A MATURITY MODEL FOR DIGITAL DATA CENTRES

W Hugo*[1]

**[1]South African Environmental Observation Network, De Havilland Crescent, Pretoria, 0001, South Africa*
Email: wim@saeon.ac.za

ABSTRACT

Digital Data and Service Centers, such as is envisaged by the ICSU World Data System (WDS), are subject to a wide-ranging collection of requirements and constraints. Many of these requirements are traditionally difficult to assess and to measure objectively and consistently. As a solution to this problem, an approach based on a maturity model is proposed: this adds significant value not only in respect of objective assessment, but also assists with evaluation of overlapping and competing criteria, planning of continuous improvement, and progress towards formal evaluation by accreditation authorities.

Keywords: Maturity Model, Continuous Improvement, Objective Assessment, Key Performance Areas, Accreditation, Preservation, Data Centers

1 INTRODUCTION AND PROBLEM STATEMENT

Digital Data and Service Centers, such as those envisaged by the ICSU World Data System (WDS), face a variety of key performance areas in respect of their operations, planning, and management – derived from a variety of sources. These may include organizational objectives, user requirements, constraints and requirements imposed by funding agencies, and, of course, the criteria set by the WDS in respect of different categories of membership. In addition, there may be local legal compliance required in respect of preservation and archiving, and technical constraints could include standards for interoperability, cataloguing, processing, and the like.

There are several management problems associated with this wide variety of requirements imposed on a Centre. Examples:

- There is an overlap, though sometimes a subtle difference, in requirements derived from multiple sources.
- Many of the requirements imposed on a Centre cannot be measured objectively, and different observers may come to different conclusions in respect of the current performance of an organization or Centre.
- Knowledge in respect of successful approaches are not easily disseminated or transferred.

2 PROPOSED SOLUTION

A solution to these, and several other smaller management challenges, may be provided by applying the principles of a 'Maturity Model' (Humphrey, 1987), analogous to the approach first proposed by Carnegie-Mellon Institute for the assessment and management of organizations involved in software creation and delivery. It provides a framework that addresses many of the management challenges that we have described thus far, and serves as a repeatable and less subjective measuring instrument to assess performance of Digital Data and Service Centers.

3 REQUIREMENTS PLACED ON DIGITAL DATA CENTRES

We will be using a hypothetical data center in the field of earth and environmental sciences to develop our solution. We assume that the data center will be distributed physically (which is increasingly the norm, and adds to the complexity of management), and that it needs to comply with typical interoperability requirements. Such a center might typically expect to

1. **Derive strategic and management objectives** from a business planning process, which, in turn, is subject to financial and other resource constraints, while presumably serving the need of one or more communities. These communities may not all be scientists, and could include the wider public, decision makers, and private enterprise;
2. **Link to a Community of Practice** that imposes constraints and requirements, with the constraints including aspects of mandate and scope of operations, and the requirements often aimed at ensuring interoperability and trouble-free access to the Centre's resources. The latter aspect may include data access policies. The Centre also generally needs to ensure that it meets the requirements of the Communities of Practice that it serves, defining appropriate products and services and service level agreements in the process;
3. **Make provision for physical and software infrastructure** to support its products and services, which may include functions of access, preservation, and processing requirements, as well as measures whereby interruption of service and risk to assets are minimized. This requirement becomes quite complex in the case of a physically distributed system, and may require the separation of archiving/ preservation arrangements from those aimed at operational data and services;
4. **Apply due diligence and sound governance** in respect of its operations, covering aspects such as independent oversight, risk management, adequate planning for long-term feasibility, and proper liaison with relevant stakeholders. There may be multiple jurisdictions that impose legal requirements and policy constraints on the Centre.

The large number of requirements and constraints deriving from the above can be arranged into an objective hierarchy (or network, since some of the objectives have multiple links to others), and each of these objectives can theoretically have a goal and current level of performance as a minimum (Brehmer, 2005). This is not new: the process is routinely performed in many private and public organizations as performance management.

The main difficulty lies with the *measurement* of the performance, which, for many of the typical requirements and constraints described above, is often performed arbitrarily and subjectively. The main purpose of this paper is to promote the use of Maturity Models to assist with objective measurement of these.

4 MATURITY MODELS APPLIED TO DATA CENTRE OBJECTIVE HIERARCHIES

The common definition of a Maturity Model is "a [framework] that describes how well the behaviors, practices and processes of an organization can reliably and sustainably produce required outcomes" (SEI, 2012). By creating such a framework, there is several side benefits that can be obtained that will be discussed in detail later on, but the obvious structure in the framework is the descriptions associated with predefined levels of performance. These levels of performance are typically designated as follows:

- **Level 1 (Initial)**: Usually associated with ad-hoc approaches, undocumented processes, and little guarantee that a given outcome can be achieved. Knowledge and capacity are centered in individuals. The organization is often ignorant of best practice and of applicable or useful standards and specifications.
- **Level 2 (Repeatable)**: Processes are documented in sufficient detail to ensure continuity and allow reliable execution by a number of participants.
- **Level 3: (Defined)**: Not only are processes documented, but they are also standardized and aligned where applicable to national or international standards and specifications.
- **Level 4: (Managed and Auditable)**: Performance metrics are being collected in respect of achievement of objectives, compliance with standards, and independent audits are performed from time to time to confirm such compliance.
- **Level 5: (Optimized)**: Deliberate process optimization is undertaken, and a regime of continuous improvement is possible.

