Comparative Study of Data Management Capability Maturity Models and Implications: Postprint

Authors: Ye Lan

Date: 2023-04-01T16:15:56+00:00

Abstract

[目的/意义] This study conducts a comparative analysis of data management capability maturity models to provide reference solutions for libraries in selecting and applying these models. [方法/过程] Through examination of model documentation, the structure of each model is introduced. A comparative analysis method is employed to evaluate seven data management capability maturity models across five dimensions: the configuration of evaluation dimensions (functional domains), organizational system of the model, evaluation criteria and rules, operability, and openness. [结果/结论] Each model possesses distinctive features. CMMI, DMM, DCAM, and China's DCMM exhibit relatively comprehensive evaluation dimensions. China's DCMM, namely the China Data Center Service Capability Maturity Model, demonstrates strong operability, and its evaluation methodology combining quantitative and qualitative approaches merits emulation. The Research Data Management Capability Maturity Model (by Professor Qin Jian's team at Syracuse University) and the Research Data Management Capability Maturity Model (by Australian ANDS) consider the practical realities of library data management and can serve as preferred models. Based on the comparative analysis and evaluation results of the models, three solutions are formulated for libraries to select and apply data management capability maturity models adapted to different contexts.

Full Text

Comparative Research and Implications of Data Management Capability Maturity Models

Ye Lan
Shenzhen University Library, Shenzhen 518060

Abstract

[**Purpose/Significance**] This paper compares and analyzes data management capability maturity models to provide reference solutions for libraries in selecting and applying such models. [**Method/Process**] Through literature review and content analysis, the structure of each model is introduced. A comparative analysis method is also used to evaluate seven data management capability maturity models from five aspects: evaluation dimension (functional area) configuration, organizational system, evaluation standards and rules, operability, and openness. [**Result/Conclusion**] Each model has its own distinctive features. The evaluation dimensions of CMMI-DMM, DCAM, and China DCMM are relatively comprehensive. The China Data Center Service Capability Maturity Model has strong operability, and its quantitative and qualitative combined evaluation method is worth learning from. The Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) and the Research Data Management Capability Maturity Model (Australian ANDS) consider the actual situation of data management in libraries and can be used as preferred models. Based on the comparison and evaluation results of each model, three schemes have been formed for libraries to select and apply data management capability maturity models in different contexts.

Keywords: data management; data governance; capability maturity model; research data

Introduction

Domestic library attention to data management began around 2011, when Yang Helin [1] first introduced the concept of "data curation" and described data management practices in American universities. Starting in 2013, domestic libraries experienced a surge of research on "data management" and "data curation," with research topics covering almost all essential elements required for implementing complete data management services, including data management needs surveys, top-level data management policies, specific data management service types and models, data management platforms and tools, data management education and training, and data management positions. This research boom continued until 2016, after which, as data management services in foreign libraries became normalized, domestic research enthusiasm also gradually stabilized, shifting toward summarization and future prospects, and contemplating future research directions.

In this wave of data management research, the evaluation of data management service capabilities remains an underexplored area. Developing a model and methodology to help university libraries identify and evaluate their data management service capabilities, thereby achieving management, improvement, and enhancement of service capabilities, is of significant contemporary importance and necessity. Using capability maturity models to assess data management service capabilities is one such innovative approach. The Capability Maturity

Model (CMM) was proposed by the Software Engineering Institute of Carnegie Mellon University in November 1986 and officially released in 1991. The CMM model adopts a staged representation, dividing software development process capability maturity into five levels: (1) Initial Level; (2) Managed Level (originally called Repeatable Level); (3) Defined Level; (4) Quantitatively Managed Level (originally called Managed Level); and (5) Optimizing Level [2]. These five levels form a progressive platform, where the process capability of each level serves as the foundation for reaching the next higher level (skipping levels is generally not allowed), and the process of continuously upgrading maturity is also a process of gradually accumulating process capability.

