Modelling Data Value in Digital Preservation

Giuseppa Caruso, Luigi Briguglio
Engineering Ingegneria Informatica S.p.A
via Riccardo Morandi 32
00182 Rome, Italy
{giuseppa.caruso, luigi.briguglio}@eng.it

Brian Matthews
Science and Technology Facilities Council
Rutherford Appleton Laboratory
Didcot, OX11 0QX, UK
brian.matthews@stfc.ac.uk

Calogera Tona,Mirko Albani
ESA-ESRIN
Via Galileo Galilei
00044 Frascati, Rome Italy
{Calogera.Tona@, Mirko.Albani}esa.int

ABSTRACT
For decades, the Earth Science (ES) community has launched missions to monitor vital phenomena of our planet and, through measurements, obtain data for improving their models. Indeed the proper characterisation of phenomena, such as desertification, artic sea ice melting, volcanic activities or earthquakes effects, requires the analysis of data acquired in a long period and the validation of correctness of scientific models. This means that digital data, especially in the ES domain, represents an important asset to be preserved over time. Despite each single ES mission's cost being quantified and supported by well documented evidence, ES organisations are not able to assess the value of data generated by those missions over time. This paper describes the rationale for and an approach to modelling the value of data/information to be preserved over long term in digital archive. This is the result of experience in the SCIDIP-ES project [16] which has considered the: i) definition of models for describing the value of digital data and related information; ii) characterisation of data/information value model through core set of key parameters and iii) identification of long term digital preservation activities that may potentially impact on key parameters and consequently on the value of digital assets. This model is being assessed in ES scenarios with data curators and archive managers.

Categories and Subject Descriptors
H.1.1 [**Information Systems**]: Systems and Information Theory – *value of information.*

General Terms
Management, Economics, Theory

Keywords
Value of Data/Information, Value Model, Sustainability, Long Term Data Preservation, Earth Science (ES).

1. INTRODUCTION
Climate change is arguably the greatest environmental challenge facing us in the twenty-first century, and this has been recognized in reports from the Intergovernmental Panel on Climate Change (IPCC) [19] and from the United Nations Framework Convention on Climate Change (UNFCCC) [20]. The consequences of a warming climate are far-reaching, potentially affecting fresh water resources, global food production and sea level. Threatening impacts on the natural environment and life on Earth for generations to come, climate change is high on political, strategic and economic agendas worldwide. This premise highlights the importance of ES studies and describes the Earth and its natural phenomena through data and models. For this purpose, ES community - which includes a wide range of scientists interested on fields related to the Earth such as physical geography, geology, meteorology, oceanography, atmospheric sciences, physics, and chemistry - acquires, processes and examines a large amount of dataset on Earth's materials, structure, history and all of the living things on it, including how and when they formed and evolved. This kind of study of the Earth helps to develop an understanding of its future and the need for careful management of its resources, and in particular, this can help to model and estimate climate change. For those reasons, for decades ES community launched missions such as Argo [21] and GRACE [22] which acquire data related to gravimetry and Mean Sea Level variations, very sensitive indexes of climate change and variability. It is also to be considered the large amount of new ES observations upcoming in the next years will lead to a major increase of ES data volumes, as well as ES datasets are characterised by heterogeneity due to different instruments and technologies mounted by each mission's satellite. It is important to highlight that validation and improvement of models cannot be successfully performed in case of "lack" or "hole" within the dataset sequence. In other words, every acquired data from the ES missions is an important asset for ES community and the whole humanity: that clarifies the importance of avoiding the lost of data related to Earth events uniquely occurred over time and space, as well as to plan and enact long term digital preservation on this asset for ensuring availability and accessibility.

An asset for an organization has, for definition, a value. While costs for generating data are widely known and documented, on the other hand, it is still an open issue for ES organizations to assess the value over time of this asset. This paper describes in Chapter 2 the existing models available from the state of art and their limits in satisfying specific needs of the ES community, especially when dealing with long-term preservation. In order to overcome those limitations, Chapter 3 introduces the experience carried out within the SCIDIP-ES project which provides an approach for adapting existing models and describes how those models have been extended. Closing remarks are reported in Chapter 4.

