INTERNATIONAL DESIGN CONFERENCE - DESIGN 2018
Dubrovnik - Croatia, May 21-24, 2018.

TOWARDS A FRAMEWORK FOR ENGINEERING BIG DATA: AN AUTOMOTIVE SYSTEMS PERSPECTIVE

Thomas J. Byrne, Felician Campean, Daniel Neagu

University of Bradford, Advanced Automotive Analytics Research Centre

Abstract
Demand for more sophisticated models to meet big data expectations require significant data repository obligations, operating concurrently in higher-level applications. Current models provide only disjointed modelling paradigms. The proposed framework addresses the need for higher-level abstraction, using low-level logic in the form of axioms, from which higher-level functionality is logically derived. The framework facilitates definition and usage of subjective structures across the cyber-physical system domain, and is intended to converge the range of heterogeneous data-driven objects.

Keywords: knowledge management, big data analysis, design models

1. Introduction
In the emerging cyber-physical systems (CPS) domain, data is the new fuel that powers decision making across the whole product lifecycle. Big data is now ubiquitous in most industry domains, gathered from a heterogeneous range of data and information sources, which present significant variation of data features such as quantities, formats, quality, and provenance. What is currently still required are integrated processes to effectively convert data into information, before rendering it contextual intelligence (CI) for operational advantage (Kutz, 2017). Within this process, the particular challenge with complex CPS is the consistent integration of data and information models with the physical system; for example, in relation to the reliability of a CPS, the physical system model and the information system representing the physical condition of the system underpins the system diagnostics and prognostics.
Using the automotive industry as example, a first challenge is to identify suitable data structures to manage information and insight from a range of data sources. This paper takes as reference case for discussion the specific problem of powertrain healthcare (with the overall aim of developing an intelligent personalised dynamic asset health management for vehicle systems). Figure 1 illustrates the range of data sources that could provide useful information, from across all phases of the system lifecycle – from product development to manufacturing, use and retirement. This includes a mix of offline data sources (e.g. durability test data, material and manufacturing process data, maintenance / warranty data, dimensional metrology data – in test, manufacturing and after use) and online sources (e.g. sensor data recorder by ECU or available as data over the air (DOTA), electronic systems

diagnostics). The exponential increase in data sources and volumes demands for significant efforts towards data governance and management.

Figure 1. Engineering data sources for powertrain healthcare assessment

Figure 2 presents an adaptation of the Data-Information-Knowledge-Wisdom (DIKW) hierarchy for the powertrain healthcare domain, illustrating the transformative journey from data to CI (Campean & Neagu, 2016). Such hierarchical structures contextualise and formalise current demands on big data expectations, including a smooth, intrinsic transformation from initial resources (data) to manageable information towards optimal ergonomic human system interaction, automated decision support with machine learning-based models, and a new level of knowledge governance (i.e. wisdom) through computational intelligence. Consequently, the initial resources that are data (coming in the form of both relational and non-relational formats from off-line as well as real-time inputs) are synthesised by nominating and recording relevant observations with quantitative and qualitative values as information for the use of experts and secondary systems. Such understanding of primary resources allows interpretation of faults (diagnostics) and their validation in a reactive form (e.g. analysis of fault records from fleet or garage data). The know-how stage of transforming information (and data) resources to knowledge permits pattern recognition strategies, assessing quality and applying machine learning (ML) models to prognostic models and personalised computational applications of data mining, such as personalised vehicle healthcare - currently the subject of extensive research and development for original equipment manufacturers (OEMs). The challenging aspect is the transformation of all three base yet hierarchical layers (DIK) to wisdom by computational intelligence in the endeavour to create model based systems as a formal ML process to define and implement sustainable resilient systems.

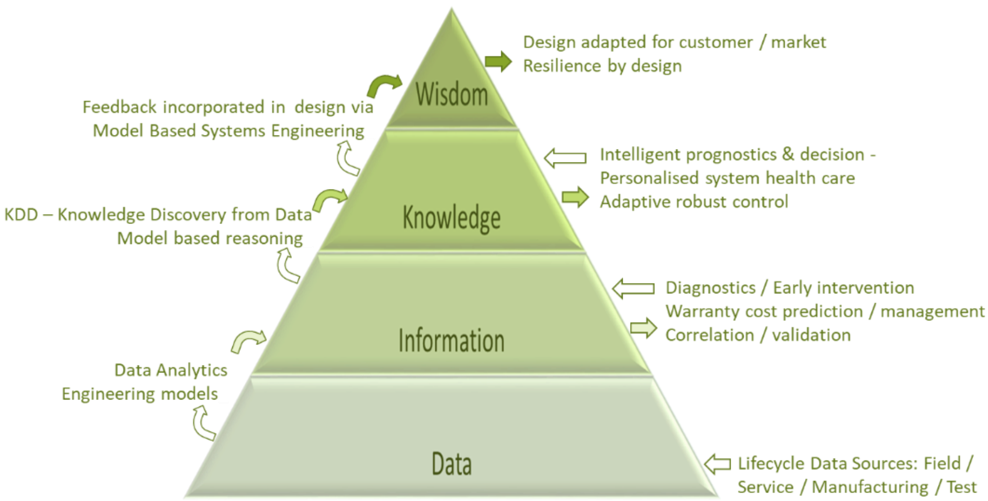

Figure 2. Transforming data to CI: powertrain healthcare example

This vision brings a holistic approach to the reliability challenge facing the automotive systems lifecycle. In theory, modern technology (building on smart sensors and Internet of Things) can monitor the condition of operational components and their environment in real-time. For instance, diagnostic analytics assess the probability of failure, while attempting to identify anomalies and mitigate system malfunction based on data over-the-air as well as historical data records. However, the current lack of multi-disciplinary integration between the physical system and the information system mean that the potential current technology offers is not fully exploited. This challenge motivates the current proposal of a framework for engineering big data.