The capability maturity model is a set of patterns and methods for software process management, improvement, and evaluation, dedicated to gradually improving management levels and continuous process improvement. Data management has process characteristics similar to software development; it is a lifecycle process composed of various activities including management plan formulation, data collection management, data description and archiving management, data processing and analysis management, data preservation management, and data discovery and reuse [3]. This makes the application of capability maturity models to data management feasible. Consequently, relevant organizations and teams in both the library community and industry have actively explored the application of capability maturity models in data management process evaluation, resulting in numerous established data management capability maturity models.

Classic enterprise data management capability maturity models mainly include the Data Management Maturity Model from the CMMI Institute (a subsidiary of Carnegie Mellon University), the Data Management Capability Assessment Model from the Enterprise Data Management Council, Gartner's Enterprise Information Management Maturity Model, and China's national standards "Data Management Capability Maturity Assessment Model GB/T 36073-2018" and "Information Technology Service - Data Center Service Capability Maturity Model GB/T 33136-2016." The most well-known library data management capability maturity models are the Research Data Management Capability Maturity Model developed by Prof. Qin Jian's team at Syracuse University and the Research Data Management Capability Maturity Model from the Australian National Data Service (ANDS). This paper comprehensively analyzes and compares seven existing data management capability maturity models from both enterprise and library contexts to provide reference solutions for libraries in selecting and applying such models.

---

1. Research Progress on Data Management Capability Maturity

Existing literature mainly conducts research from three aspects: First, introducing specific models, such as Li Bing and Bin Junzhi [4] introducing "Data Man-

agement Capability Maturity Assessment Model GB/T 36073-2018," Zhang Kai [5] and Wang Zhipeng [6] introducing the Data Center Service Capability Maturity Model (GB/T 33136-2016), and Ye Lan [7] introducing the Research Data Management Capability Maturity Model from Syracuse University (Prof. Qin Jian's team) and the Research Data Management Capability Maturity Model from Australian ANDS. Second, exploring specific applications of models, such as Li Weimian [8] referencing Qin Jian's team's model to propose an evaluation index system for repository data service capability maturity, and Dang Hongli et al. [9] introducing the CMMI Institute's Data Management Maturity Model into the library field to construct a library data management and service capability maturity model. G. Peng [10] proposed the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix (DSMM) model based on years of experience managing meteorological, environmental, and earth data at the Cooperative Institute for Climate and Satellites-North Carolina (CICS-NC) at North Carolina State University, and is committed to promoting the application and promotion of this model. Third, researching data governance maturity models from the perspective of big data or government data, such as Zhang Yujie et al. [11] studying the construction of a maturity evaluation index system for government big data governance, and Liu Guifeng et al. [12] introducing four typical foreign data governance models.

Comprehensive existing research reveals that there are still few studies providing comprehensive analysis and comparison of various data management capability maturity models, and there is no unified opinion on how libraries should select and apply data management capability maturity models. Therefore, this paper comprehensively analyzes and compares existing data management capability maturity models from both enterprise and library contexts to provide reference solutions for libraries in selecting and applying such models.

---

2. Introduction to Data Management Capability Maturity Models

Data Management Maturity Model (DMM) is developed by the CMMI Institute (a subsidiary of Carnegie Mellon University) based on the foundational principles of Capability Maturity Model Integration (CMMI) and was released in August 2014 [13].

Data Management Capability Assessment Model (DCAM) is led by the Enterprise Data Management Council (EDM Council) in North America, with participation from financial industry enterprises in its compilation and validation. Based on experience summarized from numerous actual cases, it was officially released in 2014, with a second version released in May 2019 [14].

Enterprise Information Management Maturity Model (Gartner EIM) was created by Gartner in December 2008 based on CMMI [15].

Data Management Capability Maturity Model (DCMM) was initiated

in 2014 by the National Information Technology Standardization Technical Committee, led by the China Electronics Standardization Institute, with working group members including Data Governance Workshop, Tsinghua University, China Construction Bank, and China Everbright Bank. It became the national standard "Data Management Capability Maturity Assessment Model GB/T 36073-2018" in April 2018 [4].