2. VALUE OF DATA/INFORMATION

The term, "Value" has multiple meanings, which change according to the different domains where this term is used. In this paper, the term "value" is for referring to the economic and market value of preserved information, which is seen as an asset. In economic studies, the theory of value attempts to explain the exchange "value" or "price" of goods and services.[ref.] According the Marketing approach, the "value" may be conceptualized as the relationship between the consumer's perceived benefits and the perceived costs for receiving these benefits[ref.]. From the point of view of the profit and no-profit organisations, the generation of value depends on the difference between benefits and costs derived from their activities[ref.].

The value approach followed in the SCIDIP-ES project and presented in this paper is the last one, considering that the *value of data and in particular of the preserved data is closely related to processes and activities, which are needed over time to offer the data/information to final users as well as to the activities performed on data/information by data users*. In this perspective, Benefit/Cost analysis is the starting point for the value analysis and how it changes during the whole digital object lifecycle. Thus, to achieve a better understanding of relevant current and past work on benefit/cost analysis and on the Value Analysis about information and in particular preserved information was an important step to identify existing value approaches, which could be followed by the SCIDIP-ES project.

2.1 State of the Art

The interest in digital preservation and its value is evident through the relevant related work. However, the most of the analysed research projects on digital preservation have been focused on the Cost Model and on, in particular, the estimation of their cost. Those analyses have been carried out in different domains, with a particular focus on culture heritage. It is characteristic that cost models for digital preservation take a ***lifecycle approach*** (LIFE [1], CMDP [2], KRDS [3] , ENSURE [4]). However, no common consensus has yet been reached on how the lifecycle for costing digital preservation should be structured; or on how the individual lifecycle phases should be broken down and detailed, perhaps due the high dependences of preservation costs on the range of services that an institution can offers. All the considered projects adopt the OAIS reference model [5] as starting point for the definition of digital preservation lifecycle and its breakdown but the final results of the latter are quite different among those projects, due to the different fields of application. Another unresolved or hidden issue is the development of formulas for operational cost models.

With regards the Value analysis, the studies [6][7][9] dealt with about the general value of the Information for society; they are not about the preserved data but more in general on the impact of it on the domain where it is used. However*, all are persuaded that* ***the value*** *of information* ***depend on its use and its capability to be shared****.* Keeping Research Data Safe[3] (KRDS) is the only study to consider the benefit analysis for data preservation, which also provides a Benefits Analysis Toolkit [8]. This latter has been tested, reviewed and developed further in the Keeping Research Data Safe (KRDS) Benefits Framework and the KRDS/I2S2 Value Chain and Benefit Impact Analysis tools for assessing the benefits of digital curation/preservation of research data. In conclusion, from this analysis of related work, a list of variables and parameters was defined. This paper does not include that list which is available in the project's document [11] , but it is relevant to highlight at least the typologies of variables/parameters identified. In fact, two main typologies of parameters were identified: those related to cost analysis and its definition (in this perspective it is possible to define the value of the preserved object as sum of the cost elements); the others one are general and high level parameters about digital object quality and features.

Finally, the main identified Economic Value model approaches relevant for value analysis of the Data/Information, were:

1. *Willingness Approach*: the Value of Information(VoI) measured according the willingness to pay of decision-makers (or others who use the data) where their willingness depends on the level of uncertainty and on what is at stake (amount of possible loss without information)
2. *Attribute Approach*: the value is a function of some parameters related the quality and features of the digital bject;

   VoI *= f (Usability, Shareability, Time, Accuracy, Precision, Risk, Unicity, Integrity)*
3. Historical Cost Approach: VoI as approximation of the cost of acquiring/creating/archiving/preserving it (purchase price or development cost);
4. *Present Value Approach*: information considered as an asset is valued based on the present value of expected future economic benefits.

The first two are market oriented, that means that they define the value according the value perception of who use the product or services according their features and quality, and the user's availability to pay or to do something in order to access to the asset. The last two approaches are process oriented, that means that the value of the provided product or service is defined according the process and cost for providing it as well as the produced benefits in terms of outcomes derived from an activity or work process.