At the University of Bradford's Automotive Research Centre, in the Advanced Automotive Analytics (AAA) research unit, we are developing new ways to engage big data with comprehensive computational logic. We take an object-level solution-independent approach to develop and implement a highly abstract modelling framework for engineering big data, with further research efforts focused on utilising category theory to explore dynamical systems enhancement.

The need for such an approach is justified by some of the problems mentioned below:

- Solution-neutral approaches can solve many problems at the conceptual level, but there are few good models for implementing those solutions. Addressing this deficiency will require substantial consolidation and a concerted effort to bring physical systems and information systems together.
- CPS products have become more complex, their design and development procedures are multi-disciplinary, and the demands for big data governance and management are increasing. A schema of denotation is required to represent the set of heterogeneous sources of which are shared by several divisions of the same system.
- Heterogeneous CPS modelling languages are cross-disciplinary and incompatible. Conceptual exchange is often incoherent and not possible due to domain specific preconditioning. A prevailing cross-discipline "artefact" that is the source of compatibility, is needed to reinforce the universe of discourse within the CPS design process.

The work presented in this paper is to deal with these problems based on a set of domain-neutral theories as a foundation of further research practice. In the process of engineering big data for CPS design, conceptual modelling is used to represent a model independent from an eventual solution that applies a domain-neutral modelling technique at the object-level of CPS design.

The research methodology involves a bottom-up structure of the investigation with *specific objectives*:

- A formal application of object representation in CPS, using model-based ontological reasoning, from a phenomenological view point, an ontological model of object distinguishes between the physical object and the conceptual object that is represented in several forms. This is listed as future work because of current limited space;
- A theoretical basis towards a domain-independent and omnipresent physics principles (discussed in Section 3), applied to the object-level for CPS design;
- A rigorous framework for complex multi-disciplinary design using axiomatic-based categorisation of data-objects, to consolidate the heterogeneous nature of the CPS domain;
- A discussion of the complexity of big data challenges that offers a homogeneous preconditioning strategy to confront the challenge at the constituent level (object-level), bringing big data into the desired state of distribution for data governance and management protocol and procedure.

The paper is structured as follows. Existing data management technology is reviewed in the second section of the paper, in contrast with the evolution of system modelling technology, discussed from functionality and implementation perspectives. The third section outlines the primary function for engineering big data and its expected behaviour. The focus of this paper is on proposing a framework and its structural components that will support engineering big data applications as a DIKW hierarchy, phase one of a multi part design methodology. The fundamental principles that underpin the rationale for this paper are introduced in the key definitions covered in section three, the philosophy behind the approach, alongside the framework model, its features and a narrative with explicit justification and reasoning. Following the rationalisation of the framework methodology, section four summarises the key points of the framework, and illustrates processes that occur within the object-level of the proposed framework. The paper concludes with a brief roundup of important features and the next phase of research, and discussion on the impact of Engineering Big Data on knowledge governance, management and how shareholders and researchers will benefit from strategies in the current Big Data Science field.

2. Review of data repository technology and engineering systems modelling

2.1. Data warehouse repository

Data warehouse technology supports detailed, holistic and the homogeneous, long-term management of integrated data and domain information. Developed with a particular focus on processing and providing access to large data sets, it is “a subject-oriented, integrated, time variant and non-volatile collection of data used in strategic decision making” (Inmon, 2015). In order words, a data warehouse re-structures data into organised domain information, more flexible and intelligent than a relational database, it supports aggregated atomic data values to present different granularities of information. Data warehouse information modelling for multi-dimensional information management has three general modelling schemas. The Star schema (the most common modelling paradigm) contains a large central table (flow table) with no redundancy and a set of smaller attended tables (dimension tables), one for each dimension. In a Snowflake schema (Figure 3) a variant of the Star schema model, selected dimension tables might be normalised and thus further split the data into additional dimension tables. The Snowflake schema provides an example of potential high-level data warehouse modelling solution for engineering big data. The Starflake schema combines both modelling paradigms *(*Lehner & Sattler, 2013).

