

Toward A Semantic Vocabulary For Systems Engineering

Paola Di Maio

University of Strathclyde

DMEM

75 Montrose Street, Glasgow

paola.dimaio@strath.ac.uk

paola.dimaio@gmail.com

ABSTRACT

The web can be the most efficient medium for sharing knowledge, provided appropriate technological artifacts such as controlled vocabularies and metadata are adopted. In our research we study the degree of such adoption applied to the systems engineering domain. This paper is a work in progress report discussing issues surrounding knowledge extraction and representation, proposing an integrated approach to tackle various challenges associated with the development of a shared vocabulary for the practice.

Categories and Subject Descriptors

H. Information Systems

General Terms

Management, Measurement, Documentation

Keywords

Systems Engineering, Ontology, Semantic, Wiki, Vocabulary

1. BACKGROUND

Systems engineers are not new to mechanisms for formalizing knowledge, yet when it comes to defining the systems engineering domain itself - in terms of theory and practice, vocabularies, axioms, shared conceptualizations - efforts are still fragmented and largely not in the public domain. Despite being increasingly widely practiced in all fields and across many disciplines, systems engineering is still associated with the defense sector, especially in the UK, where knowledge sharing is not encouraged, and constrained by a secretive culture. This 'cultural gap' becomes evident when comparing the amount of technical literature and resources published in the USA by the DoD, while in the UK, MoD related research is not publicly accessible nor their publication schemas visible.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WIMS'11, May 25–27, 2011, Sogndal, Norway.

Copyright 2011 ACM 978-1-4503-0148-0/11/05... \$10.00.

Thanks also to the adoption of the web as 'always on' unconstrained knowledge exchange hub, 'system of systems' architectures are taking shape at various levels (social, technical, topical, data and information levels) as the result of various combinations and meshups that constitute the fabric of the open web, and a 'global' web centric information sharing culture is trying to become mainstream. The systems engineering body of knowledge is trying to find its shape and place in this 'open web' landscape, facing many challenges:

"The discipline of systems engineering (SE) has been recognized for 50 years as essential to the development of complex systems. SE is still treated primarily as heuristics learned by each practitioner during the personal experimentation of a career. The heuristics known by each differ, as shown by the fractured development of SE "standards" and SE certification. As a result of this heuristic understanding of the discipline, it has been nearly impossible to quantify the value of SE to programs. [Sheard 2000] The differences in understanding, however, just as typically result in disagreement over the level and formality of the practices to include. [Shenhar 1997] In response to the uncertainty in the heuristics, some efforts have been made to identify "leading indicators" that provide early tracking information as to the worth of a project's SE efforts" [1]

2. THEORETICAL CONTEXT

This research is part of a three year investigation into Knowledge Sharing and Reuse in Systems Engineering, partly funded by EPSRC in the UK. The results of the exploratory phase (2009-10) indicate that most of the questions pertaining to the problem space share many of the characteristics of the 'wicked problem' [23] demanding a *systemic* approach to enquiry. Among the instruments devised to support the integration of disparate angles of the field work, a morphological box (aka Zwicky box) is developed (see Table 1). "Morphological analysis is a group of methods that breaks down a system, product or process into its essential sub-concepts, each concept representing a dimension in a multi-dimensional matrix. Thus, every product is considered as a bundle of attributes. New ideas are found by searching the matrix for new combination of attributes that do not yet exist." [24].

Our morphological field includes Theory, Methodology and Artefacts with Cognitive, Organisational and Technical aspects.

Cognitive

Language, level of skill required to use/adopt the knowledge

Organisational

Policies and management practices that promote knowledge sharing and reuse

Technical

Choice of codification, knowledge representation techniques and standards adopted

Selected elements from the morphological box are then used to navigate the research space. As explained in more detail in section 4 in this paper, the complex mix of dependencies that contributes to the challenge is reduced to a combination of governance, process, structure, and mapped to the corresponding section of the morphological box, as shown in the table below.

Table 1: Morphological Box

	Cognitive	Organisational	Technical
Theory			
Methodology		GOVERNANCE	PROCESS
Artefact	STRUCTURE		

One of the overall aims of this research is to start devising at least some rudimentary cornerstones of a 'shared knowledge model' for systems engineering, with particular emphasis on 'shared artefacts'.

Table 2: Knowledge Sharing Continuum

LIFECYCLE PHASE	KNOWLEDGE ASSET 'document, specification'	FORMAT	NOTATION/FORMALISM	SHARING MECHANISM
analysis	requirements specification	narrative structured text	natural language, pseudo-code	
design	system diagram	diagram	ER, DF, UML	
development	system specification	narrative structured text	Natural language pseudocode	image word document
installation	operating manual user guide	narrative diagrams	Natural language graphics	spreadsheet pdf html
testing	test plan	structured text	natural language charts	xml rdf
acceptance	contract	narrative	natural language	owl
support	user feedback tickets	narrative	natural language	other
fourthark				

Knowledge artefacts are considered 'explicit knowledge assets' by Nonaka and Takeuchi [25], and grouped into two categories, conceptual and systemic. One view of such artefacts follows the system lifecycle phases: for each system development phase, a number of technical document correspond that qualify as 'knowledge assets', which need to be explicit and formalised according to common standards for this knowledge to be shareable. Table 2 synthesizes the result of this analysis that associates lifecycle phase, corresponding knowledge assets, their commonly used formats, relevant notation and formalisms (as used in the practice) and their corresponding standard sharing mechanism.

