Institution: BRUNEL UNIVERSITY (H0113)

Unit of Assessment: 11 – Computer Science and Informatics

Title of case study: Improving Data Models in Operational IT Systems

1. Summary of the impact (indicative maximum 100 words)

The impact of this work stems from the provision of better quality information models, and is manifest via: (a) reduced cost through improved reuse and less rework; (b) improved system interoperability; and (c) enhanced assurance and checking that information requirements are supported by the resultant systems. The approach has been applied in commercial environments, such as Shell (UK), where it has reduced development costs by up to 50% ($1m in one case). It has also been applied in the defence environment, forming a part of underpinning standards currently being implemented by the UK and Swedish Armed Forces.

2. Underpinning research (indicative maximum 500 words)

The driver for the research was to resolve well-known but difficult problems in data and systems engineering related to the semantic interoperability of software systems and integration of the underlying data. Existing approaches were assessed as either too technology-centric (focussing on syntactic conversions and data formatting), or lacking in providing the necessary semantic accuracy in modelling real-world domains (e.g., business domains). As a consequence, an approach was sought that was underpinned by a sound ontological theory of real-world semantics that could be used across different domains and organisations to ground their business models in a common manner.

The Business Object Reference Ontology (BORO) is the approach developed through the research, the impact of which is described in this case. Theoretically informed by philosophical work on ontology and semantics, BORO derives from projects carried out in the late 1980s and early 1990s at KPMG by Chris Partridge (and subsequently a member of the Brunel team). Since the adoption of the core ideas of BORO, the underpinning research at Brunel during the period 1999-2013 has evolved the approach by: (a) empirical application to the banking and insurance sector as part of the EPSRC funded project titled “Semantic Integration Framework” (see section 2, reference [6]); (b) developing core ontological patterns used within the approach, such as geo-political regions, products, contracts and the like; (c) producing formalisations of classification systems via set theory (powertypes specifically); and (d) extending the reach of the approach to include Web services and, most recently, business processes (reflected in the EPSRC award EP/K009923/1 – section 2, reference [7]). Current work is also addressing implementation of the approach (including resulting models) via graph databases.

Specifically, BORO integrates perdurantist (or 4D) ontology, set theory, extensionalism and possible worlds theory. A primary reason for adopting these ‘kernel’ theories is to more precisely identify objects and relate them with one another (the nub of conceptual modelling). The foundational ontology is partnered with a process called ‘Content Sophistication’, which is derived from the practical application of BORO (e.g., legacy re-engineering, semantic web services, big data, and business process modelling). Content Sophistication is characterised by four core phases that: (i) partition empirical data (segmentation); (ii) identify implicit and explicit business content (interpretation); (iii) improve that content by removing differences between it and the real (business) world (sophistication); (iv) and bringing pattern fragments together (harmonisation).

The important aspect of Content Sophistication is that it is an *empirical* approach to semantic discovery. All ontological models are derived from data found in legacy systems or other sources (e.g., publicly available data, process spreadsheets, etc.). The process is akin to scientific discovery whereby a ‘theory’ of the real world domain is empirically-modelled based on what we discover in the data and *not* about the subjective assumptions that we make as modellers. The resulting ontological models are then continuously tested/evaluated against new data of further re-engineering projects – which empirically finds inadequacies of previously modelled patterns (or not) – allowing further generalisation where appropriate. As the models are repeatedly tested over time, experience shows that saturation points are achieved – models reach a stable state whereby there is a high level of confidence that they are sufficiently ‘mature’ and are capable of fitting any new data that relates to the specific ontological pattern. Over the period, the Content

Sophistication approach (as a methodological refinement of BORO) has been refined at Brunel (as can be see in references [1] and [4]), applied and extended to business process patterns (e.g., reference [7]) and, through on-going work, has semi-automated the approach and implemented ontological patterns via graph databases. Common modelling mistakes in commercial systems have also been identified (as shown in reference [2]).

Key members of the Brunel research team across the period were:

- Sergio de Cesare: Research Fellow (1999 – 2002) on GR/N01897/01; Lecturer (2002 – 2013); Reader (2013 onwards). Principal Investigator on EP/K009923/1 (2013 – 2016)
- Aseem Daga; Research Assistant (1999 – 2002) on GR/N01897/01
- Mark Lycett: project manager on GR/N01897/01; Senior Lecturer (2003-2006); Reader (2006-2010); Professor (2010-present). Co-Investigator on EP/K009923/1 (2013 – 2016)
- Chris Partridge; visiting Research Fellow 2007-present.