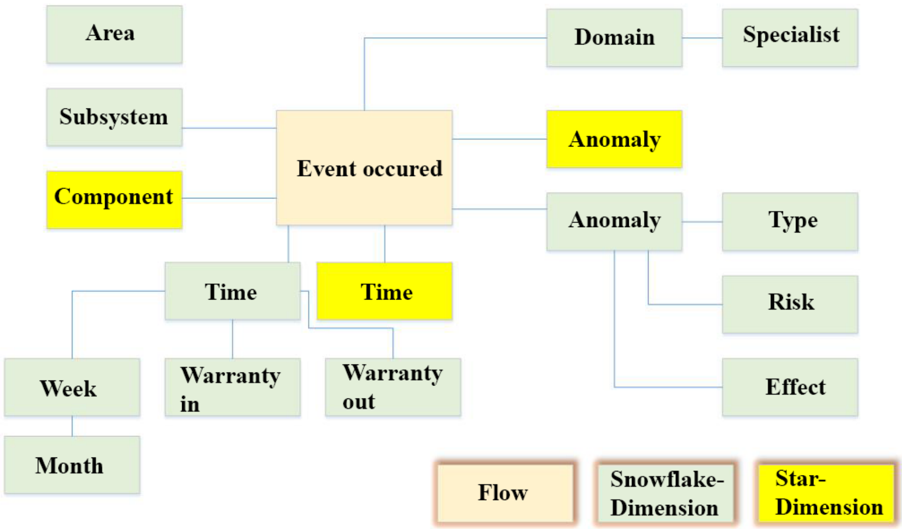

Figure 3. Information model - Snowflake-dimension schema

Also referred to as ‘fact data’ (Lehner, 2002), flow-data is responsible for most of the content in this data repository. It is stored in a ‘Flow Table’, which may contain millions of rows, with multiple primary key values and primary key functionality. Flow-data defines domain events by specifying the dimensions accessed by data values stored in the dimension tables. It is this type of CI that supports domain specific decision-making. The granularity of big data stored in the repository determines the demand on the analytics factory.

In the context of the data factory (see Section 4), dimension data is used for supporting temporal and spatial dimensionalities for a posteriori determinations by recourse to experience or experiment based on empirical particulars provided by sensors (Internet of Things), the domain of the superstructure synthetic model. It is less extensive than flow tables, non-dynamic, and de-normalised, in the sense that, dimension tables require little change. Previously-normalised databases (relational DBMS) use a similar strategy to increase performance. Reference Tables support the management of data stored in the dimensions, and can reduce the amount of data needed in the warehouse, and thereby the amount of direct access required. Derivative data is created from two or more sources of data, can be more efficiently stored and accessed, and is usually created as part of the routine that transforms data prior to storage. Another useful characteristic of this technology is the design of summary tables, which can improve query performance by allowing queries direct access to pre-calculated summaries and predefined views of data.

The unique features that support data warehouse technology (multi-dimensional, flexible, dynamically adaptable database applications) complement the proposed (data) analytics factory requirements, in the sense that its features support fundamental requirements for interpolating raw data with computational logic. The categorisation process (see Section 3), which constitute and configure the framework this paper introduces, also require multi-dimensionality to maintain the composition of the homogeneous manifold (see Section 4) and the integrity of object-level interdependency. While maintaining essential characteristics of a data warehouse, data analytics develops upon these attributes to take advantage of its structure, functionality and method.

2.2. Systems modelling techniques

A good model has a degree of variability designed to deal with system fluctuations, insofar as the context is already part of the programmed experience, according to a formal set of rules. This paper introduces an informal set of rules formulated to adapt and take an active role in system applications. Runtime applications (Morin et al, 2009; Bencomo et al, 2008; Floch et al, 2006), by virtue of holding a particular detachment from 'prior to runtime', or 'post runtime' applications suggest at least, the requirement for legitimate ad hoc in play processing. As other modelling techniques also endeavour to achieve, runtime applications advocate a real ML adaptability approach, as opposed to function modelling, or models based on the behaviour of the system, or the inputs and the outputs of subsystem states (Eisenbart et al. 2016; Eisenbart, 2014; Srinivasan et al., 2012; Pahl et al., 2007; King and Sivaloganathan, 1998). Nevertheless, model driven engineering is usually based on an understanding of how the system works, or what function it is required to deliver.

Function modelling for system analysis promotes a top-down decomposition of the main function into sub-functions. System functions can be visualised through use case scenarios, state views, subsystem interaction, actor view, effect view, and process flow views (Eisenbart, 2014; Eisenbart et al. 2016). Applications of function analysis can be classified as value analysis, failure analysis, concept analysis, artificial intelligence, and function classification. Other function modelling approaches are considered under the six headings of ontology, semantic definition of function, function representation formalism, function-context relation, decomposition and verification, and implementation in a programming environment (King and Sivaloganathan, 1998). The level of abstraction, requirement-solution, system-environment, and intended-unintended functionality, discuss function with an emphasis on the chronology of developing function definitions and function representations (Srinivasan et al., 2012).

Others introduce a concise taxonomy based on the flow of materials, energy, and information through a system (Pahl et al., 2007). It remains good practice to avail of these models to facilitate a better understanding of the requirements, the modelling and analysis of complex system architecture and aggregation of function models across multiple modes of operation, and assess and predict system interaction early in the engineering design process. Most engineering teams start to develop a functional model based on the understanding of how the system works rather than what function it is required to deliver (Yildrim et al. 2017), in this paper a solution-neutral methodology of system design is offered, by means of a rigorous framework for complex multi-disciplinary systems using axiomatic-based object-level modelling for CPS design.

