Accounting for the Value of Earth Science Data

Bruce R. Barkstrom[1], **Paula L. Sidell** [2]

[1] ***Retired NOAA***

15 Josie Lane, Asheville, NC, USA

EMail: alicebarkstrom@verizon.net

[2] *119 Sandpiper Lane, League City, TX, USA*

EMail: insomnia1@comcast.net

ABSTRACT

This paper considers a cost accounting framework for long-term preservation of Earth science information in governmental archives. It considers three standard approaches to valuing that information: net present value of economic benefits, probable cost of replicating information, or tracking the original cost of creating value. Only the latter appears viable, and even then, many of the original costs may be regarded as `sunk costs' by the time data collection begins. This means that the value of climate data is fundamentally determined by the original cost of creating the software to interpret the data and validating it. Furthermore, because this information does not depreciate in value over time, its treatment as an asset is governed by the contingent liability of loss. This conclusion suggests that the costs of preservation may be regarded as a form of 'insurance' against various kinds of losses. Because the requirements for long-term information preservation are extremely stringent, it appears appropriate to build a formal threat model for such losses. Such a model can then help prioritize various operational expenses.

Keywords: Information Value, Heritage Asset, Chart of Accounts, Contingent Liability,Information Loss Threat Model

INTRODUCTION

This paper considers the accounting of Earth science information in governmental archives where long-term preservation is a primary concern. The purpose of this study is to provide a basis for managing the resources of archives using sound accounting principles even though the information must remain useful to a designated user community over very long periods of time.

Archives or long-term repositories of information, like most other organizations, run on money. Thus, it appears helpful to use accounting structures and principles to organize the flow of funds. Table 1 presents a simple balance sheet that contains the largest categories of flows.

Debit	**Credit**
Assets [*A*]	Fund Balance [*F*]
	Liabilities [*L*]
Expenses [*E*]	Income [*I*]

Table 1. Basic Chart of Accounts.

This chart of accounts is typically applied at discrete times, whose starting and ending points are when the books are closed. For the interval between these times, Δt, the fundamental balance of accounts is

$$\Delta A_{\Delta t} + \Delta L_{\Delta t} = I_{\Delta t} - E_{\delta t}$$

In conventional language, the change in assets and liabilities during the time interval between closings must equal the amount left over after expenses are deducted from income. If there were no expenses and no change in liabilities, then all of the income would go to increasing the value of the assets.

While this simple table offers simple guidance, the interesting aspects of the accountant's view only become visible when we add detail to the simple picture. In the material that follows, we first consider the governmental accounting context, which affects both the nature of the funding flows and the concept of asset value. Second, we consider the nature of the assets within the context of the life cycle costs of Earth science information. Third, we consider the nature of liabilities. This leads us to a conclusion that is summarized in a more detailed chart of accounts that must recognize both the economic and non-economic value of Earth science information in governmental archives and repositories, as well as the fact that the primary liabilities lie in the "insurance premiums" that protect this information from loss.

THE GOVERNMENTAL ACCOUNTING CONTEXT

In this paper, we assume that an archive or repository (meaning an entity that holds data and information for longer than three years) is a governmental institution. This assumption means that income from the information it holds does not flow to the institution. Rather, the archive receives funding as a governmental entity. Within this federal governmental context, Earth science archives operate from the *general fund* of the government. In other words, for a federal Earth science archive, an increase in the "fund balance" is simply an increase in the net valuation of the data and metadata in the archive. Such an increase does not accrue to the archive's owners or managers.

A critical element of difference between a governmental archive of Earth science information and those of other kinds of organizations arises from the fact that the value of the archive, as represented by its "fund balance", derives primarily from the value users obtain through accessing the contents, rather than any intrinsic value of the tangible containers which hold it. The contents of an Earth science archive might be called the intangible residue of the intellectual capital expended by the people who create the data and metadata. By themselves, the bits that represent the archive's contents are of no value unless they can be read and understood by a designated user community.

Granof [1, pp. 2 ff.] notes that "Governments and not-for-profits differ from businesses in ways that have significant implications for financial reporting. For the most part, governments . . . provide services targeted at groups of constituents either advocating a political or social cause or carrying out research or other activities for the betterment of society. The objectives of governments . . . cannot generally be expressed in dollars and cents, and they are often ambiguous and not easily quantifiable."

