

Towards an Enhanced Data- and Knowledge Management Capability: A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

Alexslis Maindze¹, Zakwan Skaf², and Ian Jennions³

^{1,2,3}*Cranfield University, Integrated Vehicle Health Management (IVHM) Centre, College Road, Cranfield, Bedfordshire MK43 0AL, United Kingdom*

alexslis.maindze@cranfield.ac.uk

z.skaf@cranfield.ac.uk

i.jennions@cranfield.ac.uk

ABSTRACT

The creation, capturing, using and sharing of knowledge is based on data. The rate of data creation, collection, and elicitation through wide ranging experiments, simulations, observations and measurements is rapidly increasing within Integrated Vehicle Health Management (IVHM). In addition, Knowledge Management (KM), data abstraction, analyses, storage and accessibility challenges persist, resulting in loss of knowledge and increased costs. This growth in the creation of research data, algorithms, technical papers, reports and logs, requires both a strategy and tool to address these challenges. A Data Life Cycle Model (DLCM) ensures the efficient and effective abstraction and management of both data- and knowledge outputs. IVHM is characterized by prognostics and diagnostics, which depend heavily on high quality data to perform data-driven, model-based and hybrid computational analysis of asset health. IVHM does not yet have a systematic and coherent approach to its data management. The absence of a DLCM means that valuable knowledge might be lost or is difficult to find. Data visualization is fragmented and done on a project by project basis leading to increased costs. There is insufficient algorithm documentation and communication for easy transition between subsequent researchers and personnel. A systematic review of DLCMs, frameworks, standards and process models pertaining to data- and KM in the context of IVHM, found that there is no DLCM that is consistent with IVHM data- and knowledge management requirements. Specifically, there is a need to develop a DLCM based on Open System Architecture for Condition-Based Maintenance (OSA-CBM) framework.

Alexslis Maindze et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

There has been a steady growth in both scope of research as well as data extraction activities in IVHM depicted in Figure 1 below. This growth has been matched by the complexity of management and organization. With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses, IVHM needs both a data management model and a knowledge management system that facilitate the storing, organizing and sharing of its research and knowledge outputs. Such a model needs to be secure and scalable with a high level of cross-platform or domain transferability. The absence of a suitable data life cycle model means weak data- and knowledge management for IVHM. This leads to increased costs and loss of valuable knowledge thereby creating long-term uncertainties for diagnostic and prognostic management. The primary value for designing and implementing a data life cycle model includes the following:

- Enhanced and integrated requirements gathering for IVHM and IVHM data and knowledge management systems
- Increased efficiency and effectiveness of planning and handling of the growing volumes, diversity and complexity of data and data management
- Facilitate the design and development systems for high operational efficiency
- Making raw and derived data accessible to IVHM researcher and engineering operations
- Ensures the provision of secure, high quality, accurate and consistent asset data throughout its entire life-cycle
- Facilitates the retention of provenance data
- Ensuring timely, comprehensive, and secure approaches to data curation (Faundeen et al. 2013)

In this paper we carry out a non-exhaustive integrated and systematic review. We finally propose an integration of relevant elements of the Open System Architecture for Condition-Based Maintenance (OSA-CBM) framework, the International Council on Systems Engineering (INCOSE) process model, the Core Scientific Metadata (CSMD) model and United States Geological Survey (USGS) data life cycle model. The aim is to create a Data life cycle model suitable for the management of IVHM data, knowledge outputs, depth and breadth of IVHM operations, research and IVHM system requirements.

The basic foundation of Integrated Vehicle Health Management (IVHM) consists of sensing, instrumentation and signal processing. This leads to the extraction of data and features selection, paving the way for prognostics prediction algorithms. Precision and well-timed availability of the above instruments are fundamental to maintenance scheduling (Perinpanayagam 2013). The first stage of the IVHM cycle is the collection of data using simulations, observations, derivation, experiments and referencing (SODER) methods (Figure 1) about an asset. IVHM delivers value to stakeholders and reduces the cost of delivery by monitoring the health of an asset and making decisions based on the data collected (Jennions 2011). Consequently, IVHM relies predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health.

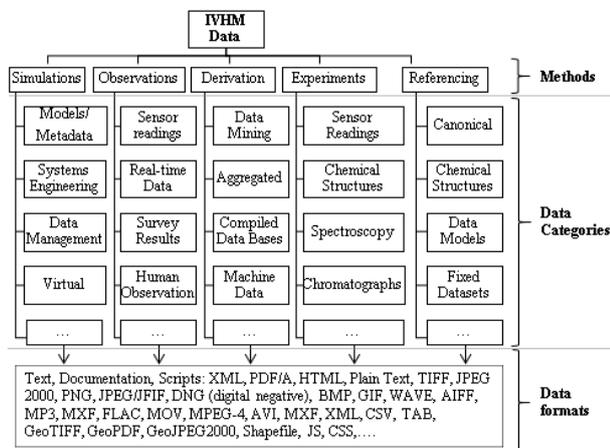


Figure 1. Data creation methods, data categories and file formats.

The data has to be accurate, complete, timely, context relevant, reliable and explicit (Dibsdale 2011). The absence of any or all of these qualities in any dataset creates uncertainties and increases the probability of misdiagnosis, modelling errors (Arahchige and Perinpanayagam 2017) and inaccurate predictions. This essentially indicates the malfunction of the IVHM system because it is “*the assembly of data related to the current and future activities of a*

critical system and transforms these data into the information and which is applied to make a functional decision” (Prajapati, Roy, and Prasad 2018:1). A data life cycle model can be defined as a “*...a formal representation of all the possible states and all the valid state transitions of a data item, when handled by a particular system or by a user application, e.g. created, duplicated, deleted, backed-up*” (Simonet et al., 2015:26). It represents the requisite actions, operations, or processes to be taken at various stages of data creation and management (Faundeen et al. 2013). Data management through its entire life cycle still presents a number of complex challenges relating to interoperability, volume, storage, data citation, and metadata standards and data provenance (Porcal-Gonzalo 2015; Beaujardière 2016; Yang, Matthews, and Wilson 2013). This perhaps explains the proliferation of discipline or domain-specific data life cycle models.

Various disciplines and organizations are creating standardized frameworks, data ontologies, standards and unique data life cycle models to suit their respective requirements. Metadata standards like Dublin core (Hsu et al. 2015), Core Scientific Metadata (CSMD) provisions the basic metadata required to enhance the search functionalities over data portals and knowledge libraries (Matthews et al. 2010), but falls short of propagating the complete provenance data. The shortfall with the CSMD is that neither does it support for “*access to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data*” as pointed out by Yang [64, p.613].

However, though metadata standards are relevant for all data and knowledge outputs and organizations, each organization or project seems to have separate requirements for their own research or projects data. This is reflected in the 17 data life cycle models identified, which all highlight the significance of metadata and standards. The models seem to differ on the depth and breadth of applicability and priority of requirements. For instance the USGS emphasized three critical cross-cutting activities namely; metadata description, quality assurance and protection from corruption or loss to be performed parallel to planning, acquisition, processing, analysis, preservation, publishing and sharing to achieve enhanced quality, understanding and long-term reuse (Plale and Kouper 2017; Faundeen and Hutchison 2017). The Digital Curation Centre (DCC) model, on the other, hand suggests metadata should comprise rules and formalized entities for automatable tools and services as well as the role of data managers and data curators in the improvement of knowledge (Plale and Kouper 2017).