Data Center Service Capability Maturity Model was led by China Merchants Bank's data center in October 2014, with 30 experts from finance, telecommunications, energy, manufacturing, and other fields completing the research over two years. It was released as the national standard "Information Technology Service - Data Center Service Capability Maturity Model GB/T 33136-2016" at the end of 2016, representing China's first evaluation standard for data center service capability [6]. In April 2019, led by the Bank of China, work is underway to promote the application for an international standard ISO/IEC PDTR 22564 "Information Technology - Data Center Service Capability Maturity Model" based on this standard.

Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) was proposed in 2011 and released in June 2014 [16]. The model has not seen new changes since its proposal.

Research Data Management Capability Maturity Model (Australian ANDS) was proposed in 2011 by the Australian National Data Service project, with a new dimension "Managing Research Data" added to the original four evaluation dimensions in 2018 [17].

Comprehensive existing research finds that most models generally consist of evaluation dimensions (functional areas), functional sub-areas, and maturity levels. Table 1 summarizes the structure of each model from three aspects: evaluation dimensions (functional areas), functional sub-areas, and maturity levels.

---

3. Comparative Analysis of Data Management Capability Maturity Models

3.1 Configuration of Evaluation Dimensions (Functional Areas) The models show significant differences in the configuration of evaluation dimensions (functional areas), reflected in the varying categories and quantities of evaluation dimensions and sub-areas. Among them, CMMI-DMM, EDM DCAM, China DCMM, China Data Center Service Capability Maturity Model, and the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) have more numerous and comprehensive evaluation dimensions and sub-areas. The Gartner EIM Model and the Research Data Management Capability Maturity Model (Australian ANDS) have evaluation dimensions that basically consist only of first-level indicators without subdivision,

making them relatively broad. The main reason for differences in evaluation dimension (functional area) configuration lies in inconsistent divisions of key data management processes across models. This is caused by the wide scope, numerous contents, and complexity of data management, as well as different evaluation objectives and focus areas of each model [18].

Despite these differences in functional areas, careful observation reveals that five major domains—data strategy, data governance, data quality, data platforms and architecture, and data security—are basically common elements across all data management maturity models, as shown in Table 2 .

3.2 Organizational System of Models The organizational system of models refers to the basis and organization method for selecting each evaluation dimension, mainly from two perspectives: (1) Selecting and organizing evaluation dimensions according to basic elements constituting data management, such as strategy and policy, organization and personnel, and technology and tools (platforms and architecture). This approach basically follows top-level strategy and policy design, achieving organizational data management objectives through effective combination of personnel and resources, attempting to select and organize evaluation dimensions from three aspects: who is responsible for data management (who), what systems and policies are involved (what), and how to technically solve data management (how). CMMI-DMM, DCAM, Gartner EIM, China DCMM, China Data Center Service Capability Maturity Model, and the Research Data Management Capability Maturity Model (Australian ANDS) all have evaluation dimensions related to strategy and policy, organization and personnel, and technology and tools (platforms and architecture). (2) Selecting and organizing evaluation dimensions according to the data management lifecycle. For example, the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) is primarily organized according to the data management lifecycle, dividing key process areas into data collection, data description and presentation, data dissemination, and storage services and preservation.

3.3 Formulation of Evaluation Standards and Rules Evaluation standards and rules develop corresponding assessment criteria for the maturity levels of evaluation dimensions (functional areas). All models are primarily based on qualitative evaluation, mainly listing the characteristics and requirements of each evaluation dimension at each maturity level according to the divided evaluation dimensions. Organizations analyze their current level in each evaluation dimension based on their status. For example, CMMI-DMM, DCAM, Gartner EIM, and China DCMM all belong to qualitative evaluation (see Table 3). The models vary in the level of detail of their evaluation standards and rules, with some conducting evaluation at the macro level and others at both macro and micro levels. Macro-level evaluation standards set evaluation criteria for each maturity level only from an overall perspective, while micro-level evaluation provides corresponding evaluation standards for each major category and

its subcategories at each maturity level. For instance, CMMI-DMM, DCAM, China DCMM, and the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) provide evaluation standards at both macro and micro levels, making their evaluation methods more detailed. The Gartner EIM Model and the Research Data Management Capability Maturity Model (Australian ANDS) only conduct evaluation at the macro level, with relatively coarse evaluation standards.

