



北京理工大学

研究数据管理的生命周期理论研究

A decorative graphic consisting of four grey chevrons pointing to the right, arranged in a horizontal line.

北京理工大学

报告人：李伟绵

指导老师：崔宇红

2014年12月19日



A. 研究意义

B. 生命周期模型归纳

C. SDM-CMM

D. 研究计划



研究意义

在研究数据管理方面，常常运用生命周期模型支持数据服务。研究数据管理过程往往是复杂的，涉及到人与人，机构，资源和活动之间的协调。

1 研究数据管理

- 识别研究数据的不同阶段
- 识别研究人员的特定需求
- 识别提供数据服务时的角色和责任

2 数据服务

- 形成一个较大的连续体
- 识别单个服务的潜在差距

3 服务提供商

- 传达服务给预期的消费者
- 确定和发展数据业务
- 服务及所需的条件或对消费者的要求之间的关系





A. 研究意义

B. 生命周期模型归纳

C. SDM-CMM

D. 研究计划



常见的8种生命周期模型（1）

2012年国外学者Alexander Ball总结了当前常见的8种数据管理生命周期模型，详情如下图所示：

模型名称	基本内容	提出机构
DCC管理生命周期模型	核心:数据;第一层:描述和（管理）呈现信息;第二层:保存规划;第三层:团体活动参与;第四层:数据管理和长期保存;最外层活动包括:概念化→创建或接收数据→评估和选择→存储→获取\再利用→改变	英国JISC和e-science核心计划
I2S2理想化科研活动生命周期模型	基本阶段包括：研究提出的发展;同行审查;实验执行;处理，分析和对数据的解释，最终以各种形式报告和公布的研究成果。 除此之外，还包含一些理想化的阶段，包括：评估和质量控制;元数据;存储，归档，保存和管理，知识产权等	英国JISC
DDI结合的生命周期模型	研究概念→数据收集→数据处理→数据存档→数据发布→数据发现→数据分析→数据再利用	DDI联盟

常见的8种生命周期模型（2）

模型名称	基本内容	提出机构
ANDS 数据共享动词	①创建；②存储；③描述；④识别；⑤注册；⑥发现；⑦获取；⑧开发	澳大利亚国家数据服务（ANDS）
DateONE 数据生命周期模型	①计划；②收集；③保证；④描述；⑤保存；⑥发现；⑦整合；⑧分析	新墨西哥大学图书馆；国家生态分析合成中心；
英国数据存档 (UKDA) 数据生命周期模型	①数据创建；②数据加工；③分析数据；④保存数据；⑤数据访问；⑥数据再利用	英国埃塞克斯大学
Research360 机构研究生命周期模型	①计划和设计；②收集和获取；③解读和分析；④管理和保存；⑤发布和出版；⑥探索和再利用	巴斯大学
科学数据管理的能力成熟度模型	①数据获取，加工和质量保证；②数据描述和表示；③数据发布；④知识库服务/保存	雪城大学Jian Qin 和Kevin Crowston



小结

1

学科领域差异

如：DataONE数据生命周期模型是从生态学角度提出的；

DDI结合生命周期模型、英国数据存档(UKDA)数据生命周期模型是从人文、社会科学的角度的提出的；

科学数据管理的能力成熟度模型是从“小科学”角度提出的。

人文社会科学数据的研究方法具有一定的通用性，且该类研究数据量较小，数量载体较为一致，可用一些简单的方法、策略即可完成管理。

“小科学”研究具有数据来源、结构复杂，缺乏通用性等特点，故针对“小科学”学科尚未建立明确的研究数据管理政策，而且在“小科学”的实践较少。



小结

2 适用性较低

研究中心等机构的研究领域比较专一，所以其模型都是符合特定领域的科学数据的生命周期，对其他学科数据适用性较低。

3 均涉及通用阶段

如：数据收集、数据处理、数据发布

Alexander Ball认为，Kevin Crowston和Jian Qin将软件工程的能力成熟度模型引入科研数据管理中，可能是实践中最有效的模型。下面对其进行重点介绍。