Taking the DataOne data life cycle model to illustrate the variance, though it represents all the classical stages of the data management life cycle, all parts of the life cycle are not

mandatory (Plale and Kouper 2017; Allard 2014; DataONE 2017; Harrison 2013; Hidalgo et al. 2017; Pouchard 2015). This implies that the number of selected stages are dependent on the type of project or project requirements. Furthermore, although the USGS addresses the weakness in most data life cycle models identified, it can be used in diverse settings despite having been developed mainly for USGS science data. The requirements on which this is developed are not consistent with small research facilities like the IVHM Centre Labs or the University as a whole. None of the models identified seems to satisfy the scenario of IVHM research, operations and systems.

The core of IVHM is the capture of data and analysis to establish an early detection of anomalies and an advance indication of future failures, state or distinctive characteristics of current assets based on their current state. The conduct of prognostics and diagnostics is reliant on high-quality reliable data. This quality assurance can be determined by the data management life cycle. Data life cycle models are often integrated with software services and policies (Plale and Kouper 2017). It is therefore fundamental to consider associated processes involved in the planning, designing and developing of the software services and policies.

The INCOSE system engineering process [V-Model] is suitable for the development of any system. It is vital in the scenario of IVHM data life cycle model and knowledge management system. This is suitable for the IVHM data life cycle and system development because of the restricted duration and scope of this research project. The V-model is best suited for these kinds of projects that have well-defined length and scope, consistent technology, and a clear and well documented technical and functional specifications (INCOSE 2017).

Below is the presentation of our review and analyses leading to the IVHM Data life cycle model.

1.1. Methodology: The Systematic Review

Following a scoping study, the search string depicted in Figure 2 was implemented. As a consequence of the diverse and qualitative studies reviewed, a narrative synthesis was applied. A narrative synthesis is “an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis.”(ten Ham-Baloyi and Jordan 2016:5). This study identified 19 data life cycle models with varying degrees of complexity, composition and depth (Figure 3). According to Pouchard [2015, p.180] “Data life cycle models present a structure for organizing the tasks and activities related to the management of data within a project or an organization. ” Data lifecycle models represent a description of “data objects through a set of time ordered stages” (Plale and

Kouper 2017:95). The data life cycle models vary in steps or phases from one organization to another.

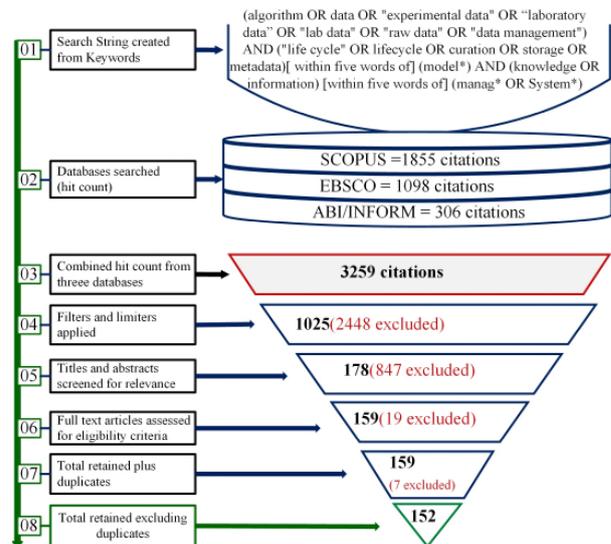


Figure 2: Scoping study and search string implementation

The longest model has nine phases, and the shortest has five phases. However, these data life cycles are not necessarily cyclical but rather functional as some phases run parallel across the entire life cycle.

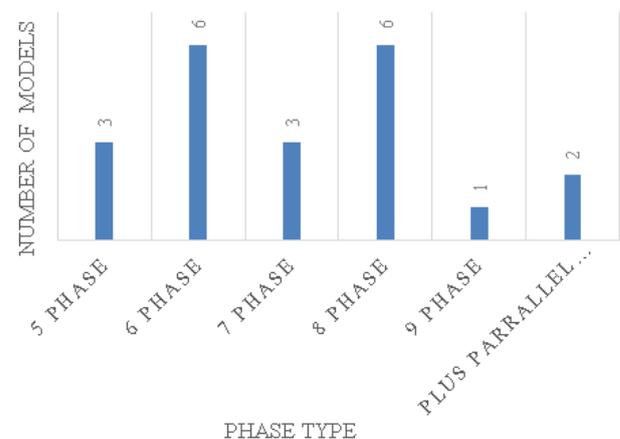


Figure 3: Data life cycle models by phases

The data life cycle models were then grouped by their respective numbers of data life cycles, i.e. five-phase, six-phase, seven-phase, eight-phase and nine-phase models. The review revealed nine standards and frameworks relating to data life cycles. In the reviewed studies, the data life cycle has been described as the set of activities that affect the short and the long-term preservation of datasets through a system from planning, creation, maintenance, re-use and purging (Simonet, Fedak, and Ripeanu 2015;Beaujardière 2016).

1.1.1. Five-phase data lifecycle models

Three five phase models were identified; CRUD (Create, Read, Update and Delete), Enterprise Data life cycle model and Michigan State University (MSU) Records life cycle. These models make use of the classical elements of the data life cycle. The CRUD model provides a flexible data life cycle; it only obliges creation, storage and destruction while living use, sharing and archiving optional. The Enterprise Data life cycle model is a closed life cycle, and its emphasis on destruction limits the availability of data in the long-term (Arass, Tikito, and Souissi 2017). The Michigan State University (MSU) life does not feature sharing, classification and analysis.

1.1.2. Six-phase data lifecycle models

Six data life cycle models with six phases were identified; (i) United States Geological Survey (USGS) Science Data Lifecycle Model, (ii) University of Virginia, Steps in the Data Life Cycle, (iii) International Leader in Data Stewardship (ICPSR) Data Lifecycle, (iv) UCSD Libraries Data Lifecycle, (v) Generic Lifecycle Model, and UK Data Archive Data Lifecycle.

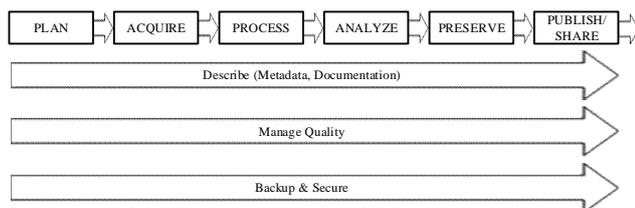


Figure 4 : USGS Science Data Lifecycle Model (Faundeen et al. 2013)

According to Faundeen and Hutchison (2017), data lifecycle models are fundamental to communication and data management and ensures adequate long-term preservation and accessibility. After reviewing more than 50 data lifecycle models [the latter four above inclusive], they came to the conclusion that none of the existing models was entirely consistent with the USGS data management requirements. It was imperative to USGS that their functional processes and workflows were adequately captured in a model.

Furthermore, like other organizations, the USGS developed its own data management lifecycle with the aim of reducing complexity and removing redundant or irrelevant steps or phases that were not in sync with their scientific workflows and processes. The USGS opted for a linear and easily operated illustration of their new model (Figure 4). The model (Faundeen et al. 2013; Faundeen and Hutchison 2017) included the basic classical data lifecycle phases and laid emphasis on three parallel phases; metadata, quality management, backup and Security.