The China Data Center Service Capability Maturity Model combines qualitative and quantitative evaluation. The model establishes maturity grading rules for evaluation elements, capability items, capability sub-areas, and capability areas in terms of qualitative aspects, while also introducing weights for quantitative evaluation of each evaluation element, capability item, capability sub-area, and capability area. The quantitative evaluation decomposes eight evaluation elements and indicators from seven capability elements—personnel, technology, process, resources, policy, leadership, and culture—to assess the score of each capability item. The weighted average of indicator values forms the maturity of a single capability item, and the maturity of a single capability item is weighted and averaged to derive the overall service capability maturity (maturity derivation method: capability elements → capability items → capability sub-areas → overall capability).

3.4 Operability The main factors affecting model operability include the level of detail in evaluation standards and rules, whether application and operation guidelines are provided, whether there is a broad user base, whether there is a dedicated maintenance organization, and the openness of model-related materials. CMMI-DMM, DCAM, China DCMM, and the China Data Center Service Capability Maturity Model have relatively comprehensive functional areas and provide evaluation standards for each functional area at each maturity level, with detailed evaluation methods, thus possessing strong operability. However, the specific implementation process is relatively complex. Because these models involve numerous functional areas, some with sub-areas, conducting maturity assessments on all content is a huge undertaking. For example, DCAM includes 7 capability areas, 31 capability sub-areas, and 106 capability items, making the assessment process extremely complex. Similarly, the China Data Center Service Capability Maturity Model includes 3 major capability areas, 11 capability sub-areas, and 33 capability items, and evaluating all 33 capability items from eight evaluation elements involves enormous workload. The Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) provides evaluation standards from four aspects—commitment to perform, ability to perform, activities performed, and process assessment—for each of the five key process areas of research data management, which also has certain complexity in implementation. The Gartner EIM Model and the Research Data Management Capability Maturity Model (Australian ANDS) have relatively rough evaluation methods, conducting evaluation only at the macro level without detailing to each functional area, which may result in weak operability

due to unclear evaluation standards during actual evaluation. Overall, whether currently operable or not, all models need to provide application and operation guidelines to enhance operability.

3.5 Openness Through comparison, we find that the openness of each model still needs improvement (see Table 4). Most models cannot obtain the full text for free and require purchase. Many models adopt a membership system, where only formal members can access the assessment solutions provided by the model, such as CMMI-DMM, DCAM, and Gartner EIM [20]. Some models have no dedicated website and are rarely updated and improved, resulting in low openness. Models with better openness include China DCMM, China Data Center Service Capability Maturity Model, Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team), and Research Data Management Capability Maturity Model (Australian ANDS), which provide free access to the full model text.

---

4. Overall Evaluation of Data Management Capability Maturity Models

All seven models compared in this paper evaluate the maturity of organizational data management capabilities from an institutional perspective, with each model having its applicable scope and functional characteristics. In terms of applicable fields, the CMMI-DMM model covers the widest range of industries, including IT, aviation, finance, and government. EDM DCAM, China DCMM, and Gartner EIM models are mainly oriented toward enterprises, including finance, power, IT, and other industries. The China Data Center Service Capability Maturity Model applies to enterprise internal data centers and commercial data centers (IDC, cloud computing centers, Internet enterprise data centers), but is primarily oriented toward banks. The Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) and the Research Data Management Capability Maturity Model (Australian ANDS) are mainly applicable to personnel responsible for providing research data infrastructure within university institutions (such as management, researchers, and libraries).