3. The rationale for a methodology based on universal principles of categories

In this section, we argue qualitatively that when a system satisfies a theoretical basis for CPS towards certain preconditions, the system is stable and controllable within a specified framework. Developed around how we structure the world in our mind, rather than how the world is structured, this methodology is based on human thought processes, as a juxtaposition of two logically derived mathematical structures grounded in Newtonian physics. The categories of thought are identified as original Newtonian concepts, so that the conclusion is that Newtonian science is dealing only with subjective structures of our thinking. It is an attempt to identify the inner resources of the mind and isolate all the empirical particulars, and evaluate what is left. Even if we remove from experience everything that belongs to the senses (empirical particulars), there remain nevertheless certain original concepts, and certain judgements derived from them, which must have had their origin entirely a priori, independent from experience. With this rationale, the categories of logical kinds of judgement are formed, as shown in Table 1.

Sensing precedes understanding, insofar as the faculty of sensing is distinguished from the faculty of thinking. Our perception of objects comes from two things, the raw material world that we experience through our senses, and the form that the mind gives to that experience. Perceptual experience is formed from structured sensual experience. If empirical input comes to us as atomistic perceptions, somehow it gets sorted and ordered. Therefore, our faculties provide a structure to unify sense experience, while the mind provides structural principles that enable us to conceptualise what goes on in the world of perceptual experience. What the understanding does, it formulates judgements about perceptual experience using a structured set of categories. The categories are simple ways in which we think, which give order in the mental world, whereby the mind is the active contributor that structures experience and thought. The a priori categories are the universal principles applied to data-objects in the analytic model. Other data-objects, representing the physical components of the powertrain system, domain specific analytical specialists, or more enduring information data such as warranty data and engine test data, which alter only in quantum leaps, are also interpolated with the universal principles, providing an a priori continuity across the entire CPS domain.

3.1. Object-level categorisation in the cyber-physical domain

The a priori categorisation of data-objects distinguish analytical contextualisation from the a posteriori experimentation process, which is controlled in the synthetic model built on top. The initial analytic model supports real-time data acquisition processing, categorisation classification and association processing (i.e. associating real world data to data-objects of which the categories apply), conjoining the ‘data’ layer with the ‘information’ layer shown in Figure 2. The analytic model provides the preconditions that make categorisation possible, while the synthetic model provides the preconditions for conceptual understanding. Just as the forms structure perception, the categories of understanding, shown in Table 2, give structure to information systems at the object-level of understanding, satisfying a theoretical basis for CPS towards a stable and controllable set of preconditions within the specified framework: the way that we structure the world, rather than the way the world is structured. The categories are defined through reasoning; “if these are the ways in which we understand things, the ways in which we classify our experiences, then it’s natural that if you can lay out a classification of different kinds of judgement we make, that those judgements are likely to embody the a priori categories” (Kant, 2008). The forms of perception meet the categories of understanding, the perception that comes into the mind, and then the understanding that gets a hold of it. Perceptions are particulars, the categories are universal. If every single representation stood by itself, every particular sense idea, simple idea, isolated from the others, nothing like what we call knowledge could ever arise, because knowledge forms a whole of representation, connected and compared with each other.

Table 1. Logical kinds of judgement (Kant, 2008)

Quantity of Judgement	Quality	Relation	Modality
Universal	Affirmative	Categorical	Problematical
Particular	Negative	Hypothetical	Assertorical
Singular	Infinite/Indefinite	Disjunctive	Apodictical

Table 2. Categories of understanding (Kant, 2008)

Of Quantity	Of Quality	Of Relation	Of Modality
Unity	Reality	Of Inheritance & Subsistence	Possibility - Impossibility
Plurality	Negation	Of Causality & Dependence	Existence – Non-existence
Totality	Limitation	Of Community	Necessity – Contingence

We can relate the concept of time to all the categories using axioms (see Section 4); what we develop then is an abstraction, a temporalised conception of cause and effect, or of substance (see Table 2), whereby the cause must be the concurrent with, or antecedent to the effect; the idea of substance is the idea that something ‘is’, it has an ‘enduring identity’, continuity in time. The categories in relation to time provide a schema, whereby our pure sensuous concepts depend on some schemata of objects relative in time. The schemata therefore is nothing but a priori determinations of time according to rules, or ways of thinking about time according to rules, and these apply to all data-objects, following the arrangement of the categories, relate to the series in time, the content in time, the order in time, and finally, to the complex or totality in time.

4. Design axioms - a priori determinations of time according to rules

This section discusses a rigorous framework for complex multi-disciplinary design using axiomatic-based object-level modelling. Logical axioms are statements that are taken to be true within the system of logic they define. Axioms of the categories are arithmetic-based statements that serve as a starting point from which other, more complex mathematical statements, are logically derived. The axiomatic arrangement of concerted data-objects in a CPS, accumulate continuous temporal representation of form and event, in the sense that it reduces the virtual experience to the science of geometry and arithmetic, through which the representations of a determinate space and time are generated. On this basis is it possible to advance the synthesis of representation as a temporal mode, through the composition of the homogeneous manifold. The origin of our axiomatic system lies in Kant's (2008) schemata of the understanding, outlined below.