1. Plan: The organization should identify the resources, methods, techniques, functional and technical system requirements and generate both plans for either data acquisition, data entry or signal reception and data management.
2. Acquire: This is the data capture phase which can either data acquisition, data entry, signal reception or all three activities combined.
3. Process: Raw as well as derived data verification, organization, transformation, integration, and extraction in appropriate format takes place in this step.
4. Analyze: This encompasses demonstrable quality requirements fulfilment, data analytics, modelling and evaluation test results as well as methods and activities carried out to facilitate definitions of facts, identification of forms and trends, developing interpretations and testing hypotheses.
5. Preserve: Data storage for Long-term access and reuse. The purpose of this phase is the guarantee long-term preservation, ease of search and retrieval, accessibility and usability of the data. This step employs multi-copy/storage locations, long-term usefulness, accuracy and consistency, information security, metadata and file formats.
6. Publish/Share: Put together quality assured, metadata rich, platform or system-agnostic data, with relevant security safeguards and share with interested parties of stakeholders.
7. [Parallel to phases 1-6] Describe (Metadata, Documentation): Establishes an obligation to create and upgrade metadata on any or all the stages of the lifecycle including the documentation of usage in specific systems, applications and settings.
8. [Parallel to phases 1-6] Manage Quality: Mandatory to accurately undertake data collection, handling, processing, usage, and maintenance across all the phases of the scientific data lifecycle, is the use of protocols and methods. This implies effective and efficient quality assurance and quality control.
9. [Parallel to phases 1-6] Back Up and Secure: Regularly create image backups of both files and data bases on either onsite or offsite devices. Access control and other security measures must be taken to prevent accidental data loss and data corruption.

The USGS data lifecycle model encapsulates the activities and steps in the latter four models. However different research activities or projects will use some or all elements of the data lifecycle in dissimilar ways. The data lifecycle management is influenced by the requirements of a particular project or organization (Faundeen and Hutchison 2017).

1.1.3. Seven-phase data lifecycle models

The key feature of these models is the inclusion of discovery, knowledge repository and reuse. There is a noticeable absence of access control and security element. The Geospatial Data Lifecycle is a flexible non-linear and non-sequential comprising the classical Define, Evaluate, Obtain, Access, Maintain, Use/Evaluate and Archive stages but lacks the discovery and reusability elements. These may not be weaknesses in themselves as each life cycle model has been designed to cater for to the needs of their respective organizations. As noted in Yu and Wen(2010) not all elements of the data lifecycle are applicable to all contexts and scenarios.

1.1.4. Eight-phase data lifecycle models

The DataOne data lifecycle model is one of eight phase data lifecycle models designed by the National Science Foundation (NSF) exclusively to focus on the phases field, or laboratory data goes through rather than the role of the person on data(Allard 2014). The model is developed as a foundation for the development of tools and services for data-intensive sciences and for encouraging best practices (Pouchard 2015; Arass et al. 2017; Plale and Kouper 2017; DataONE 2017).

The experimental geomorphology data life cycle model addresses the complexity in analysis, storage and search and retrieval posed by increasing volumes of laboratory data(Hsu et al. 2015). Hsu et al., (2015) review data management practices and challenges for experimental geomorphology, established that the lack of; rules for metadata integration, documentation of workflow, data storage, motivation and training were impeding significant amounts of data sharing and re-use. These challenges mean that accessibility or availability of experimental or laboratory data is limited to the research or project lifecycles, and therefore undermining reproducibility and quality control. They suggested that efficient and effective data management and sharing should consider the entire data lifecycle, including metadata set information, disciplinary information, Quality information and readiness for reuse.

The Data Documentation Initiative (DDI) version 3.0 mostly linear combined lifecycle model integrates data application perspective and social science research data. Metadata requirements are grouped into five comprising conception of the study, collecting of data, logical data encoding structure, physical data encoding structure and archiving.(Ball 2012)

The DCC curation lifecycle model (Faunden and Hutchison 2017; Ball 2012; DCC 2008) is the model with the highest number of phases with emphasis on preservation, archiving and management of data and publications for long-term availability and reuse. It

concentrates on data curation and serves as a planning tool for data creators, curators and users. The model features a complete lifecycle action, sequential actions and occasional situational actions (Plale and Kouper 2017). The model presents indispensable phases in the life cycle by using circles with a common center. The phases include: These are: (i) Conceptualize (Conceive & Plan); (ii) Create or Receive (Generate metadata); (iii) Appraise and select (Quality and Governance); (iv) Ingest (Move to Storage or Archive); (v) Preservation action (Apply retention schedules); (vi) Store (Secure Storage); (vii) Access, use and reuse (Access policy and control); (viii) Transform (Migration), plus (ix) Sporadic Actions: Migrate, Dispose, Reappraise and Dispose.

The DCC lifecycle model provides an applied framework that can be loosely categorized into different levels and areas of curation such as technology level operations, bit-level preservation routines and metadata curation. The technology level operations involve migration, backup, indexing, and system upgrades. Bit-Level preservation routines include data-recording, checksum reporting, error-correcting and character replacement and file format registries. The metadata curation level defines the content and context of digital elements. This level fulfils the technical, descriptive, structural and preservation metadata requirements(Parry 2016; Sabharwal 2015). One of many technical standards for metadata is the Dublin Core. Activities that generate rich footnotes and significance to images in Digital humanities scholarship and teaching practices are given three levels of Curation.

Level one (L1) focuses on digital files and technology used in their preservation while levels two and three denotes the exercise of intellectual control and scholarly processing respectively. Furthermore, underlying these levels are aspects of metadata schema such as Dublin Core, interoperability standards like the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), file format and data encoding, network characteristics and reliable hardware and software systems(Sabharwal 2015).

1.1.5. Standards and frameworks

In addition to some of the standard mentioned above, Figure 5 below shows four main standards commonly used to support data management through the data life cycle. The design of data lifecycle models takes into consideration relevant and appropriate standards/frameworks to support respective domain data. Some relevant standards identified in the literature included the Open System Architecture for Condition Based Maintenance (OSA-CBM), the Core Scientific Metadata Model (CSMD), INCOSE systems engineering Management process (the V-model) and the National Oceanic and Atmospheric Administration (NOAA) Environmental Data Management (EDM) Framework. Each plays an important role in the data management lifecycle.

The Open System Architecture for Condition Based Maintenance (OSA-CBM) is an ISO-13374 compliant (Erkki Jantunen et al. 2017; Felke et al. 2010) “*standard architecture for moving information in a condition-based maintenance system*”(MIMOSA 2018). The ISO 13374 provides the standards for condition monitoring and diagnostics of machines and machine systems – Data processing, communication and presentation. The OSA-CBM framework (Figure 5) facilitates the integration different software and hardware components in order to decrease costs, enhance interoperability, boost competition, merge design changes, and stimulate collaboration in condition-based maintenance (Löhr and Buderath 2014; Sreenuch, Tsourdos, and Jennions 2013). According to (Choudhary, Perinpanayagam, and Butans 2016) OSA-CBM is fundamental in the creation data identities [‘data CVs’] in the form of metadata that includes attributes such as id, site, time, alert, algorithm type, name, description and others.