He continues "If the financial statements of a government . . . incorporate only monetary measures, such as dollars and cents, they cannot possibly provide the information necessary to assess the organization's performance. For an organization to report properly on its accomplishments, it must augment its financial statements to include nonfinancial data that relate to its objectives." He comments later [1, p. 5] "state-of-the-art budgets establish a [basis for evaluating accomplishments] by indicating not only how much will be spent on a particular activity, but what that activity will achieve. A post-period assessment can then focus not only on whether the entity met its revenue and expenditure projections, but, equally important, on whether it attained what was expected of it. Evaluators can then assess organizational efficiency by comparing inputs (such as dollar expenditures) with outcomes (results)."

The key point of this discussion is that in order to achieve the goals of this paper, we need ways of expressing both the non-economic objectives of federal Earth science archives and the economic processes that affect the achievement of these objectives.

THE GOVERNMENTAL ACCOUNTING CONTEXT

In order to properly account for the nature of the accounting transactions, we need to deal with both the *basis of accounting* and its *measurement focus*. The basis determines when transactions and events are recognized on the entity's books. The focus determines what resources are involved in the transactions. Generally, the federal government is moving toward a full accrual accounting basis, which means that a transaction appears on the books when it has its substantive economic impact. This also implies that its measurement focus will be on all of its economic resources. The alternative is a cash accounting basis. In this case, the transaction is only recognized when the cash related to the transaction is received or paid. The focus of a cash basis is clearly upon the cash, not upon the total assets and liabilities of the entity involved.

Granof [1, p. 260ff] notes that "Most assets should be depreciated over their estimated useful lives in a manner that is 'rational and systematic' (i.e., using one of the methods commonly used by businesses). However, governments, like businesses, do not have to depreciate inexhaustible assets, such as land, works of art, or historical treasures." It is critical to note that because Earth science observations are unique and therefore irreplaceable, they are also non depreciable.

Information Life Cycle

For our purposes, Earth science information may be viewed as developed using a five-stage life cycle:

1. Acquisition of measurement resources, such as instrument development
2. Data processing operations, including data validation and reprocessing
3. Transfer of processed data from production and validation teams into archives
4. Archival curation, including transformational migration and reprocessing
5. Providing user access and access transformation

These phases may overlap to some extent. For example, a multi-instrument climate data record may have procurement of new instruments going on while the current generation of resources is sending down data that the production and validation teams are processing into products that are sent to Earth science archives. Table 2 provides a rough categorization of the assets or operations that need to be included in the list of resources, as well as the accounting treatment we expect for each category.

In this table, we use the abbreviations

- For Activity
    - **T**: Activity produces a **Tangible** asset, such as hardware, software source code, or documentation
    - **I**: Activity produces an **Intangible** asset, primarily information that may be used by researchers, decision makers, or other data users
- For Expected Interval
    - **LOM**: Life Of Mission, which may be taken as ten years for a single satellite or instrument, and which may increase to more than a century for operational environmental observation capability
    - **LOD**: Life Of Data, which NARA defines as 75 years beyond scientific research use. Here, we take this time period to be 200 years – give or take
- For Accounting Basis
    - **M**: Modified Accrual accounting basis
    - **F**: Full Accrual accounting basis
- For Account Type
    - **DC**: Development and Construction
    - **O**: Operations
    - **HS**: Hardware and Software, including accounts for initial capitalization, depreciation, and refresh/upgrades
    - **SD**: Specialized scientific software Development

Information Asset Valuation

The accounting categories in Table 2 provide some unexpected insight into the sources of asset value. Several expense accounts, such as launch vehicles or instruments would normally be valued during construction as constituting inventory items. However, after launch, the vehicle is a `sunk cost' – usually literally. The same is true of the instruments and satellites themselves after they stop operating.

Activity	Expected Interval [yr]	Accounting Basis	Account Type
Satellite Missions			
Instrument Dev. [T]	5	M	DC
Instrument Char. & Cal Facil. [T]	5	F	DC
Instrument Char. & Cal. [I]	0.2	M	O
Instr.-Sat. Integration	0.5	M	O
Sat.-Vehicle Integration	0.5	M	O
Initial Sat./Instr. Checkout	0.2	M	O
Sat./Instr. Ops.	LOM	F	O
In Situ Data Networks			
In Situ Data Site Dev. [T]	2	M	HS
In Situ Data Site Ops.	LOM	F	O
Science Data Production			
Science Algorithm Dev. [T]	LOD	F	SD
Science Data Validation [I]	LOM	F	O
Science Data Production [T]	LOM	F	O
Science Data Prod. Facility [T]	LOM	F	HS
Archival Activities			
Archive Ingest [I]	LOM	F	O
Archive Curation [T]	LOD	F	O
Archive Facilities [T]	LOD	F	HS
Archive Operations	LOD	F	O
User Access			
User Access Evolution [I]	LOD	F	SD

Table 2. Earth Science Life-Cycle Activities and Accounting Basis.