2.2 Limits of current models

The state of the art analysis gave an overview of the available and more used approaches about the value analysis as well as provides for the SCIDIP-ES project an early idea about their advantages and limits according the project needed. In this perspective, it is possible to highlight the following aspects:

- Most of the value models analysed may not be applied to Preserved data, because they are mainly focused on cost analysis.
- Those models are not addressing the benefit provided by the data itself, that is considered an important aspect for the ES community and consequently for the SCIDIP-ES purpose.
- Moreover, current experiences are not considering the whole lifecycle of digital data which may impact on its value.

Starting from the models identified, it becomes important to adapt and extend them, for the specific purpose of the project.

In order to achieve this goal, the SCIDIP-ES team proposes to adapt and extend:

- The Historical Cost approach by adopting for the cost analysis ABC (Activity Based Costing) model and introducing a benefit framework for the benefit analysis;
- The Attribute Approach by introducing the SCIDIP-ES core set of preservation parameters, which allow the definition of the value of data/information and the impact on this value due to activities performed on data during the whole lifecycle.

3. VALUE OF PRESERVABLE DATA

The proposed model aims to bring together both to the process oriented approaches and to the market one. In this perspective, this section offers more details about the Cost/benefit framework as a process oriented approach as well as on the extension of the attribute approach as a market oriented approach.

3.1 Tailoring Benefit/Cost Analysis

Cost-benefit analysis takes into account the positive and negative aspects related to a case to be evaluated. Those aspects must be expressed in terms of a common unit of value, which conventionally is money. That represents a limit for measuring the benefits generated from the long term digital preservation activities in the scientific domain, since currently most of the data and information are freely available for users. Thus the benefit analysis proposed in this paper suggests measuring them following an approach based on the identification of the general impact on the community and society. With regards the cost analysis based on the Activity based Costing Model, the main effort, to tailor it, was to define an activities Framework for digital preservation relevant for scientific organisations.

3.1.1 Analysing Data Benefits

The following section will go deeper into the benefits of the data product. This approach starts from the analysis of the KRDS [3] benefits model, before passing to a more systematic model to be applied to data product relevant to scientific data.

The KRDS model of benefits [8] defines 3 dimensions: outcomes, timescales and beneficiaries as a framework to evaluate the benefit of a data product. Outcomes are then divided into:

- Direct benefits: positive impacts obtained in a data curation activity.
- Indirect benefits: negative impact avoided by investing in a data curation activity.

The guide to the benefits framework then goes on to discuss how this framework might apply in particular instances. This gives particular instances of outcomes which might apply; however, these are an unstructured list of potential outcomes.

In the SCIDIP-ES project a more systematic characterisation of the outcomes is proposed which could be applied to a data product within a research data scenario. This approach can then be combined with the rest of the KRDS approach to provide a more detailed analysis of the potential benefits accruing from the preservation of a data product.

This approach can also be compared with that of Whyte and Wilson [14] who identifies seven general criteria for retention (*Relevance to Mission; Scientific or Historical Value; Uniqueness; Potential for Redistribution; Non-Replicability; Economic Case; Full Documentation*). Again, while these are useful, they are not comprehensive, and do not in general capture the intentionality behind the criteria which may lead data archivist to identify additional benefits not covered within these definitions, or provide measurable criteria.

The nature of the benefits can be analysed by considering two main categories of benefits: *Utility* and *Substitutability*. These categories approximately correspond to KRDS's direct and indirect benefits.

Substitutability factors are those which assess whether an alternative data set of an acceptable quality which can be used in place of the data can be accessed if it is needed, if the archive's copy is not available. If a reasonable substitute can be accessed elsewhere, or generated afresh at a reasonable cost (for example at a lower cost than continuing to preserve the data), then the benefit of keeping a copy of the data within the archive is likely to be lower.