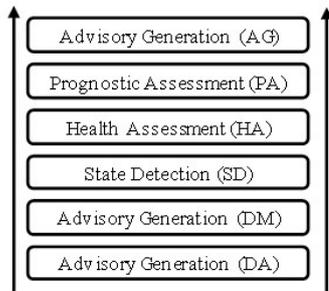


Figure 5: OSA-CBM Functional Blocks(MIMOSA 2018; E Jantunen, Junnola, and Gorostegui 2017)

Its functional capability includes human-machine interface provisioning information accessibility. Its focus on interoperability and metadata makes its relevance in the data lifecycle models extremely pertinent. Figure 5 above displays the basic Architecture of OSA-CBM and the following paragraphs explain the data flow between the layers;

Data Acquisition (DA): This block uses either transducers or sensors to pick up physical manifestations and convert to clean digital signals for computerization to extract relevant information. The DA block is essentially a server of cleaned digital signal data.

Data Manipulation (DM): The DM block processes raw data from the DA block by means of mathematical algorithms, generating computed and the virtual sensor readings which are stored in a database.

State Detection (SD): In this block, the resulting data from the DA and DM blocks are compared to known profiles for any discrepancies, and if so, identify the profile associated with the data.

Health Assessment (HA): While taking possible faults into consideration, historical health trends, functional status and load history are analyzed to diagnose faults and current health situation.

Prognostic Assessment (PA): Using current data and projected usage, current health conditions of assets and the remaining useful life (RUL) are forecast using either model-based (physical phenomena of degradation), data-driven (pattern recognition and machine learning algorithms) or hybrid approach (a combination of statistical data and physical phenomena) to get the best outcome.

Advisory Generation (AG): This is the decision support layer that provides recommendations on steps and movements that need to be undertaken to optimize the useful health of the system in consideration.

Presentation Block (Human Interface): The presentation layer displays health valuations, prognostic valuations, or decision support recommendations and alerts (Lebold et al. 2002; Redding 2011).

The OSA-CBM framework provides the parameters for systems architecture required for the successful application of Integrated Vehicle Health Management (IVHM) to a product(Redding 2011). Thus, the structure of OSA-CBM architecture provisions the foundation for evaluating important IVHM technologies and database standards (Goebel 2011). This is significant because IVHM is a data-driven and data acquired from transducers and sensors directs much of the thinking surrounding it. In addition data management, data integrity, data quality are imperative for features or faulty conditions extraction-fundamental in condition-based maintenance (CBM) and Prognostic health management (Goebel 2011; Dibsdales 2011). IVHM systems require the capability to organize and manage small as well as large data sets in linked tables to promote the easeful appreciation and deliver a comprehensive language for data definition, retrieval, and update. Therefore, making data management an imperative competence in the operations room in particular and IVHM in general (Dibsdales 2011).

Unlike OSA-CBM the Core Scientific Metadata (CSMD) model (Figure 6) is designed for use in large scale laboratory-centered scientific facilities to represent data acquired from scientific experiments or structural sciences(Yang et al. 2013). The Science and Technology Facilities Council (STFC) generated the model. It is designed around the hierarchical concept of scientific studies and investigation which are usually characterized by experiments, measurements, simulations, modelling and observations. The outcomes typically include three phases – Phase One being the acquisition of raw data which is analyzed in Phase two to create derived data which is eventually published in phase three. The Core Scientific Metadata (CSMD) model mainly involves experimental data acquisition and partially automated creation of metadata.

The CSMD model supports interoperability of data management and accessibility, and facilitates cataloguing, data curation and data reuse for medium and to long-term(Matthews et al. 2010).

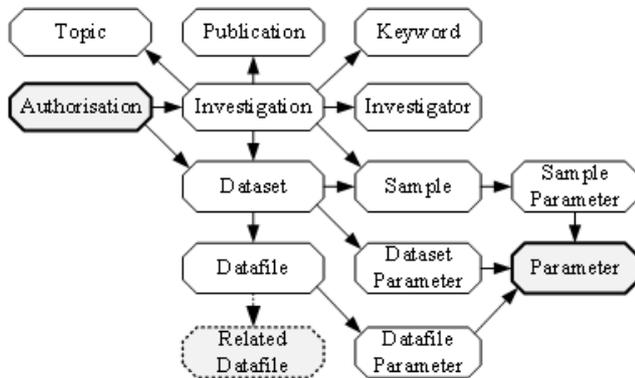


Figure 6: The CSMD model (Matthews et al. 2010)

The main elements of the CSMD model (Matthews et al. 2010) include:

Investigation: This is the most important entity of the study that specifies title, abstract, dates, data collection tools, facility and the unique identifiers referencing the particular model.

Investigator: Stakeholders of the study; Main researcher, support researcher, sponsors, institutions and their roles.

Topic and keyword: This include managed and unmanaged terms used in interpreting and cataloguing the research.

Publication: Assigns references to publications linked with the research.

Sample: Detailed data on study research sample. Unique details such as precautions on the toxicity of elements, and attributes relating to tagging of samples, substance annotation are captured in the model.

Dataset: Research projects can include more than one dataset on which diverse and multiple samples are analyzed. Research activity can include raw datasets on which analyzed datasets are subsequently inserted.

Datafiles: In the CSMD, this takes the form of physical data objects stored on physical storage disk drives(Yang et al. 2013). Its metadata includes name, version, location, data format, time created, modified by and time modified, and other details such as Check-sum.

Parameter: Defines explicit aggregates like pressure, temperature, volume or scattering angle connected to the research. These aggregates can be used to describe sample parameters, study environment or the variables being measured. Parameters can be linked at various levels with datasets and metadata elements.

Authorization: The CSMD model can specify access controls on investigation and datasets.

In a study of the problem of managing provenance of derived data in scientific research, Yang et al (2013) found that although the initial CSMD model provisioned accessibility, usage and reuse of experimental raw data, it did not "support access to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data"(Parry 2016:613). In other words, the original CSMD model recognizes the sources' provenance of the derivative data but fails to describe the transformation provenance. They emphasized the significance of keeping track of previous work and the need for a resilient data management tool and computational workflows that would capture the flow of data, raw data to derived data through to final publication.

They recommended that data trails generated during analysis should be captured to ensure reliable reproducibility research outcomes – an essential element in valid scientific research. Because much of the data in scientific facilities is generated in large volumes and at significant costs, repetition of data collection is not a viable or preferred option and therefore any bid to replicate results. Therefore, replication of test results would be best achieved through the re-analysis of already collected raw data.

To demonstrate the validity of their proposition, they extended the CSMD model to account for derived data and to record the data analysis process enough for each of their use case. They extended the CSMD model to a software agnostic one that contains resultant data product to include a description of transformation provenance which is not covered in the existing model. They carried a pilot implementation with experimental scientists, with annotations employing the ICAT data catalogue scheme. They identified five fundamental factors for capturing provenance data.

They include the following:

- the *data objects* involved
- the *programs* that produce or consume data objects
- the *ordering* of the programs
- the *parameters* to the programs; and
- the *people*: This refers to those who drive the programs and therefore fundamental for accountability, security, attribution and archival processes. This element is excluded from the extended model

They proposed six items as an extension to the CSMD model as follows:

1. Adding a *SoftwareExecution* subclass of investigation: The subclass is for modelling the executions of one data analysis task in the process.

2. Linking a program to a *software execution*: This is a runtime notion associated a single software program, one or more data files and zero to any number of parameters that drive the program and the outputs from the data files and parameters. The persistent and catalogued aspects of derived data provenance trail are often determined by researchers.
3. Linking software executions with datasets: Software

In conclusion, Yang et al., (2013) all observed that their proposed extended model was domain agnostic though developed for solving structural science data management problems from large scale facilities, it can be used to resolve issues of derived data management in many disciplines as well as in limited size scenarios like university research laboratories [such as the IVHM lab].