The same accounting approach appears to be applicable with respect to such expense categories as hardware integration, satellite operations, and ground system operations.

The question then is `what constitutes the residual value of Earth science data, metadata, and the information that accrues from them?' This is a difficult question to answer, although the basic answer is `because the data and information are useful.'

One part of the difficulty lies in the fact that most of the archived data is a public good, established at public expense. A second difficulty arises because this data and its information value lie in three categories of use:

1. Immediate use for warning systems, short-term decision making, and information flows, such as weather forecasting
2. Relatively short-term use for scientific process understanding, such as creation of data sets that enable the scientific research community to improve physical process parameterizations
3. Long-term statistical use, including measurement of trend and extreme value distribution

The use of data for warning systems is closely tied to governmental activities, including crisis planning, disaster management, and recovery. Short-term decision making may involve non-governmental economic elements, as well as some governmental activity. There are clearly two value systems at play here: one dealing with activities designed to make a profit from the data use, the second intended to provide a public good, particularly dealing with avoiding the loss of life and property.

The longer term data and information uses add an additional value system: the improvement of scientific understanding. This component is very difficult to place in the same value system as the value of profit-making economic activity, or the protection of life and property. We note that the issue of valuation is key to the issue of prioritizing budgets, an issue that has consumed a very large amount of discussion in the literature. There are a number of sophisticated analyses [2, 3] that attempt to create a "currency" known as "utility" that allows comparability between different systems. [4] suggests use of information metrics. Current research, e.g. [5], suggests using auction theory. Indeed, peer reviews of measurement system proposals may operate as a form of auction that is capable of considering the `science value' of data sets without probing deeply into the economic return from a project.

Rather than attempting a valuation based on such first-principles approaches, the accounting standards suggest three approaches:

1. Valuing the archive information based on the cost of producing it and incorporating it into the archive – which we might describe as the acquisition value of these assets
2. Valuing the archive information based on the cost of replacing it
3. Valuing the archive information based on the future value of the use of these assets

Based on the previous discussion, the accounting standards suggest that many of the tangible items, such as launch vehicles and hardware are to be regarded as sunk costs for this valuation. The cost of production then boils down to the direct costs of producing the software that produces the useful information and of validating the data values. In essence, these costs boil down to the cost of data producer time for the production of the production software.

The second approach to asset valuation for long-term archival is similar to the approach used by the insurance industry: the replacement value of similar goods. In the case of many kinds of archived assets, it is not clear how one could directly replace unique and strictly irreproducible items. We do not base the insurance value of the "Mona Lisa" upon the cost of producing a "new and essentially indistinguishable" copy using modern technologies. Such an approach is also not directly relevant to unique observations of the Earth, since there is no way to recreate missing observations.

The third approach assumes that the data and information in the archive produces a flow of future value. The normal approach for such evaluation would be to discount future benefits. This approach might be appropriate if the benefits were strictly economic, as they would be for a profit-making enterprise. Of course, the economic situation of the past year illustrates the difficulties of economic forecasting, as does a recent review [7] of the legal standing of climate models, as well as models of the economy. [8] emphasizes the difficulty of assigning a discount rate for this kind of forecasting.

[9] has the following common-sense advice: "a balance sheet will be prepared utilizing a variety of valuation methods – the selection is normally based on the nature of the item and the relevance and reliability of the method of accounting for that item." [9] continues by noting "To be relevant, information about an item must have feedback value and/or predictive value for users and must be timely. Information is relevant if it has the capacity to make a difference in the decisions of owners, investors, creditors, or other interested parties." "To be reliable, information about an item must be representationally faithful, verifiable, and neutral. Information is reliable if it is sufficiently consistent in its representation of the underlying resource, obligation, or effect of events; and sufficiently free of error and bias to be useful to owners, investors, creditors, and others in making decisions." [9] then notes that "If two methods are equally relevant and reliable, then the method with the lowest cost to the preparer would probably be chosen."