Utility factors consider the value of the data for re-examination and reuse in the future. Thus if the Utility of the data is high, then the benefit of the data is high. Considering data utility further, clearly the data is more valuable if the data is *desirable,* that is it requested, re-examined and reused in the future, especially in new contexts and new situations. Data may also have more beneficial impact if it is *reusable*, that is presented in a manner which encourages re-examination and reuse; if it is easier to comprehend and to integrate with other data and computing systems, it is likely to be reused, and thus have a higher utility. To this end, some instances of the types of evidence for the benefits of data in terms of both substitutability and utility have been identified, together with some guidelines on metrics which might be used by a data archive to measure such evidence. Those evidences and metrics bring together the concepts to estimate in terms of benefits, the gross value of the data. It is important, anyhow to highlight that often such metrics are subjective and difficult to measure, especially for a long time in the future. For brevity, we omit a comprehensive treatment here; Table 1 gives some examples of evidence of Data Desirability.

Table 1. Data Desirability Metrics

Evidence	Description	Metric
Data requests	Number of requests for the data arising from the user community.	Number of user requests. This can be also measured by a percentage of the funding which is supporting the user community (e.g. future research grants).
Data Citations	Citations of the data within refereed published literature.	Number of citations to data (or a reference paper for the data), weighted by the impact factors of the citing papers.
Research grants	Future research grants which cite or request access to the data. This is evidence that the data remains relevant in an active research area.	Percentage of the value of research grant.
Commercial data access	Sales of access to the data or added value products using the data.	Value of sales of the data or derived products.
Patents	Use of the data leads to commercial patents.	Number of patents arising (and an estimate of their

		value e.g. use in products).
Products	Use of the data leads to commercial patents.	Value of sales of products.
Influencing decisions makers	Use of the data by government or other agency to either: - influence policy (e.g. included in IPCC report) - directly influence action (e.g. monitoring of volcanic ash and flights)	Citation of data in policy documents. Estimate of value of policy or action.

3.1.2 Analysing Data Costs

Estimating the cost for long-term digital preservation has received attention from many organisations (e.g. companies, digital libraries, research data centres) who are interested in preserving for their data. In particular, in Earth Science domain, the importance of digital preservation is given by some data attributes as the non replicability of the acquisition process within the same conditions (i.e. satellite or airborne data). This interest is because a sound cost model should lead industries to better understand economic impact of digital preservation. Despite that, cost modelling for long-term digital preservation is a relatively new area of study. Many research projects analysed above (e.g. Life Cycle Information for E-Literature (LIFE)[1], Keeping Research Data Safe (KRDS)[3] and NASA's Cost Estimation Tool (CET) [15]), dealt with the cost model. Those existing studies are related to specific projects, institutions or materials and therefore difficult to transfer into other contexts. That is due to the particularity of the costs of preservation which are determined for specific digital assets using specific technologies, at a specified level of reliability and so on. From that perspective, it may be possible to follow the approach and high level model of others experiences, while tailoring them according the specific case requirements.

For the SCIDIP-ES project's needs for costs analysis it has been decided to follow an approach **based on Activities Based Costing (ABC) model**, which seems the most frequently used approach for the cost analysis. This is a costing methodology that identifies activities in an organization and assigns the cost of each activity with resources of all products and services according to the actual consumption by each activity. In that perspective, it is powerful tool both for cost assessment and for better understanding organisation processes. For such reason, this method is very useful to: i) identify and eliminate or modify production or service processes that are ineffective; ii) support an economic analysis of the adoption of new production or service processes. The first step in designing an ABC system is to conduct an activity analysis to identify the resource costs and activities of the organisation. The activity analysis identifies the work performed by the organisation to carry out its operations. Consequently, activity analysis includes gathering data from existing documents and records, as well as collecting additional data using questionnaires, observations, or interviews of key personnel. In our specific experience, we have identified the activities through two ways:

- In the first part of the analysis, the high level activities have been defined according the past experience of other projects which provided their cost models and some approaches for the breakdown of the activities for organisation committed in the digital preservation (e.g.: LIFE, KRDS, ENSURE);
- Then a re-adjustment and an identification of other lower level activities more related to Science domains has been carried out through internal discussion and analysing in particular the current digital preservation process inside ESA (European Space Agency).