The extended model captures both the data source and its

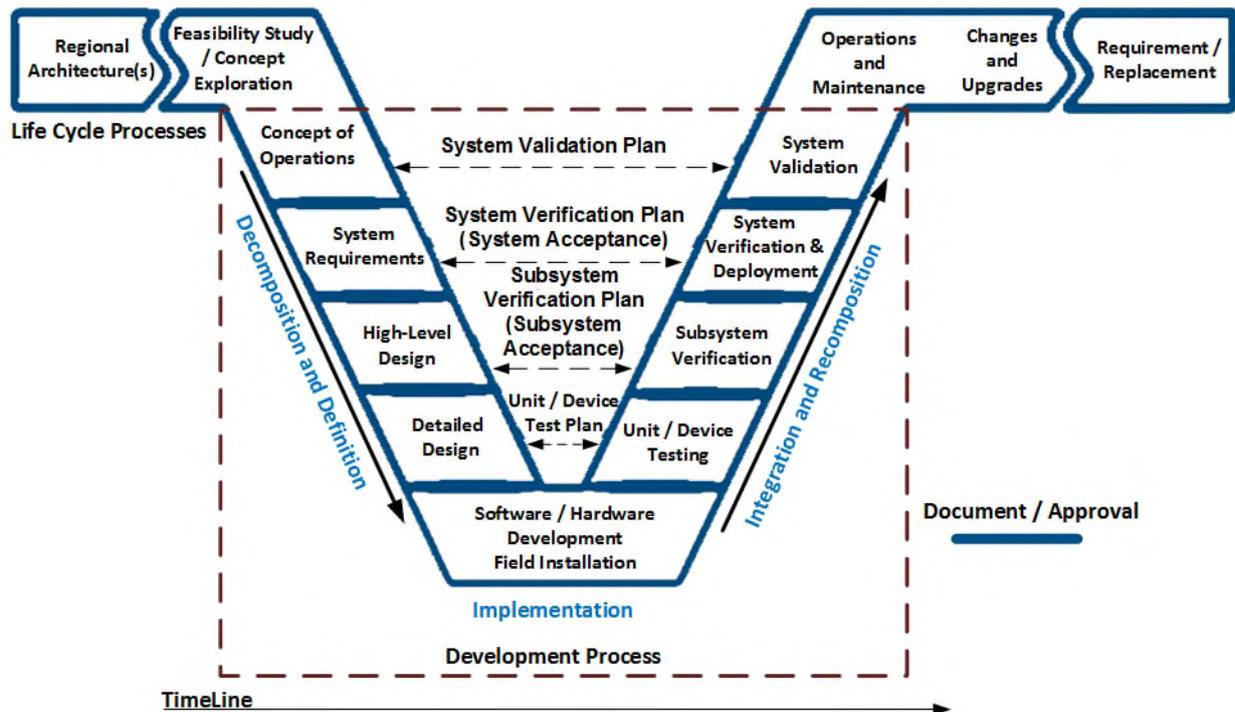


Figure 7: Systems Engineering Process - V-Model

executions are linked to input datasets and output datasets from an execution of a program and associating multiple software executions to input datasets extended to include many to many relationships between investigator and dataset, to capture their context in the provenance process.

4. Associating parameters with a software execution: Parameters must be linked with no less than one of many possible programs [can take zero or more parameters] executions corresponding to unique datasets but with assorted output datasets and runtime parameters.
5. Re-introducing the study: A study represents an amalgamation of associated investigations and a channel for relating SoftwareExecutions to each other and other types of investigations.
6. Introducing study nesting: Nesting an investigation inside one or more larger ones.

provenance – the transformation it has gone through in the lifecycle. The ICAT prototype does not yet allow for the propagation of the complete provenance of output data without unpacking the datasets to facilitate the querying of used transformations for researchers and neither does the extended proposed CSMD model. In addition, there is still scope for improving the software and hardware environment which have equally not been covered in this proposition.

International Council on Systems Engineering (INCOSE) Systems Engineering V-Model [Verification and Validation model] (Figure 7) is a process for ensuring effective and efficient satisfaction, high quality, trust worthy, cost efficient and schedule customer as well as stakeholder requirements throughout a systems lifecycle. It illustrates the product lifecycle from foundation to obsolescence or destruction (INCOSE 2015). The V-model lifecycle is a waterfall-like step-by-step process implementation that ensures predictability, stability, reproducibility, and substantial surety.

“Systems engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem” INCOSE (2015: 11).

This definition covers the data lifecycle from planning, creation to retirement disposal. A key strength of the systems engineering process is the consideration of the complete lifecycle of a project during the project development phase. It depicts the ten basic steps involved in the conception, planning, functional and technical specifications, and the implementation of a system. The implementation can vary from system to system in terms of timescales, costs and predictable lifespan. The essential of the V-Model illustrates the gradual advancement from requirements specification, system/product implementation, to verification processes.

The left part of the V-model characterizes problem identification. It depicts the process of functional and technical requirements gathering with evolving granularity, for product systems and subsystems, including components and the relationship amongst them. The left part of the V-Model comprises artefact abstraction, selection, and the design description of the product. The right part of the V-Model (Figure 8) depicts activities relating to the running and preserving, changing and enhancing, and eventual obsolescence or replacement of the system (INCOSE 2015; INCOSE 2017).

The Systems Engineering process can be used in the development of any product or system. The process can be summarized into the following seven tasks: State the problem (concept studies), Investigate alternatives (concept development), Model the system (preliminary design), Integrating the system (final design), Launch the system (fabrication), Assess performance (verify components /performance), and Re-evaluate (demonstration & Validation) (Jacobs 2015).

2. COMMONALITIES BETWEEN DATA LIFECYCLE MODELS, OSA-CBM, STANDARDS

Data management through its entire lifecycle still presents a number of complex challenges relating to interoperability, volume, storage, data citation, and metadata standards and data provenance (Porcal-Gonzalo 2015; Beaujardière 2016; Yang et al. 2013). This perhaps explains the proliferation of discipline or domain-specific data lifecycle models. Various disciplines and organizations are creating standardized frameworks, data ontologies, standards and unique data lifecycle models to suit their respective requirements. Metadata standards like Dublin Core (Hsu et al. 2015), Core Scientific Metadata (CSMD) provisions the basic metadata required to enhance the search functionalities

over data portals and knowledge libraries (Matthews et al. 2010), but falls short of propagating the complete provenance data. The shortfall with the CSMD is that neither does it support for “access to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data” as pointed out by Yang (2013:613).

However, though metadata standards are relevant for all data and knowledge outputs and organizations, each organization or project seems to have separate requirements for their own research or projects data. This is reflected in the 17 data lifecycle models identified, which all highlight the significance of metadata and standards.