On the Category of Quantity

Software applications with the processing acumen to characterise the quantity of data-objects as *universal*, *particular*, or *singular*, contain and represent the synthesis of time itself, in the successive determination of data-objects, whereby:

- *Unity* is defined as the state of objects being united or joined as a whole in a determined time;
- *Plurality* is defined as an object containing several diverse elements in a determined time; and
- *Totality* is nothing else but plurality contemplated as unity.

On the Category of Quality

Software applications with the processing acumen to characterise the quality of data-objects as *affirmative*, *negative*, or *infinite*, are defined as the synthesis of data-objects with the representation of time, whereby:

- *Reality* is existence in a determined time;
- *Negation* is the zero quantity of something in so far as it does not fill time, in a determined time; and
- *Limitation* is the quantity of something in so far as it fills time, is exactly this continuous and uniform generation of the reality in time, as we descend in time from a certain degree, down to the vanishing thereof, or gradually ascend from negation to the quantity thereof.

In the Category of Relation

Applications with the processing acumen to characterise the relationship between objects as *categorical*, *hypothetical*, or *disjunctive*, are defined in the context of the relationship of data-objects to each other in all time, whereby:

- *Inheritance* or *subsistence* is the permanence of the real in time, the idea of substance is the idea that something 'is', it has an 'enduring identity', a continuity in time;
- *Causality* of a thing as the real which, when posited, is always followed by something else, the cause must be the concurrent with, or antecedent to the effect; and
- *Community* is the reciprocal causality of substances in respect of their events, is the coexistence of the determinations of the one with those of the other - reciprocity of action and reaction.

In the Category of Modality

Applications with the software processing acumen that characterise the schema of modality and the categories, time itself, as the correlative of the determination of a data-object, whether it does belong to time, and how. Objects are characterised as *problematic*, *assertoric*, or *apodictic*, in the sense that:

- *Possibility* is the accordance of a synthesis of different representations with the conditions of time in general (e.g. opposites cannot exist together at the same time in the same thing, but only after each other) and is therefore the determination of the representation of a thing at any one time;
- *Existence* in a determined time; and
- *Necessity* is the existence of a data-object in all time.

Temporally oriented data-objects establish an accord between the analytic and synthetic model, whereby high-level computational statements, adhering to these axiomatic concepts of time, advance the synthesis of representation as a homogeneous manifold across the analytic synthetic threshold. Arithmetic being the science of time (the form of the inner sense), and geometry being the science of space (the form of the outer sense), retain the character of a temporal hypothesis referring to all possible relationships between data-objects and events, determined in pure a prior form.

This paper discusses strategies of using axioms for systems, in order to synchronise the empirical particulars of the homogeneous manifold, to be considered as aggregates, i.e. as a collection of previously given parts with their contextually relevant data-object, as an 'enduring identity', a continuity in time. The axiomatic categorisation of data-objects is a set of constructs and rules to combine those data-objects; the temporal hypothesis is a method of procedures by which the axiomatic categorisation system can be used; a data-object is the product of the modelling process; and the Automotive Analytics Factory is the setting in which the modelling can occur. The quality of conceptual modelling is believed to have an enormous impact on information systems analysis and design (Recker & Bjorn, 2008).

The first stage of engineering big data is transforming heterogeneous datasets into homogeneous subject oriented information (see Figure 4). Only thereafter, can we begin to convert big data from information into contextually relevant knowledge. Following that, the superstructure component of this framework is the complementary body of work, its function is illustrated in Figure 5 (though not discussed in detail because of limited space). This paper is focused on the substructure of the framework, the analytic model. Defined as the substructure with the superstructure (synthetic model) built on top. The concepts used to express the former are oriented around the natural conjugate of the powertrain system, its physical components, material properties, and mechanical functions. The latter superstructure, distinguishes the permanent from the changeable characteristics of the CPS. The effect of the synthetic model is determinable by recourse to experience or experiment based on empirical particulars provided by sensors (e.g. Internet of Things) to the information system at predetermined intervals, i.e. the real-time data representing the homogeneous manifold of the physical system. This covers constructs required in engineering big data, the categorisation procedure is developed to consolidate the physical particulars with their contextually relevant data-objects.

5. Proposed framework for engineering big data

The proposed framework is a comprehensive methodology developed to interpolate data-objects with low-level logic, maintaining logical associations with higher level functionality. For example, engineering big data with computational logic for applications in more complex logical statements in higher level system operations. For illustration purposes, the case of engineering big data to assist in the development of a powertrain healthcare system is depicted in Figures 4 and 5. In particular, the category of relational judgement (Table 1) is explained in context with the axiomatic agent of the object manifold. Figure 4 describes in 3 phases how heterogeneous datasets are processed into information:

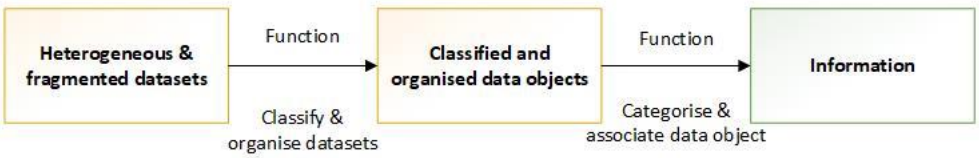

Figure 4. Dataset information transfer function model

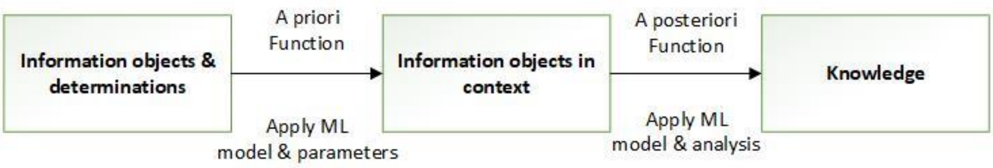

Figure 5. Information to knowledge transfer function model

Prior to operational processing, the four categories charged with supplementing contextual data-objects with omnipresent principles, first associate empirical datasets (engineering big data) with their physical counterpart components of the powertrain system, which are in turn categorised under the same

omnipresent structures that enable us to contextualise observed events. This homogeneous association of data-objects is the a priori distribution state of the analytic factory.
Associations are used to represent a wide range of connections amongst sets of objects. Such relationships are modelled by using associations and aggregations denoting hierarchical aspects in a community of objects coexisting as determinations of the one with those of the other.
Initial processing occurs in the substructure (the 'analytic model' – see the lower segment of Figure 6), of which functions are illustrated in Figure 4. For the a priori distribution state described in phase 1, denoted as θ, *physical state parameters* are based around:

- The physical interaction between components, e.g. a touch condition of a controlled clearance;
- The energy is transferred from one component or medium to another across the interface;
- The information transferred at the interface;
- The material used, or is there, or there is a material exchange at the interface;
- The functional interdependence between physical components of the powertrain system.

In phase 1; classification procedures in the analytic model associate heterogeneous datasets with a community of homogeneous data-objects (categorised data-objects), inheriting aggregations with other components, thus other empirical datasets, through their parent component hierarchical structure. For example, a *Sensor* contains many *Datasets*, which in turn contains many *Values*. A *Powertrain* is composed of physical *Components*. A *Component* contains a number of *Sensors*. Sensors represent the physical condition of individual powertrain components. As a community, they represent the powertrain system. In phase 2; categorisation commutes each dataset towards the a posteriori distribution (ML) process via the a priori distribution state. Through the dataset data-object association, recorded changes in the physical condition of components, and their environment, are processed in phase 3, by means of one or a combination of physical state parameters (reciprocity of action and reaction), of which functions are illustrated in Figure 5. The determinations of which reappraise life prediction evaluations:
In phase 3; the a posteriori distribution procedure occurs in the superstructure (the 'synthetic model' – see the higher segment of Figure 6), built on top of the substructure. The a posteriori distribution process distinguishes the permanent from the changeable characteristics of the system, i.e. physical components from modes of physical components, in the sense that modes are determined by recourse to dynamic datasets and experimentation. For example, the Bayes equation (4.1) is the subjective distribution associated with θ after performing the experiment, and is the result of a synthesis between the a priori information and the sensor data. The associate between the a priori state and the flow data streaming from over-the-air is essentially the homogeneous manifold of the whole product lifecycle.

$$f(\theta/x) = f_p(\theta) = \frac{f_a(\theta)f(x/\theta)}{f_1(x)} \qquad (4.1)$$

The Bayesian process of combining the a priori information with the experimental data is described as follows: let x be the vector of the times-to-failure and $f(\theta/x)$ its probability density function (Catuneanu & Mihalache, 1989), a function of a continuous random variable, whose integer value across the lifecycle of the vehicle is provided by the homogeneous manifold (sensors/a priori state).
The motivation for this framework derives from an unintended consequence that has emerged in CPS design. The dual domains of discourse with incompatible 'object' representation, disjointed at the object-level of CPS design are the physical component and the data-object. The object-independent concept is applied by inserting arithmetic-based axiomatic rules into data-objects. Computation is the action of mathematical calculation, mathematics have composition, and the axioms provide composition to data-objects in the form of temporalised concepts (arithmetic). This research paper aims to form an interconnectivity between data-objects and determinations of data-objects, whereby upon the notion of permanence, rests the concept of change, and only the condition, or quality thereof changes, thereby, the integrity of object connectivity is predetermined and omnipresent.
In a data warehouse raw data sits on a virtual shelf (i.e. cube), just as raw materials are stored in a warehouse; in an analytics factory model, raw data is manufactured with predefined logic, based on the categories and axioms, used to support further more complex mathematical logic in higher level domain applications such as machine learning-based modelling. Current technology affords us the means but

not the wherewithal for autonomous ML, in the sense that, the tools currently available are deficient in effect. State-of-the-art modelling frameworks are reliant on sophisticated algorithms to retrieve data for operation use. In effect, engineering big data takes raw data out-of-the-box and into the realm of application, in other words, it no longer merely sits redundant in a data warehouse, rather, in an analytics factory, it sits pre-processed with a collection of association attributes established in advance.