3. References to the research (indicative maximum of six references)

[1] de Cesare, S., Foy, G., Partridge, C. (2013), Re-engineering Data with 4D Ontologies and Graph Databases. *CAiSE 2013 Workshops*, LNBIP, Springer-Verlag Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-38490-5_29

[2] Lycett. M. and Partridge, C. (2009), The Challenge of Epistemic Divergence, *Communications of the ACM*, **52** (6), pp. 127-131. http://dx.doi.org/10.1145/1516046.1516079

[3] Bell, D., de Cesare, S., Lycett, M., Iacovelli, N. and Merico, A., (2007), A Framework for Deriving Semantic Web Services, *Information Systems Frontiers*, **9** (1), pp. 69-84. http://dx.doi.org/10.1007/s10796-006-9018-z

[4] Daga, A., De Cesare, S., Lycett, M. and Partridge, C., (2005), An Ontological Approach to Sophisticating Legacy Business Content, *38th Hawaii International Conference on Systems Sciences,* January 3-6, Big Island, Hawaii.

[5] Partridge, C. (2005), Business Objects Re-Engineering For Re-Use*,* 2nd edition, The BORO Centre, London.

[6] EPSRC award grant reference GR/N01897/01: Semantic Integration Environment (Paul and Macredie), October 2000 – October 2003, £1.34 million.

[7] EPSRC award grant reference EP/K009923/1: Empirical Modelling of Business Process Patterns with Ontologies (de Cesare and Lycett). June 2013 – June 2016, £398k.

4. Details of the impact (indicative maximum 750 words)

As noted above, much of the impact of the research comes from the practical application of the Content Sophistication process in empirical settings. Its application allows general ontological patterns to be developed that, in turn, improve the application of BORO in subsequent empirical settings. So, application provides model-based solutions to specific areas from one perspective, but in the background develops more general patterns that are applicable both within and across domains of application. Products, for example, might be a model applicable within a given context (e.g., a Unisys banking system). Naming or geo-political regions (instances of which were developed through the research undertaken at Brunel) are examples of patterns that are cross-domain.

Four interconnected areas of impact arising from the research are detailed below, along with evidence/indicators of the significance and reach of each impact area:

1. At *Shell Oil* (see corroborating source S1) the evolved BORO approach resulting from the Brunel research was applied to conceptual modelling of the downstream oil and gas production process between 2006 and 2008 (including follow-on work to assess and support the integration of several local SAP instances into a global instance). The significance of the work was in the provision of a reference model for Shell covering the Global downstream business (from oil tanker through refinery to petrol pump). Test application projects based on the reference model showed a good

level of reuse, providing time and cost savings on those projects. The reach of the work is evidenced in the standardisation of asset data in the oil and gas industries – the 4-dimensional aspects of the model are core to ISO15926 (see corroborating source S1). It is estimated that use of the downstream model allowed *Shell* to reduce the cost of development by at least 50% ($1m).

2. The underpinning research had an impact on *ISO 15926*, a standard for data integration, sharing, exchange and hand-over between computer systems (see corroborating source S1). The impact is primarily through the standards adoption of 4D (perdurantist) data modelling, stemming significantly from the practical application of the outcomes of the Brunel research in the Oil and Gas industry (e.g., at Shell). Work in relation to ISO 15926 was carried out across the period 1994 to 2006; the standard has been in place across the eligible period 1.1.2008 to 31.7.2013.

3. The underpinning research has had an impact on the work of the *International Defence Enterprise for Exchange Group (IDEAS*) (see corroborating source S2), which provides a data exchange format for military Enterprise Architectures. Again, the influence is through the adoption of 4D data modelling, though the group is also explicit publicly in its adoption of BORO as the means of ontology development. Work in relation to IDEAS started in 2006 and has been used as a semantic extension for the Ministry of Defence and U.S. Department of Defense Architecture Framework (MoDAF and DoDAF) (see corroborating source S3).

4. *MODEM (MODAF Ontological Data Exchange Model*) is the most current foundational framework for enterprise architecture, providing a common way to plan and organise information about structures, behaviour and capability. The ontological data exchange mechanism is based on IDEAS so, again, the use of BORO is explicit in these frameworks, representing a global impact in the defence area. Development is being led by the Swedish Armed Forces (with the MoD, DoD and Canadian Defense force participating), who propose benefits in terms of improved operational capability for collaborating forces, greatly reduced maintenance costs and increased choice and quality of tooling environments (see corroborating source S4).

5. Sources to corroborate the impact (indicative maximum of 10 references)

S1 – Letter received from Reference Data Architecture and Standards Manager for Shell Downstream.

S2 – http://www.ideasgroup.org/5Methodology/ provides an explicit acknowledgement of BORO as a foundation for the IDEAS Group work.

S3 – https://www.gov.uk/mod-architecture-framework : the document 'Ontologies and their use in MoDAF', available as part of the MoDAF Architectural Framework documentation via this link, is explicit in its acknowledgement of BORO as a foundation.

S4 – Letter received from Head of Architecture, Frameworks and International Co-operation, Swedish Armed Forces Headquarters Supreme Commanders Staff, Joint Development Department.