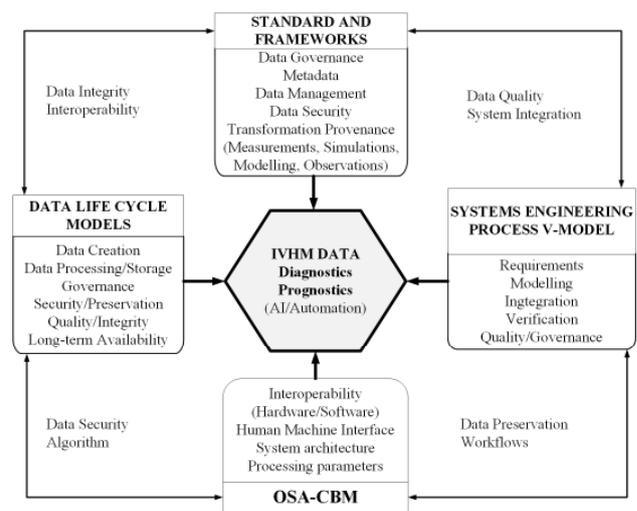


Figure 8: Integration parameters

The models seem to differ on the depth and breadth of applicability and priority of requirements. For instance the USGS emphasized three critical cross-cutting activities namely; metadata description, quality assurance and protection from corruption or loss to be performed parallel to planning, acquisition, processing, analysis, preservation, publishing and sharing to achieve enhanced quality, understanding and long-term reuse (Plale and Kouper 2017; Faundeen and Hutchison 2017).

The Digital Curation Centre (DCC) model suggests metadata should comprise rules and formalized entities for automatable tools and services as well as the role of data managers and data curators in the improvement of knowledge (Plale and Kouper 2017). Taking the DataOne data lifecycle model to illustrate the variance, though it represents all the classical stages of the data management lifecycle, all parts of the lifecycle are not mandatory (Plale and Kouper 2017; Allard 2014; DataONE 2017; Harrison 2013; Hidalgo et al. 2017; Pouchard 2015). This implies that the number of selected stages are dependent on the type of project or project requirements. Furthermore, although the

USGS addresses the weakness in most data lifecycle models identified and can be used in diverse settings despite having been developed mainly for USGS science data. The requirements on which this is developed are not consistent with small research facilities like the IVHM centre Labs or the University as a whole. None of the models identified seems to satisfy the scenario of IVHM.

The core of IVHM is the capture and analysis of data to establish advance indication of a future failures, state or distinctive characteristics of current assets in current the state. The conduct of prognostics and diagnostics is reliant on high-quality, reliable data. This quality assurance can be determined by the data management lifecycle. Data lifecycle models are often integrated with software services and policies (Plale and Kouper 2017). It is therefore fundamental to consider the associated process involved in the planning, designing and developing of the software services and policies.

The INCOSE system engineering process [V-Model] suitable for the development of any system, is vital in the scenario of IVHM knowledge management system and data lifecycle model. This is suitable for the IVHM data lifecycle and system development because of the restricted nature of this project. The V-Model is best suited for these kinds of projects that have well-defined length and scope, consistent technology, and a clear and well documented technical and functional specifications (INCOSE 2017).

The conceptual model above sheds light on the relationship between standards, data lifecycle models, OSA-CBM and the INCOSE systems engineering process model. It depicts the commonalities that are required for the IVHM Data lifecycle model.

Integrated Vehicle Health Management (IVHM) is shaped by prognostics and diagnostics that rely predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. The data has to be accurate, complete, timely, context relevant, reliable and explicit (Dibsdale 2011). IVHM is data-centric and driven by the OSA-CBM data model. The centrality of data for IVHM in the short, medium and long-term diagnostics whether it be historical asset health trends, operational status, load history, or fault identification, necessitates a data lifecycle model or a hybrid model consistent with OSA-CBM.

Of all the existing data lifecycle models, there is none that is consistent with the requirements of IVHM data and knowledge management requirements; that integrates OSA-CBM which is absolutely imperative to IVHM. The OSA-CBM model is based on the concept of metadata and interoperability that requires persistent visibility and traceability of data (Choudhary et al. 2016) within and across diverse platforms, systems and devices, and therefore making data provenance a fundamental requirement. This

has not been explicitly covered in any of the data lifecycle models. A new model is needed, and one that that integrates data acquisition with signal reception as well as data entry in order to accommodate the role of the systems or device operators for IVHM. In the next section, we describe the various phases of the proposed IVHM data lifecycle model.

3. THE INTEGRATED VEHICLE HEALTH MANAGEMENT DATA LIFECYCLE MODEL

The proposed IVHM Data lifecycle model (IVHM-DLCM) (Figure 9) is a hybrid that integrates relevant standards, frameworks and models that fit the profile of IHVM research and engineering activities. The IVHM-IVHM Data lifecycle model is scalable and can be used in diverse IVHM activities of all depths and breadths. It takes into consideration interoperability, integrity, quality, security, provenance and preservation of data throughout the lifecycle.

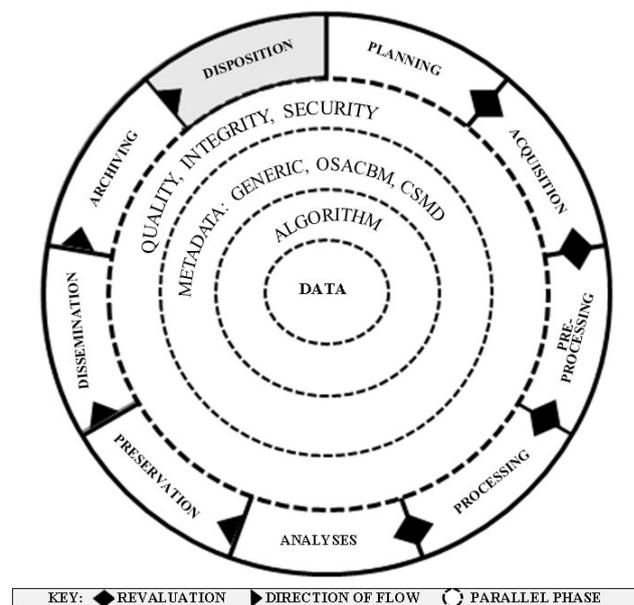


Figure 9: The IVHM Data Lifecycle Model

The IVHM-DLCM has nine discrete phases with three other phases running parallel. There also the revaluation process that runs parallel to phases One, Two, Three and Four.

3.1. Distinct Phases

3.1.1. Planning

The planning phase is the conception and beginning of simulations, observations, derivation, experiments and referencing (SODER) activity. In this phase, the research or engineering task is defined and planned – resources and planned deliverables for each phase are explored and explained. The requirements for success, quality, integrity and security are defined, including systems requirements.

The metadata parameters both generic and standard are considered in this phase. The data management plan, retention schedules and requirements specification are some of the recommended deliverables in this phase. In this phase, we also do the selection of the sensor modules and other accessories like software tools required for the SODER activity. File formats, data storage and sharing plans are developed at this stage.

3.1.2. Acquisition

The data acquisition phase represents data creation from scratch or the extraction of existing raw or derived data for reuse. In this stage, observational, experimental, simulation data creation takes place as well as the retrieval of existing datasets, derived or reference data for reuse. Data governance standards and best practices ensuring integrity, security, quality and metadata are considered. The quality of the process is also vital as it has a direct impact on the quality of the data created or retrieved for reuse.

3.1.3. Pre-Processing

The preprocessing phase aims to flag out of range data values, missing values to mitigate the risk of making decisions based on misleading results. It is fundamental to IVHM machine learning activities. It is used to clean the original signal by eliminating noise and to improve object component condition. In other words, this phase represents the low-level computation of sensor data and constitute a key element of the OSA-CBM architecture. Sensor data is transformed into an understandable format. This is the stage where data cleansing –detecting and correcting mistakes, incomplete, inaccurate, irrelevant and incorrect records from datasets. This step employs best practices to ensure that data is free of inconsistencies, correct, usable and reliable.