LONG-TERM PRESERVATION EXPENSE ACCOUNTS

Digital preservation creates a non-standard context for accounting and economic valuation. In this section, we extend the discussion of context and suggest a systematic, quantitative framework for dealing with the necessities of long-term preservation. The philosophy of our approach follows standard accounting principles for heritage assets: if a scientifically valuable artifact retains its value without depreciation, then threats to the usefulness of that asset should be treated as impairments of the asset's value. Because most of these threats are probabilistic in nature, they may be treated as creating contingent liabilities for which expenditures that reduce that risk are equivalent to insurance premiums. Our strategy for estimating liabilities follows the principles suggested in [10]. There are three processes required for this strategy:

1. Identify the threats to the asset value
2. Estimate the probability of loss and the probable value of the loss if the threat materializes
3. Develop an affordable strategy to mitigate the risk

We can identify seven risks that become critical for longterm preservation:

1. Institutional instability and funding flow changes
2. Physical disaster
3. Operator error
4. Media, hardware, and software errors
5. Loss of context, including software obsolescence
6. IT Security incidents
7. Evolution of hardware and software,

Long-term preservation is difficult. We suggest that the normal period for preservation is two hundred years. Suppose that the probability for loss in a year is p. This means that the probability of survival at the end of the first year is $1 - p$. In the second year, the probability of loss is $p(1 - p)$; that of survival is $(1 - p) - p(1 - p) = (1 - p)^2$. Likewise, the probability of survival after the third year is $(1 - p)^3$, and so on. If we extend this calculation to two hundred years, the probability of survival into the two hundred and first year is $(1 - p)^{201}$. Table 3 provides some numerical values for this last measure of survival probability.

Probability of Loss per Year, *p*	**10%**	**1%**	**0.01%**
Probability of Survival to Year 201	6.4×10^{-10}	0.13	0.98

Table 3. Probability of Content Survival for 200 Years..

All seven of the identified risks appear to have loss probability rates above five percent per year. Clearly, we have a far to go in reducing the probability of loss to an acceptable level for long-term preservation of this critical public information resource.

A REFINED BALANCE SHEET AND CONCLUDING REMARKS

We summarize our results in table 4, which provides a refined chart of accounts for an Earth science archive.

The view we have taken suggests three areas in which further research is necessary:

1. Formal *valuation modeling* that allows for the fact that Earth science data may have both scientific value and economic value, where the former valuation is non-depreciable and both valuations are used in budgetary prioritizations
2. Formal *threat modeling* to produce well-founded loss probabilities and sensible risk reduction guidance
3. *Evaluation of future cost profiles* that include the impact of exponential declines in hardware and media costs, as well as the (usually mild) rates of increase in personnel cost

4.

Debit	**Credit**

Assets
    Data
    Metadata
    Documentation
    Human Capital
General Assets
    Buildings and Related Assets
	Capitalization

	Liabilities
	    Loss of Context
	    Format and Software Obsolescence

Expenses [*offsets to avoid asset impairment*]
    Hardware Evolution
    Software Evolution
    Replication
    Automation and Automation Testing
    Operations
    Power and Air Conditioning
    Preservation Planning
    Administration
	Income

Table 4. Refined Chart of Accounts for an Earth Science Archive.

REFERENCES

[1] – M. H. Granof: Government and Not-For-Profit Accounting: Concepts & Practices, 4th Edition, Wiley, Hoboken, NJ (2007).

[2] – T. L. Saaty: Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process, RWS Publications, Pittsburgh, PA (2006).

[3] – R. L. Keeney and H. Faiffa: Decisions with Multiple Objectives: Preferences and Value Tradeoffs, Cambridge University Press, Cambridge, UK (1993).

[4] – D. W. Hubbard: How to Measure Anything: Finding the Value of Intangibles in Business, John Wiley, Hoboken, NJ (2007)

[5] – Y. Shoham and K. Leyton-Brown: Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, Cambridge University Press, Cambridge, UK (2009)

[6] - Consultative Committee on Space Data Systems (CCSDS), Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. January 2002. This Recommendation has been adopted as ISO 14721:2003, http://ssdoo.gsfc.nasa.gov/nost/isoas/ref model.html.

[7] – D. A. Farber: Climate Models: A User's Guide, UC Berkeley Public Law Research Paper No. 1030607, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1030607 (2009)

[8] – N. Stern: Stern Review on the Economics of Climate Change, UK Treasurer's Office, available at http://www.hm-treasury.gov.uk/sternreveiw_index.htm (2007)

[9] - Farm Financial Standards Council: Guidelines for Valuation Methods, (2006) available at http://www.ffsc.org/guidelines/valuation.htm

[10] – B.W. Boehm, B. W.: Software Risk Management: Principles and Practices, IEEE Software, Vol. 8, pp 32-41 (1991)