For each high level activities group, two other levels of sub-activities were defined. The activities classes and groups are significant for economic assessment of the different parts of the overall system which brings a product or a service to a customer. This activities model represents the most important part in the ABC model application. The high level activities are conceptually based on the OAIS reference model [5] following the approach of many other projects (e.g. [1][2][4][17][18]) engaged in the cost analysis. The lower levels are more related to science organizations; they form a guide for users and can be contextualised to the structure and language of the organisation.

The Figure 1 shows the first and second level of the identified activities proposed for our cost analysis scope.

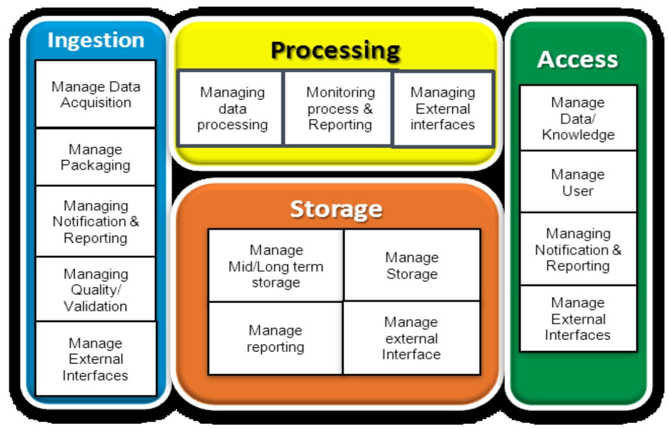

Figure 1. LTDP Activities 1st and 2nd levels –ABC approach

3.2 Extending Attribute Approach: Core Parameters

From the analysis of the parameters identified in the state of the art as well as from discussion within the SCIDIP-ES project, is come to light that some data's features are very relevant to explain the engendering of value of data itself. The selected features which are by us called "core parameters" are a core set of five parameters, which qualifies the value of data/information according the Data Users. They are defined as "factors that characterise the preserved digital object, which could impact on the utility perception of who needs and uses the digital object". Consequently they influence decisions on data use by data users impacting on the benefits generation.

On the other hand, providing over time digital data with the required degree of core parameters means to be aware on the organisational activities and resources (e.g.: technologies, know-how), which impacting on them as well as to be able to leverage on activities and resources for achieve the required levels of those parameters.

Those parameters are defined as follow:

1. ***Availability***

Availability is the property that a data is available for long-term use and at the time it needs to be utilised.
Data availability (sometime related to the concept of timeliness [12]) is one of the most frequent data quality dimensions that must be managed. According Vermaaten, Lavoie and Caplan [13], in order to ensure availability, the digital object must be ingested into, and subsequently maintained by, a preservation repository.

2. Accessibility

Accessibility is the ability to access data from some system and/or entity. Accessibility requires rights and/or permissions to access the data, technology (i.e. hardware and software) to access the data and the related documentation necessary to understand the data itself. In some case the data could be available but its access is not possible or not easy. This reduces its value for the interested community because becomes difficult to use it.

3. Integrity

Data Integrity is defined as the ability to ensure that data is not altered or destroyed in an unauthorized manner. This complies to the ISO:14721:2003 OAIS definition [5].
Usually we could say that enforcing data integrity ensures the quality of the data. Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle. The data integrity is very important in particular in the business, administrative and legal domains as well as in science and research because this feature assures the reliability and trustworthiness of result derived from data itself.
Data integrity imposes a strong commitment on the organisation involved in the data curation and preservation, by adopting well defined rules of actors involved in the processes, as well as standards and procedures. But to provide data assuring its integrity allows improving the utility for the Data users and consequently the benefits.

4. Completeness

Data completeness is defined as the degree of data to be provided with all the comprehensive and correct information in order to facilitate future discovery, access, and reuse. That includes any description on the resource's provenance and the context of its creation and use. This is a data quality dimension dealing with how complete the data is. In any data resource, it is essential to meet requirements of current as well as future demand for information. Data completeness assures that the above criterion is fulfilled.

5. Usability

ISO defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."
The usability improves the capability to compare, correlate and aggregate set of data. Usually usability of set of data is assured by the adoption of common standard and methods. In terms of process cost, of course providing usable data means to have defined preservation plan, standard and method agreed with community.