These levels of performance are, of course, generic, and needs to be translated into corresponding descriptions for each of the objective hierarchy elements applicable to a Data or Services Centre. The example in Figure. 1 deals with 'Meta-Data Interoperability'. Deriving similar descriptions for each performance level across all relevant objectives in the hierarchy leads to a comprehensive 'Maturity Matrix'.

There are several side benefits and additional uses of this approach, in addition to being able to identify the level that most closely matches current performance (and in the act of doing so, making an objective and repeatable assessment):

- Organizations often do not know where to start. By having access to a maturity matrix, it is possible to evaluate a feasible entry point;
- The matrix can, and should, contain the benefit of prior experience – and each entry may be supported by best practice, standards, guidelines, and specifications.
- It can assist multiple organizations with roughly the same objective hierarchy to align and pursue a shared vision (for example, in the ICSU WDS).
- It assists with planning the next level of performance as a set of explicit, measureable objectives and to priorities such actions that may be needed to achieve it.
- It serves to define a level of performance across a collection of objectives, and as such can be used to envision the requirements imposed by certification or audit authorities, for example, by defining the level of performance required to be certified as a 'trusted digital repository'.
- It provides and relatively objective way to compare the performance of organizations, should the need arise to do so.

Maturity Matrix or Framework						
		Initial Phase	Repeatable	Defined	Managed/ Auditable	Optimised
		Level 1	Level 2	Level 3	Level 4	Level 5
IN	Meta-Data Inter-operability	No interoperability	Standardised software allow standards-based exchanges on an ad-hoc basis with external applictions.	Meta-data interoperability requirements are defined and translated into a set of portal functions and services.	Tasks such as harvesting from participating portals are routine, automated, and can be managed by using portal functionality.	Meta-data duplicates are eliminated, update frequency is managed, optmised harvesting paths, and links are continuously tested.

- **Figure 1.** Example of a Maturity Matrix entry for a Specific Objective

5 CONCLUSION

Hence, such an approach can be useful to establish all of the following:

1. Current level of performance;
2. A set of internal objectives and self-assessment against these;
3. A set of future goals and milestones to support a process of continuous improvement;
4. A quality assurance program;
5. Accreditation and external audit mechanisms.

Current work will be extended in the near future to develop specific matrix entries for a wide variety of input requirements, based on the scope discussed in the paper. The intention is to establish this as a community resource that can be edited by any number of collaborators with a view to its refinement, validation, and extension, thereby serving the ICSU WDS specifically, and scientific data systems and services in general.

6 ACKNOWLEDGEMENTS

The ideas and concepts described in this paper derive from an informal discussion group on Digital Data Preservation, that in addition to the author, includes Dr Lucia Lötter of the Human Sciences Research Council, South Africa, Dr Heila Pienaar of the University of Pretoria, and Dr Martie v Deventer of the Council for Scientific and Industrial Research, South Africa. The work is supported in part by the National Research Foundation of South Africa and the South African Environmental Observation Network.

7 REFERENCES

Brehmer, B (2005): "The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control", *10th International Command and Control Research and Technology Symposium.* Retrieved December 2011 from http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/365.pdf,

GEO (2007): "*Tactical Guidance for Current and Potential Contributors to GEOSS", Document 24, Group on Earth Observations*, Retrieved December 2011 from http://earthobservations.org/documents/portal/24_Tactical%20Guidance%20for%20current%20and%20potential%20contributors%20to%20GEOSS.pdf

Humphrey, W.S (1987): *"Characterizing the Software Process- A Maturity Framework", CMU/SEI-87-TR-11, a Technical Report prepared for SEI Joint Program Office*. Retrieved December 2011 from http://www.sei.cmu.edu/reports/87tr011.pdf

ICSU-WDS (2010), "*Certification of World Data System Facilities and Components*", Retrieved December 2011 from http://icsu-wds.org/images/files/Certification_summary_23Oct2010.pdf

OAIS (2009): "Reference Model for an Open Archival Information System", Draft Recommendation for Space Data System Standards", CCSDS Secretariat, Retrieved December 2011 from http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf

SEI (2011): *"Getting Started", Online guidance published by the Software Engineering Institute at Carnegie-Mellon University,* Retrieved December 2011 from http://www.sei.cmu.edu/cmmi/start/

8 APPENDIX: DETAILED OBJECTIVE HIERARCHY

The following sources were applied in deriving an integrated objective hierarchy for a digital data and service center serving the Earth and Environmental data community:

1. Requirements imposed by the **Governance Framework** of the host organization, that includes aspects such as risk management, proper systems engineering and adherence to community requirements, proper technology planning, and the like.
2. Requirements that need to be met for acceptance into the **World Data System** and eventual accreditation (ICSU-WDS, 2010).
3. **GEOSS Architecture** and Interoperability requirements (GEO, 2007).
4. **Open Archival Information System** (OAIS) requirements (OAIS, 2009).

The objectives derived from these sources can be arranged and collated into a hierarchy. The table indicates the source of each objective, and further makes an assessment of its likely contribution to the generic software engineering goals of **Availability**, **Usability**, and **Reliability**. Each of these objectives can be expanded into 5 descriptive levels of performance – part of our future work. A detailed matrix can be obtained as a downloadable spreadsheet: See
http://data.saeon.ac.za/documentation/it-governance/governance/G328.4.1%20Governance%20Matrix.xlsx