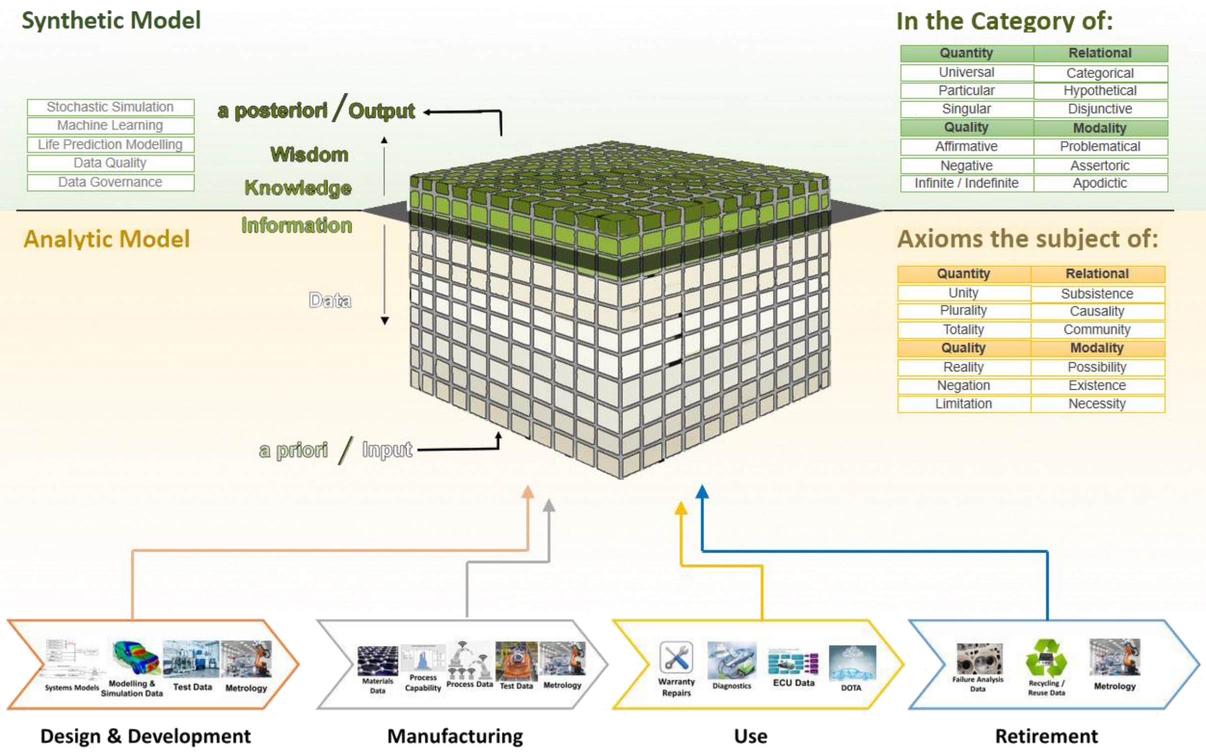

Figure 6. The automotive analytics factory model

Endowed with universal principles that define the quantity, quality, relationship, and modality of domain objects, the analytics model is a centralised and integrated data resource, like an index catalogue, it is searchable, with regional, rural, and remote system connectivity. The motivation of this object-level methodology is to reduce processing time and effort through the synchronous construction of contextually intelligent data-objects, creating a symbiotic relationship and co-dependency amongst system objects by engineering big data them to retain the character of a temporal hypothesis. The analytic model is the substructure, modelled on the principles that govern sensory perception. The synthetic model is the superstructure, modelled on the principles that govern contextual understanding (see Figure 6). Albeit based on a philosophy grounded in pure mathematical structures and Newtonian physics, the initial objective is to apply mathematical logic to data structures, and in stage two, to design the superstructure to support autonomous ML methodologies based on conceptual understanding.

Figure 6 illustrates the holistic teleological endeavour of the automotive analytics factory. The subsystems that support the model deal with the ongoing day-to-day operations. It is within this function that we apply design, development, and manufacturing resources, including warranty data, diagnostics data, field operational parameter data, durability test data, engine development test data, manufacturing data, and external data. Stochastic data is assessed under supervisory conditions before applying new ML strategies and models to the automotive healthcare domain. Such applications are static, and change only in quantum leaps. In order words, the ML models remain constant form day-to-day, and only change after domain specialists observe, analyse and direct such activity. Models adapted for the specific purpose of individual vehicles monitor the condition of powertrain functionality and of its environment. In our example models measure powertrain function reliability, evaluate potential actions for recourse, and perform other domain specific tasks. Model governance monitors tactical ML decisions, whom makes them and holds domain supervisors accountable for altering models and developing new insight.