3.1.4. Processing

In this phase meaningful and relevant is extracted in suitable formats from raw datasets created after preprocessing for future use. This step involves activities such as validation, aggregation, summarization, sorting, classification and validation. It includes the conversation of data in usable and desired forms and formats. It can take the form mechanical, manual or digital processing.

3.1.5. Analyses

This represents activities like organization, interpretation and presentation of data. It involves statistical data analytics, simulations, modelling and other computation activities that reveal trends, facts, faults and tests theories and assumptions. In this step raw data is transformed into information and communicated to the stakeholders.

3.1.6. Revaluation: Phase 1 – Phase 5

Revaluation is one of the most important tools in The Systems Engineering Process. Re-evaluation observation of outputs and using the information to modify the system, the inputs, the product or the process (INCOSE 2017). It takes place in the first five phases. Feedback is collected at each phase on a situational basis to continually improve the steps and eliminate problems. The loops are used specifically when issues are identified as the lifecycle moves from one phase to the next.

3.1.7. Preservation

Preservation includes steps and processes for active data storage for the duration that it might be needed. It also involves access control and backup for security. These are the security actions taken to reduce the chances of data corruption and data loss. They also include submission to reliable data repositories.

3.1.8. Dissemination

This step includes the preparation and dissemination of datasets, derived data as well as findings or outcomes to relevant stakeholder communities. It improves accessibility as well as being a recommended best practice.

3.1.9. Archive

This the last step before disposition. It represents the retraction of data and related outputs from active circulation. The retention schedule is applied to the data and only accessible on demand.

3.1.10. Dispose

The data has reached the end of its useful and potential useful life. It is purged at this stage. It is securely disposed of to ensure that there can be no unwarranted access.

3.2. Parallel Phases

3.2.1. Metadata: Generic, CSMD, OSA-CBM

The strategy for metadata generation and documentation is referenced throughout the data lifecycle. As the data transitions from one phase to the next, it might also change platforms and devices making this phase particular important for addressing provenance issues. The metadata phase draws on the best generic metadata created by systems, the core scientific metadata model and the OSA-CBM built in metadata to create a scalable metadata model within the lifecycle. This allows for easy contraction and expansion research and engineering activities. It accommodates small to large scale data generating activities.

3.2.2. Data Integrity, Quality and Security

Data integrity involves creating, processing and maintaining the assurance, accuracy, consistency and completeness of data throughout its lifecycle. The content and meaning of data is maintained throughout its lifecycle. This also includes compliance with statutory requirements. Quality represents the use of best practice protocols and methods of collecting and organizing data that ensures its accessibility, completeness, validity, accuracy, consistency, availability and timeliness. Security involves the protection of data from unauthorized access to modify, use, delete and disclose. It includes protection against theft, breach of agreements, data protection laws and unintended or hateful modification. The computer system security, physical security and file security are all part of this step.

3.2.3. Algorithms

Refers to the algorithms and agents that automate research and engineering tasks. They are maintained as datafiles and at the same time they represent agents within the data- and knowledge management system.

4. CONCLUSION

We have identified key data lifecycle models (DLCM) and frameworks and found that though they had some of the elements for IVHM data, they lacked some essential ones. Thus, sustaining the cycle of data and knowledge management issues – creation, quality, storage, security and provenance. In the review, we found that the USGS DLCM encapsulated most data lifecycle models. The USGS reviewed more than 50 DLCMs to develop their one, and therefore was chosen as the ideal Data Life Cycle for integration. The strength of the proposed data lifecycle lies in the integration of key elements of the OSA-CBM framework, the CSMD, the engineering process model to create a scalable model that fits the depth and breadth of IVHM Research and Engineering operations. This model supports the design and implementation of protocols for effective and efficient data management. It provides a foundation for Data- and knowledge management system requirements.

ACKNOWLEDGEMENT

This Project is sponsored by Boeing and hosted by Integrated Vehicle Health Management (IVHM) Centre at Cranfield University, United Kingdom.

REFERENCES

Allard, Suzie. 2014. "Evaluating a Complex Project: DataOne." P. 436 in *Research Data Management: Practical Strategies for Information Professionals, Charleston Insights in Library, Informat*, edited by J. M. Ray. Purdue University Press.

Arahchige, Buddhi and Suresh Perinpanayagam. 2017. "Uncertainty Quantification in Aircraft Gas Turbine Engines." Pp. 1628–38 in *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*. Vol. 232. IMECHE.

Arass, M. E., I. Tikito, and N. Souissi. 2017. "Data Lifecycles Analysis: Towards Intelligent Cycle." *2017 Intelligent Systems and Computer Vision, ISCV 2017*.

Ball, A. 2012. "Review of Data Management Lifecycle Models." (February):15.

Beaujardière, Jeff De La. 2016. "NOAA Environmental Data Management." *Journal of Map & Geography Libraries* 12(1):5–27.

Choudhary, R., S. Perinpanayagam, and E. Butans. 2016. "Design and Analysis of Communication Model for Implementation of CBM Systems Based on OSA-CBM Framework." Pp. 1–7 in *2016 IEEE Aerospace Conference*.

DataONE. 2017. "Data Life Cycle." *Best Practices* 1. Retrieved December 21, 2017 (<https://www.dataone.org/data-life-cycle>).

DCC. 2008. "The DCC Curation Life-Cycle Model."

Dibsdale, Charlie. 2011. "Integrated Vehicle Health Management Operations Rooms." Pp. 113–23 in *Integrated Vehicle Health Management: Perspectives on an Emerging Field*, edited by I. K. Jennions. SAE International.

Faundeen, J. L., T. E. Burley, J. A. Carlino, D. L. Govoni, H. S. Henkel, S. L. Holl, V. B. Hutchison, E. Martin, E. T. Montgomery, C. C. Ladino, S. Tessler, and L. S. Zolly. 2013. *The United States Geological Survey Science Data Lifecycle Model*. South Dakota: U.S. Dept. of the Interior.

Faundeen, John and Vivian Hutchison. 2017. "The Evolution, Approval and Implementation of the U.S. Geological Survey Science Data Lifecycle Model." *Journal of EScience Librarianship* 6(2):e1117.

Felke, Tim, George Hadden, David Miller, and Dinkar Mylaraswamy. 2010. "Architectures For Integrated Vehicle Health Management." *AIAA Infotech@Aerospace 2010* (April).

Goebel, Kai. 2011. "Basic Principles." Pp. 55–65 in *Integrated vehicle health management: perspectives on an emerging field*, edited by I. K. Jennions. SAE International.

ten Ham-Baloyi, Wilma and Portia Jordan. 2016. "Systematic Review as a Research Method in Post-Graduate Nursing Education." *Health SA Gesondheid* 21:120–28.

Harrison, Kate. 2013. *A Guide to Data Management in Ecology and Evolution*. edited by K. Harrison. British Ecological Society.

Hidalga, Abraham Nieva de la, Barbara Magagna, Markus Stocker, Alex Hardisty, Paul Martin, Zhiming Zhao, Malcolm Atkinson, and Keith Jeffery. 2017. *ENVRI Reference Model News*.