3.3 Preservable Data Value Model

The SCIDIP-ES Value model (fig. 2 and 3) in order to overtake the mentioned limits of the other models (par. 2.2), has tailored the benefit/cost analysis, extending it with the adoption of the attribute approach. The inclusion of them in that model is important since they identify the quality level required for guaranteeing the usage of the data over time, at which are closely related the generation of benefits as well as of the organizational costs. The former is performed by the proposed benefit framework as well as by the data activities analysis for the cost analysis based on the ABC model

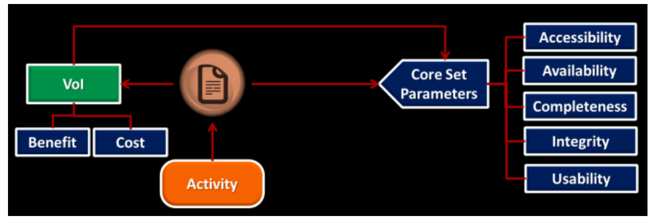

Figure 2. Value of Preservable Information Model

In this perspective, that model also takes in consideration the main relevant users: data provider/manager and data user. For this reason, the model has been extended by considering the specialization of activities carried out/controlled by two users, as shown in Figure 3.

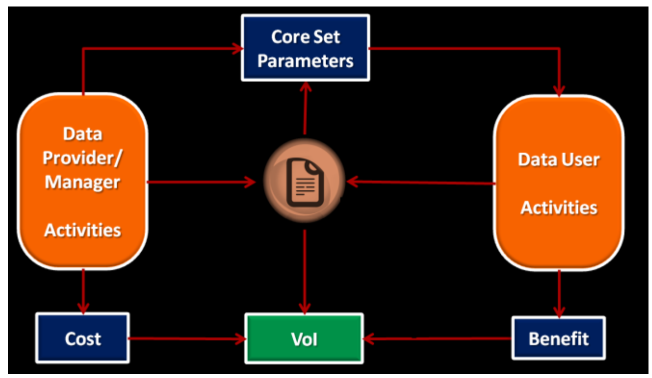

Figure 3. SCIPID-ES Value Model

This model is centred on the data and starts from the consideration that data has a value (VoI), which is determined and impacted from benefits and costs. The benefits are generated by usage of data, while the costs are generated from the activities performed from the data provider/managers in order to maintain and provide the data itself. Moreover, the data is characterised by attributes (so-called parameters), which impact on the utility perception of the data user. Indeed, that latter will decide to use data according to values assumed by core set parameters and acceptance thresholds/criteria. It is important to highlight that acceptance thresholds/criteria may differ between different organizations, based on their internal policies and objectives. However, improving those attributes, according this model, means to increase the probability that data /information will be used over the time, increasing consequently the possibility to generate more benefits. On the other hand, the data provider/manager activities impact on values assumed by core parameters for each data set provided. Consequently, data provider/managers should keep in mind those core parameters when plan or perform activities and choose resources (e.g.: technologies) for preserve digital data.

4. CONCLUSION

The paper addresses the issue of assessing the value of digital asset for ES community, that is the huge amount of data available from a variety of ES missions and preserved in ES archives. Of course, this is a crucial point also for other fields as Social Science, Bioinformatics, Astronomy, Particle Physics, Medicine and Health, where the quantity of information that will be stored in digital form will increase dramatically.

This amount of data has to be preserved and the most difficult task to be performed by data owners is the assessment of its value. It cannot be derived from just the cost of missions, because that is a component which takes into account the only generation aspect, while beyond data generation it has to be considered the whole lifecycle and performed activities on data itself. On this perspective, this paper has described the existing models from the state of the art for assessing the value and those models have been analysed for identifying limitations in supporting data owners. Consequently, in order to overtake those limits, it has been described the proposed approach for adapting the existing models, mainly based on historical cost approach (process oriented). Moreover, it has been enriched by including the benefit framework and by analysing the contextualised activities for cost definition, according to ABC model. Finally, the model has been extended by characterising the data through a core set of parameters which may potentially impact on value of data itself.