In terms of comprehensiveness of evaluation dimension (functional area) configuration, CMMI-DMM, DCAM, and China DCMM show certain similarities in their evaluation dimension configuration, basically following the basic elements constituting data management—strategy and policy, organization and personnel, and technology and tools (platforms and architecture)—to select and organize evaluation dimensions, with relatively comprehensive functional areas covering the five basic domains of data strategy, data governance, data quality, data platforms and architecture, and data security. Particularly, China DCMM not only absorbs the industry-recognized components but also combines the actual situation of domestic data development by adding three independent capability items: "data standards," "data security," and "data application,"

making its evaluation dimensions (functional areas) practically valuable for reference. The Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) and the Research Data Management Capability Maturity Model (Australian ANDS) have relatively simple evaluation dimensions (functional areas).

In terms of detail of evaluation standards and rules, CMMI-DMM, DCAM, China DCMM, and the China Data Center Service Capability Maturity Model not only set evaluation standards for each maturity level from an overall perspective but also provide corresponding evaluation standards for each major category and its subcategories at each maturity level from a micro perspective.

From an operability perspective, the China Data Center Service Capability Maturity Model has strong operability, with its standard document providing detailed evaluation methods and steps. Additionally, its quantitative and qualitative combined evaluation method is worth learning from. The Research Data Management Capability Maturity Model (Australian ANDS) also has strong operability due to its relatively simple evaluation dimension (functional area) configuration.

Common problems across models include: (1) The boundaries between maturity levels are not clearly defined and have certain ambiguity, making it difficult to distinguish the maturity level. Although many models have developed detailed evaluation standards and rules for each evaluation dimension at each maturity level, there is generally a lack of substantive and clear distinguishing standards between levels within the same evaluation dimension. (2) Lack of guidance on the assessment process and application guidelines, including specific assessment steps, methods, and case studies, resulting in many models being developed but not substantially applied. There is an urgent need to provide application guidelines describing the assessment process, methods, and maturity level calculation methods to further promote model implementation.

---

5. Recommendations and Implications for Libraries Selecting and Applying Data Management Capability Maturity Models

Through comparative analysis of existing data management capability maturity models, we have identified models with relatively comprehensive evaluation dimensions (functional areas), detailed evaluation standards and rules, and good operability. Libraries can select one or a combination of several models to conduct data management capability assessments. Based on the comparison and evaluation results of each model, we have formed three solutions adapted to different contexts (see Table 5).

Since the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) and the Research Data Management Capability Maturity Model (Australian ANDS) have already considered the actual situa-

tion of data management in libraries during their development, focusing more on data management service areas that libraries can provide, such as metadata and storage facilities, libraries can use these two models as the foundation for assessment.

Scheme 1 uses the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) as the main model, applicable to institutions that focus on data lifecycle management or have complete data lifecycle management. Considering that this model is primarily qualitative, it can be combined with the China Data Center Service Capability Maturity Model, utilizing its quantitative evaluation method to form a comprehensive assessment method combining qualitative and quantitative approaches. Additionally, the Research Data Management Capability Maturity Model (Syracuse University, Prof. Qin Jian's team) is not perfect in certain data management functional areas, such as lacking dedicated evaluation dimensions for "data strategy," "data platforms and architecture," "data quality," and "data security," and can be combined with CMMI-DMM and China DCMM models to expand data management domains.

Scheme 2 uses the Research Data Management Capability Maturity Model (Australian ANDS) as the main model. Its evaluation dimensions (functional areas) and evaluation standards and rules are very concise and easy to operate, applicable to institutions that are just starting data management or have relatively simple data management processes. This model can also be combined with the China Data Center Service Capability Maturity Model, utilizing its quantitative evaluation method to conduct assessment using a qualitative and quantitative combined approach.

Scheme 3 is applicable to institutions that only evaluate specific aspects of data management, such as only assessing data strategy, data quality, data security, data application, or other individual aspects. Institutions can select one or several models from the seven models that involve the aspect being evaluated and conduct assessments by integrating the evaluation rules of each model for that data management aspect.

Data management capability maturity models are important tools for initiating, maintaining, and supporting data management. By analyzing existing data management capability maturity models, we find that each model has its applicable scope, advantages, and disadvantages. Selecting a model during the data management initiation or planning phase and using it to regularly evaluate data management activities is conducive to promoting the achievement and development of organizational data management objectives. This paper analyzes and compares important data management capability maturity models both domestically and internationally, and provides preliminary solutions for how libraries can select models, hoping to be helpful for libraries conducting data management capability maturity assessments.