- Hsu, Leslie, Raleigh L. Martin, Brandon McElroy, Kimberly Litwin-Miller, and Wonsuck Kim. 2015. "Data Management, Sharing, and Reuse in Experimental Geomorphology: Challenges, Strategies, and Scientific Opportunities." *Geomorphology* 244:180–89.
- INCOSE. 2015. *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*. 4th ed. edited by W.-E. DaVid d., R.-E. Garry J., F. KeVin J., H. DoUgLaS R., and Sh.-C. THomaS m. John Wiley and Sons, Inc.
- INCOSE. 2017. *Guide to the Systems Engineering Body of Knowledge (SEBoK)*. v1.9 R.D. edited by R. Adcock. Hoboken, NJ: The Trustees of the Stevens Institute of Technology.
- Jacobs, Stuart. 2015. "Systems Engineering." Pp. 31–60 in *Engineering Information Security*, edited by S. Jacobs. John Wiley & Sons, Inc.
- Jantunen, E, J. Junnola, and U. Gorostegui. 2017. "Maintenance Supported by Cyber-Physical Systems and Cloud Technology." Pp. 708–13 in *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)*.
- Jantunen, Erkki, Urko Zurutuza, Michele Albano, Giovanni di Orio, Pedro Maló, and Csaba Hegedus. 2017. "The Way Cyber Physical Systems Will Revolutionise Maintenance." Pp. 145–48 in *30th Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2017)*. 10 to 13, Jul, 2017. Preston, United Kingdom.
- Jennions, Ian K. 2011. "Introduction." Pp. 1–7 in *Integrated vehicle health management: perspectives on an emerging field*, edited by I. K. Jennions. SAE International.
- Lebold, Mitchell, Karl Reichard, Carl S. Byington, and Rolf Orsagh. 2002. "OSA-CBM Architecture Development with Emphasis on XML Implementations." *Maintenance and Reliability Conference (MARCON)* 6–8.
- Löhr, Andreas and Matthias Buderath. 2014. "Evolving the Data Management Backbone: Binary OSA-CBM and Code Generation for OSA-EAI." Pp. 1–10 in *EUROPEAN CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY 2014*. PHM.
- Matthews, Brian, Shoaib Sufi, Damian Flannery, Laurent Lerusse, Tom Griffin, Michael Gleaves, and Kerstin Kleese. 2010. "Using a Core Scientific Metadata Model in Large-Scale Facilities." *International Journal of Digital Curation* 5(1):106–18.
- MIMOSA. 2018. "Open System Architecture for Condition-Based Maintenance (OSA-CBM)." *OSA-CBM*. Retrieved December 2, 2017 (<http://www.mimosa.org/mimosa-osa-cbm>).
- Parry, Michael J. 2016. "Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections." *The Electronic Library* 34(6):1055.
- Perinpanayagam, Suresh. 2013. "Chapter 2 Sensors, Instrumentation, and Signal Processing." Pp. 5–25 in *Integrated Vehicle Health Management: the Technology*, edited by I. Jennions. Pennsylvania: SAE International.
- Plale, Beth and Inna Kouper. 2017. "The Centrality of Data: Data Lifecycle and Data Pipelines." Pp. 91–111 in *Data Analytics for Intelligent Transportation Systems*. Elsevier.
- Porcal-Gonzalo, Maria C. 2015. "A Strategy for the Management, Preservation, and Reutilization of Geographical Information Based on the Lifecycle of Geospatial Data: An Assessment and a Proposal Based on Experiences from Spain and Europe." *Journal of Map & Geography Libraries* 11(3):289–329.
- Pouchard, Line. 2015. "Revisiting the Data Lifecycle with Big Data Curation." *International Journal of Digital Curation* 10(2):176–92.
- Prajapati, A. K., B. K. Roy, and R. Prasad. 2018. "A State of Art Review of Integrated Vehicle Health Management System." Pp. 1–5 in *2018 3rd International Conference for Convergence in Technology (I2CT)*.
- Redding, Louis. 2011. "Integrated Vehicle Health Management: Perspectives on an Emerging Field." Pp. 17–26 in *Integrated vehicle health management: perspectives on an emerging field*, edited by I. K. Jennions. SAE International.
- Sabharwal, Arjun. 2015. *Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections*. Chandos Publishing.
- Simonet, Anthony, Gilles Fedak, and Matei Ripeanu. 2015. "Active Data: A Programming Model to Manage Data Life Cycle across Heterogeneous Systems and Infrastructures." *Future Generation Computer Systems* 53(Supplement C):25–42.
- Sreenuch, T., A. Tsourdos, and I. K. Jennions. 2013. "Distributed Embedded Condition Monitoring Systems Based on OSA-CBM Standard." *Computer Standards & Interfaces* 35(2):238–46.
- Yang, Erica, Brian Matthews, and Michael Wilson. 2013. "Enhancing the Core Scientific Metadata Model to Incorporate Derived Data." *Future Generation Computer Systems* 29(2):612–23.
- Yu, Xiaojun and Qiaoyan Wen. 2010. "A View about Cloud Data Security from Data Life Cycle." *2010 International Conference on Computational Intelligence and Software Engineering* (4072020):1–4.

BIOGRAPHIES



Alexslis Maindze is a final year researcher at the Integrated Vehicle Health Management (IVHM) Centre at Cranfield University. Winner of the Flemish Inter-University Council scholarship, 2002 and the PHM Doctoral Symposium award 2018. He has more than 7 years' experience in industry. He has successfully implemented knowledge and information management projects working for Converteam (now GE Energy) and Lancaster University, Plan International Ltd, the National Health Service (NHS), and Scottish and Southern Energy Network (SSEN). He received a B.Ed in Curriculum Studies and Teaching/Geography from the University of Buea, Cameroon. He received his MS in Human Ecology from Vrije Universiteit Brussels in 2004, and an MS in Knowledge and Information Management from the University of Gent, Belgium. His research interests include data, knowledge and information management, IVHM enabled information analysis and integration, IVHM application in operations maintenance and supply chain including the foundation for enhanced availability, capability and readiness.



Zakwan Skaf is currently a Lecturer in Diagnostics and Prognostics in the Integrated Vehicle Health Management (IVHM) at Cranfield University. Before joining Cranfield University, he held several positions with Loughborough University, Sheffield University, Warwick University and the University of Manchester. Zakwan received his B.S. degree from the faculty of Mechanical Engineering in 2001, and his MSc and PhD degrees from the University of Manchester, Control Systems Centre. He is an active researcher with over 35 publications in many world class peer-reviewed journals and conferences. His research interest involves control engineering, data analytics, machine learning, condition monitoring, diagnostics, and prognostics related to aerospace, railway, marine, energy, and automotive applications.



Ian K. Jennions' career spans over 40 years, working mostly for a variety of gas turbine companies. He has a Mechanical Engineering degree and a PhD in CFD both from Imperial College, London. He has worked for Rolls-Royce (twice), General Electric and Alstom in a number of technical roles, gaining experience in aerodynamics, heat transfer, fluid systems, mechanical design, combustion, services and IVHM. Ian moved to Cranfield in July 2008 as Professor and Director of the

newly formed IVHM Centre. He is on the editorial Board for the International Journal of Condition Monitoring, a Director of the PHM Society, Chairman of SAE's IVHM Steering Group, contributing member of the SAE HM-1 IVHM committee, a Chartered Engineer and a Fellow of IMechE, RAeS and ASME. He is the editor of five SAE books and co-author of another one.