This model is being assessed in ES scenarios with data curators and archive managers, in order to carry out an economic sustainability analysis of: i) the Long Term Data Preservation (LTDP) in the ES domain as well as, ii) the developed SCIDIP-ES Infrastructure which provide a set of services and toolkits for managing digital preservation of ES-data.

5. ACKNOWLEDGMENTS

Work partially supported by the European Community under the Information Society Technologies (IST) program of the 7th FP for RTD - project SCIDIP-ES, ref. 283401.

6. REFERENCES

[1] LIFE project - Life Cycle Information for E-Literature (2005-2010) http://www.life.ac.uk/

[2] CMDP Project - Cost Model Digital Preservation . 2010-2012. http://www.costmodelfordigitalpreservation.dk/

[3] KRDS Project - Keeping Research Data Safe. 2007- 2010. http://www.beagrie.com/krds.php

[4] Ensure Project - Enabling kNowledge Sustainability Usability and Recovery for Economic value .2011-2014. http://ensure-fp7-plone.fe.up.pt/site/

[5] The Consultive Commitee for Space Data Systems: Reference Model For An Open Archival Information System (OAIS), CCSDS 650.0-M-2, 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf

[6] Molly K. Macauley. The Value of Information: A Background Paper on Measuring the Contribution of Space-Derived Earth Science Data to National Resource Management. Discussion Paper 05–26. May 2005. http://rff.org/rff/Documents/RFF-DP-05-26.pdf

[7] D. Moody, P. Walsh. Measuring The Value Of Information: An Asset Valuation Approach. European Conference on Information Systems (ECIS'99). 1999.

[8] Charles Beagrie Ltd. Guide to the KRDS Benefits Framework. Keeping Research Data Safe. report, v3. July 2011. http://www.beagrie.com/KRDS_BenefitsFramework_Guidev3_July%202011.pdf

[9] S. Schwolow , M. Jungfalk . The Information Value Chain Strategic Information Management for Competitive Advantage . Bachelor's Project at Copenhagen Business School. 2009. http://www.informationvaluechain.com/wp-content/uploads/information-value-chain.pdf

[10] Mark Bide, Mark Bide & Associates - Business Models For Distribution, Archiving And Use Of Electronic Information: Towards A Value Chain Perspective - A Study For ECUP+ . 1999. ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/businessmodels.pdf

[11] G. Caruso, B. Matthews, B. Polsinelli . Parameters for Long Term Preservation and data sustainability model. SCIDIP-ES-DEL-WP31-D31.1. April 2013

[12] Heinrich, Bernd, Klier, Mathias. A Novel Data Quality Metric For Timeliness Considering Supplemental Data. 17th European Conference on Information Systems 2009. http://www.ecis2009.it/papers/ecis2009-0656.pdf

[13] Sally Vermaaten, Brian Lavoie and Priscilla Caplan. Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment. September/October 2012 D-Lib Magazine http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html

[14] Angus Whyte and Andrew Wilson. Appraise & Select Research Data for Curation. Digital Curation Centre and Australian National Data Service "working level" guide, 25 October 2010. http://www.dcc.ac.uk/resources/how-guides/appraise-select-data

[15] NASA's Cost Estimation Tool - http://opensource.gsfc.nasa.gov/projects/CET/index.php

[16] SCIDIP-ES project – SCIence Data Infrastructure for Preservation with focus on Earth Science. 2011-2014. http://www.scidip-es.eu

[17] Anna S. Palaiologk, Anastasios A. Economides, Heiko D. Tjalsma, Laurents B. Sesink : An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS , International Journal on Digital Libraries, Volume 12, Issue 4, pp 195-214, September 2012. http://link.springer.com/article/10.1007%2Fs00799-012-0092-1

[18] Beagrie, N., Chruszcz, J., & Lavoie, B. : Keeping research data safe: A cost model and guidance for UK universities. Copyright HEFCE 2008. http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf.

[19] IPCC - Intergovernmental Panel on Climate Change. http://www.ipcc.ch/

[20] UNFCCC - United Nations Framework Convention on Climate Change. http://unfccc.int/

[21] ARGO project . 2000-on-going http://www.argo.net/

[22] GRACE – Gravity Recovery and Climate Experiment, since 2002. http://www.csr.utexas.edu/grace/