目录

CONTENTS

- A. 研究意义
- B. 生命周期模型归纳
- C. SDM-CMM
- D. 研究计划

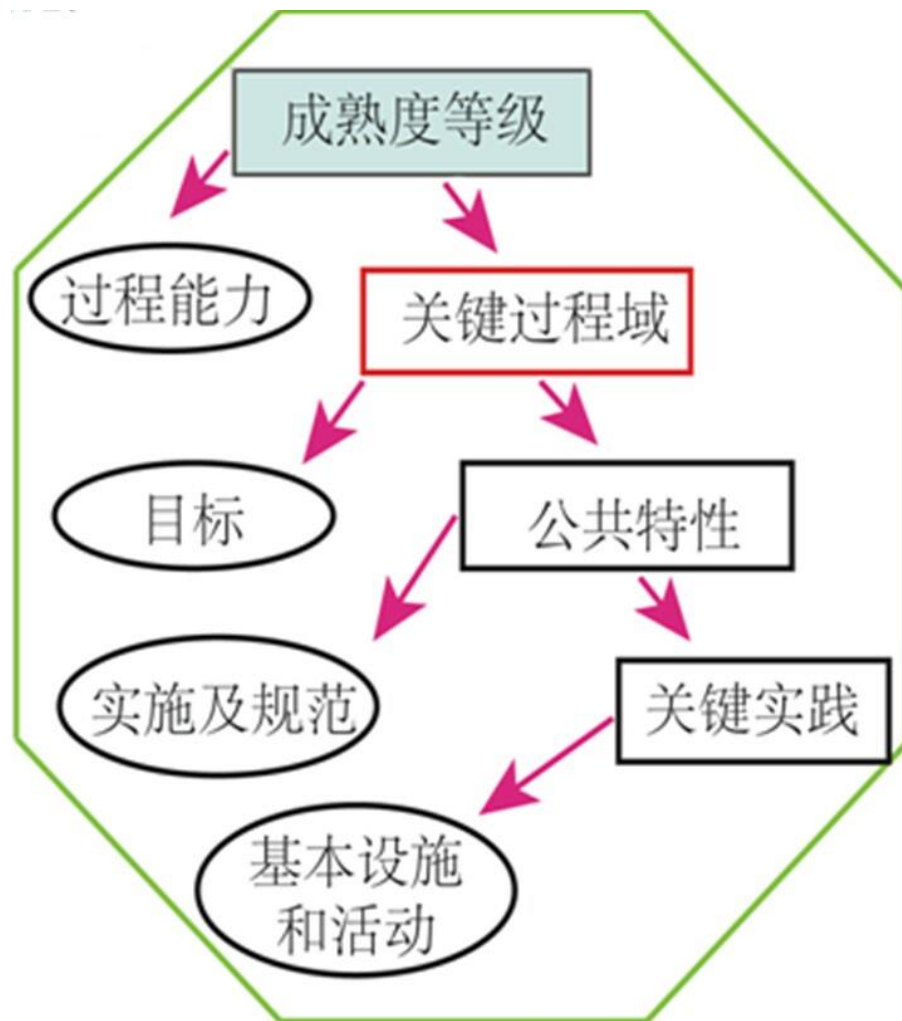


能力成熟度模型 (CMM)

背景

最初的能力成熟度模型 (CMM) 是卡内基梅隆大学软件工程研究所 (SEI) 支持改善软件开发的组织可靠性而开发的。旨在帮助开发者通过确定他们当前的过程成熟度，识别最关键的问题，来选择过程改进策略，从而提高软件质量和过程。

SEI-CMM 内部结构 (如右图所示)