References

[1] Yang Helin. Data Curation: New Exploration of University Libraries in the United States[J]. Journal of Academic Libraries, 2011(2): 18-21, 41.
[2] CMMI Product Team. CMMI for Development Version 1.3[EB/OL]. [2019-10-15]. https://resources.sei.cmu.edu/asset_{files}/TechnicalReport/2010_{{{005}}*{{001}}*}{15287}.pdf.
[3] Ding Ning, Ma Haoqin. Comparative Study of Foreign University Scientific Data Lifecycle Management Models[J]. Library and Information Service, 2013, 57(6): 18-22.
[4] Li Bing, Bin Junzhi. Data Management Capability Maturity Model[J]. Big Data, 2017(4): 29-36.
[5] Zhang Kai. Research and Application of Data Center Service Capability Maturity Model[J]. Financial Computerizing, 2015(9): 56-59.
[6] Wang Zhipeng, Zhang Can. Research on International Technical Report of Data Center Service Capability Maturity Model[J]. Information Technology & Standardization, 2017(6): 62-65, 73.
[7] Ye Lan. Analysis of Research Data Management Capability Maturity Model[J]. Document, Information & Knowledge, 2015(2): 115-123.
[8] Li Weimian. Research on Research Data Management Service Evaluation Based on Lifecycle Theory[D]. Beijing: Beijing Institute of Technology, 2016.
[9] Dang Hongli, Tan Haibing. Data Management Maturity Model Based on DMM and Its Application in Service Evaluation[J]. Modern Information, 2017, 37(9): 118-121.
[10] Peng G, Privette J L, Kearns E J, et al. A unified framework for measuring stewardship practices applied to digital environmental datasets[J]. Data Science Journal, 2015, 13: 231-253.
[11] Zhang Yujie, An Xiaomi, Zhang Guoqing. Construction of Maturity Evaluation Index System for Government Big Data Governance[J]. Information and Documentation Services, 2018(1): 28-32.
[12] Liu Guifeng, Qian Jinlin, Lu Zhangping. Comparison of Foreign Data Governance Models[J]. Library Tribune, 2018(11): 18-26.
[13] CMMI Institute. Data Management Maturity (DMM)[EB/OL]. [2019-09-25]. https://www.cmmiinstitute.com/data-management-maturity.
[14] EDM Council. DCAM: the data management capability assessment model[EB/OL]. [2019-09-25]. https://edmcouncil.org/page/aboutdcamreview.
[15] The Gartner enterprise information management framework[EB/OL]. [2019-10-04]. https://blogs.gartner.com/andrew_{white}/files/2016/10/On_{{{site}}}_{{poster}}}.pdf.
[16] Qin J, Crowston K, Kirkland A. A capability maturity model for research data management[EB/OL]. [2019-10-17]. http://rdm.ischool.syr.edu/xwiki/bin/view/CMM+for+RDM/WebH
[17] ANDS. Creating a data management framework[EB/OL]. [2019-10-17]. http://www.ands.org.au/guides/creating-a-data-management-framework.
[18] Data Management maturity models: a comparative analysis[EB/OL]. [2019-10-04]. https://datacrossroads.nl/2018/12/16/data-management-maturity-models-a-comparative-analysis/.
[19] DMM model at a glance-Chinese Language (Digital Version)[EB/OL]. [2019-09-30]. https://cmmiinstitute.com/getattachment/eb414209-110e-491b-

b45c-c5b2fa6a5d54/attachment.aspx.
[20] Han Yi, Wang Yinyu, Yang Wei. Comparison of Classic Data Management Capability Maturity Models and Strategic Application Route Suggestions for Control-oriented Groups[J]. Journal of Shanghai Ship and Shipping Research Institute, 2018, 41(3): 72-79.

Note: Figure translations are in progress. See original paper for figures.