 134, 3473 (Jul. 1961), 161–164.

[55] Whyte, A. and Pryor, G. 2011. Open Science in Practice: Researcher Perspectives and Participation. *International Journal of Digital Curation*. 6, 1 (2011), 199–213.

[56] Wilson, J. and Jeffreys, P. 2013. Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford. *International Journal of Digital Curation*. 8, 1 (2013), 235–246.

[57] 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.