SDM-CMM

关键过程域1：数据采集、加工、质量保证

- **关键实践**：获取数据；内容准备数据进行存储、分析和分配；保证数据质量

关键过程域2：数据描述和表示

- **关键实践**：开发和应用元数据规范和模式；创建数据集的描述和语义元数据等

关键过程域3：数据传播

- **关键实践**：识别和管理数据产品；鼓励分享；分发数据等

关键过程域4：知识库服务和保存

- **关键实践**：存储、备份和保护数据；执行数据迁移；构建数字保存网络等

1

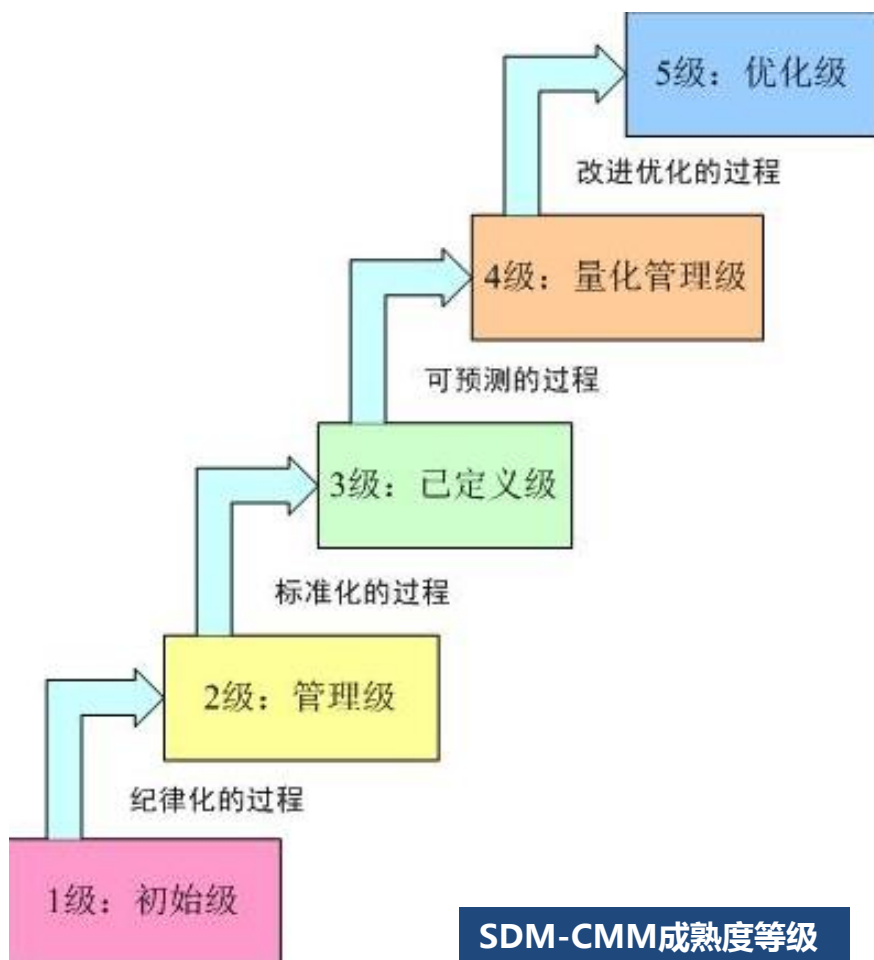
2

3

4



SDM-CMM成熟度等级（1）



1级：初始级

- CMM的初始水平描述一个没有定义和稳定的过程的组织。

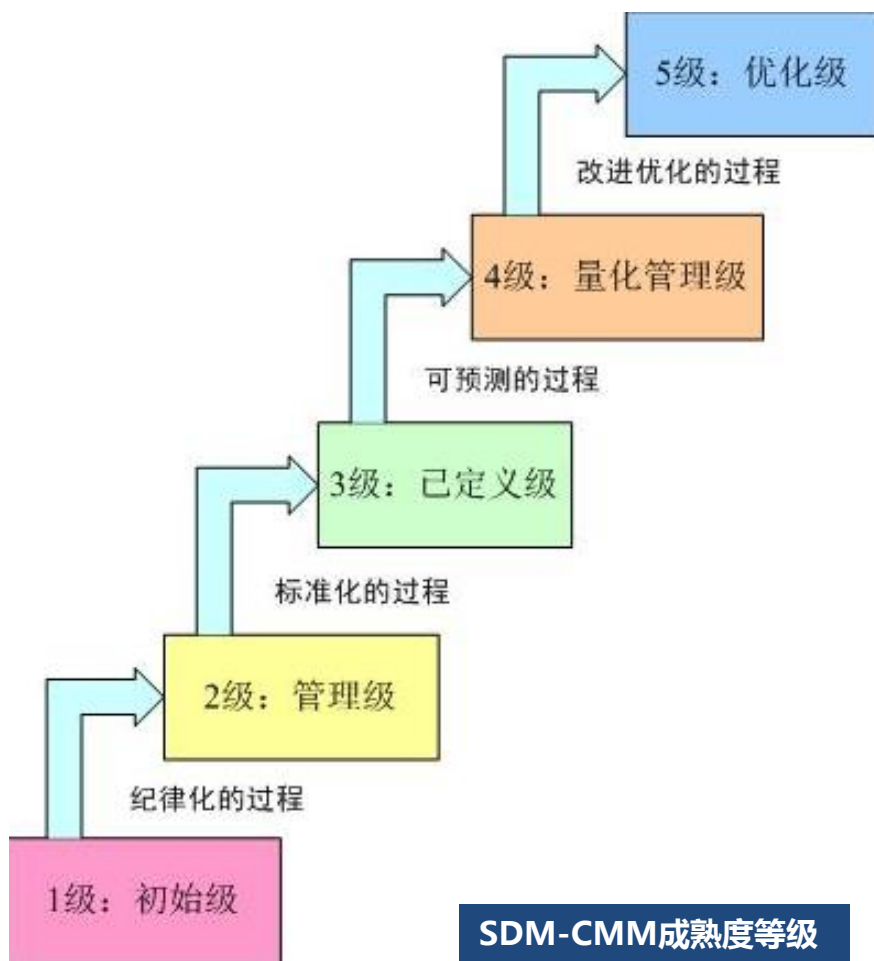
2级：管理级

- 项目中项目管理过程是通过政策和程序建立

3级：已定义级

- 整个组织过程记录,针对特定的项目定制并应用。

SDM-CMM成熟度等级（2）



4级：量化管理级

- 已为产品和过程量化质量目标。过程执行和数据是系统地收集和分析评估过程。

5级：优化级

- 组织致力于过程改进:缺陷识别和弊端主动处理。

SDM-CMM

该模型为实施和管理数据管理实践提供了指导，但仍旧处于初步阶段，存在以下不足之处：

- 在选择文章时采用“立意抽样”方法，该类抽样结果受研究人员的倾向性影响大，一旦主观判断偏差，则容易引起抽样偏差；不能直接对调查总体进行推断。
- 模型中描述的关键实践和关键过程域处于初步状态，有待对这些描述进一步详尽和扩展。
- 模型的重点是数据集的创建，缺少描述任何相关数据发现或重用的实践。





目录

CONTENTS

A. 研究意义

B. 生命周期模型归纳

C. SDM-CMM

D. 研究计划



研究计划



CCM-RDM体系

1. 提出CCM - RDM
2. 建立CCM-RDM测度和评价指标体系

RDMS实践

1. 调研当前高校RDMS实践
2. 开展对当前实践的定量测度和定性评估，帮助机构指定政策和路线图



Thanks!