6. Conclusion and further work

In this research paper we have addressed a multi-disciplinary problem in CPS design that fosters contradictory presuppositions emerging form the information systems and the physical systems domain. Addressing the problem at the conceptual level, this paper proposes a solution-neutral approach that confronts the challenge at the object-level of design, introducing a domain-independent constituent-based framework. The motivation behind the framework is driven by the need for a higher level of abstraction that supports explicit uninhibited continuity, bounded by a digital twin of the physical system and cross-system applications. Further work will begin towards a formal application of object representation in CPS, using model-based ontological reasoning, from a phenomenological view point. The primary objective towards developing an ontology of an object, is to establish a cross-disciplinary "artefact" that will reinforce the universe of discourse. This paper presents a theoretical basis towards applying the domain-independent artefact to operational systems, using an omnipresent set of constructs and rules. Our proposal offers a rigorous framework for multi-disciplinary design (see Section 5) using an axiomatic-based (see Section 4) categorisation technique (see Section 3), in what we refer to as engineering big data. These processes take place in the analytic model of the Analytics Factory. The analytic model is the substructure of the framework, developed to consolidate the heterogeneous disposition of the CPS domain, through a series of operational functions (see Figure 4). The superstructure, built on top, is the synthetic model, developed to determine by recourse to experience, acquired from empirical particulars provided by sensors, of which functions (see Figure 5) are the subject of further research. Alongside the ontological model of an object, future work will also focus on the axiomatic concepts of time, to advance the synthesis of representation as a homogeneous manifold across the analytic synthetic threshold, in the sense that the axioms retain the character of a temporal hypothesis in the analytic model, referring to all possible relationships between objects and events in the synthetic model, and a temporalised conception of cause and effect.

Acknowledgement

The research presented in this paper is part of the Intelligent Personalised Powertrain Healthcare research project, in collaboration with Jaguar Land Rover.

References

Bencomo N., Grace P., Flores C., Hughes D., Blair G. (2008). "Genie: Supporting the Model-Driven Development of Reflective, Component-Based Adaptive Systems," Proc. 30th Int'l Conf. Software Eng. (ICSE 08), ACM Press, 2008, pp. 811-814.

Bruegge B., Dutoit A. H, (2010). "Object-Oriented Software Engineering Using UML, Patterns, and Java". Publ. Prentice Hall, 3rd edition, pp. 121 - 170.

Campean, F, Neagu, D. (2016) "Intelligent Personalised Powertrain Healthcare: an R&D Roadmap for Advanced Powertrain Reliability", AAA White Paper, Automotive Research Centre, University of Bradford.

Catuneanu V. M., Mihalache A. N., (1989). Reliability Fundamentals. Elsevier Science Publishers.

Eisenbart,B.(2014). Supporting interdisciplinary system development through integrated function modelling. PhD thesis. University of Luxembourg.

Eisenbart, B., Gericke, K., Lucienne, T.M., Blessing, L.T.M., & Timothy, C.M. (2017). "A DSM-based framework for integrated function modelling: concept, application and evaluation", Res in Eng Design 28(1), 25–51, DOI: 10.1007/s00163-016-0228-1..

Erden, M. S., Komoto, H., Van Beek, T. J., D'Amelio, V., Echavarria, E., Tomiyama, T., (2008), "A review of function modelling: Approaches and applications", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 22, pp. 147-169. DOI: 10.1017/S0890060408000103

Floch J., Hallsteinsen S., Stav E., Eliassen F., Lund K., Gjorven E. (2006). "Using Architecture Models for Runtime Adaptability", IEEE Software, March 2006, pp. 62-70.

Inmon, W. H. (#), "Data Warehousing and Online Analytical Processing", In: Han, J., Kamber, M., Pei, J. (Ed.), Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, USA, pp. 126-128.

Kant, I. (2008). The Critique of Pure Reason: Of the Schematism of the Pure Conceptions of the Understanding. Cosimo, New York, 2008 (Originally published 1787). Book II. Chap. I, pp. 100 - 106.

King, A.M., & Sivaloganathan (1998). Development of a methodology for using function analysis in flexible design strategies. Journal of Engineering Manufacture 212(3), 215–230.

Kutz M. (2017), Contextual Intelligence – How thinking in 3D Can Help Resolve Complexity, Uncertainty and Ambiguity. Palgrave Macmillan, pp. 9-20. DOI: 10.1007/978-3-319-44998-2

Lehner, W. (2002), Datenbanktechnologie fur Data-Warehouse-System, Dpunkt-Verlag, Germany.

Lehner, W., Sattler, K. (2013), Web-Scale Data Management for the Cloud, Springer-Verlag, New York.

Morin B., Barais O., J´ez´equel J., Fleurey F., Solberg A. (2009) "Models at Runtime to Support Dynamic Adaptation", IEEE Computer Society, 2009, pp.46-53. DOI: 10.1109/MC.2009.327

Pahl, G., Beitz, W., Feldhusen, J., & Grote, K.H. (2007). Engineering Design: A Systematic Approach, 3rd ed, Springer, London, https://doi.org/10.1007/978-1-84628-319-2.

Recker J. C., Bjorn N. (2008). "Epistemological perspectives on ontology-based theories for conceptual modelling", Applied Ontology 3(1-2):pp. 111-130. DOI: 10.3233/AO-2008-0045

Spivak, D. I. (2014), Category Theory for the Sciences, MIT Press, Cambridge, Massachusetts.

Srinivasan, V., Chakrabarti, A., & Lindemann, U. (2012). "A framework for describing functions in design". Design Methods: Proc. 12th Int. Design Conf., pp. 1111–1122, Dubrovnik, Croatia, May 21–24.

Yildirim, U., Campean, F., Williams, H. (2017). "Function modelling using the system state flow diagram", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol 31, No. 4, 413-435, DOI: 10.1017/S0890060417000294.