3. TOWARD A SE KNOWLEDGE MODEL

Structuring and exposing information models is required to facilitate knowledge exchange and reuse on the web, however information architectures and knowledge models, also referred to as 'ontologies', in systems engineering seem to exist within very tightly controlled organisational boundaries.

Table 3: Summary of provisional survey results

	1	2	3	4	5	6
Sysmo sysmo.org		unknown	no	yes	no	no
Sweet /sweet.jpl.nasa.gov/		unknown	no	yes	yes	no
U Texas		unknown	unknown	yes	no	not known

1. Ontology name, URL 2. development method declared? 3. implementation independent formalisation? (UML, for example) 4. OWL encoding? 5 Accessible on the web? 6 Open License?

A systematic survey has been started as part of this project, data collection is currently underway [2]. Some of the common problems with existing SE ontologies, or fragments:

1. They are 'top down' – one single researcher/team working for a commercial company or other organisation proposes one single ontology – this results in a single view of the world being proposed
2. They are not publicly available, accessible and/or do not benefit from a public license to facilitate its reuse (not 'open')
3. The ontology development process is not declared (the method followed to develop the ontology is not explicit nor documented) and there is no provenance tracking to explain how and where from formalisms such as axioms or lexemes contained in the ontology are derived.
4. The ontology is tightly coupled with its encoding, for example, many ontologies on the web only exist in OWL. (the ontology web language standard published by the W3C consortium) [3]. No implementation independence possibly results in heavy encoding bias and not good practice. See Gruber 'Minimal Encoding Bias' principle.[4].

An ontology, intended as shared conceptual schema, should be understood and to some extent even manipulated by a variety of stakeholders with different skill sets (not necessarily skilled in the implementation language of choice) and used as part of a variety of architectures with different implementations. Implementation Independence is a well established principle in systems science, and it constitutes the foundation of many system models, as demonstrated in the early days of computing by Child[17]; in Codd [18] the principle applies as 'data independence and in Model Driven Architectures translates into 'platform independence' [19].

One of the difficulties in developing highly specialised semantic web artifacts is that not only 'ontologists' from different schools can sometimes disagree about the choice of representation method and approach, but often the subject matter specialists - in our case systems engineers - who are not necessarily familiar

with the ontology development method or encoding of choice (nor want to be) become excluded from participating the ontology development process, when it becomes dominated by the prevailing ontological 'sect', rather than being an open, transparent, accessible collaborative process. A variety of ontology engineering approaches are available, and different systems and systems engineering ontology efforts are being discussed, and possibly initiated, however to the author's knowledge there is no realistic time frame for this to happen. To ensure that subject experts maintain their ability to understand the knowledge model, and can contribute expertise to its development and deployment – a rigorous but minimal approach to distributed ontology development was presented at INCOSE EUSEC [5], which synthesizes best practices in the field while making ontology development process agile, integrated, iterative and open to a variety of stakeholders. Another possible method to derive the backbone of a knowledge model sometimes referred to as 'bottom up' consists of performing automated knowledge extraction from the corpus (body of knowledge). A discussion of challenges and the way forward constitutes the main focus for the rest of this paper.

4. MOVING TO A WIKI

A 'systems engineering' ontology may be a long term, open ended project, of certain risk and uncertain feasibility, given the massive scale and heterogeneity of practitioners and sources knowledge. However there is enough consensus among practitioners deriving a basic vocabulary of reference for the systems engineering body of knowledge could be both feasible and desirable. The NASA [6] and INCOSE [7] respective Handbooks of Systems Engineering have been instrumental to the evolution and standardization of the contemporary practice of systems engineering worldwide, and represent knowledge reference for the profession. These invaluable resources however to date are still 'hard coded' and shared on the web as 'PDFs', which are not the format of choice for semantic representation of information and knowledge. One of the obvious starting points toward the 'quantification' of the SE body of knowledge is to publish the respective sources of knowledge to a wiki, ideally, a wiki platform that supports semantic structure. A wiki is a standard online distributed collaborative environment, by default capable of supporting flexible and easily implemented vocabularies and taxonomic hierarchies. In the case of semantic wikis a basic taxonomic structure is provided by a controlled vocabulary to help the consistency of the terminology and the semantic structuring of the concepts. In addition to a lexical definition of what the term means, values and properties are associated to each term to assert its relation to other terms. Different semantic wiki environments offer different advantages. We evaluated to some extent some of these platforms (Semanticmediawikiticmediawiki, Knoodl Referata) The Knoodl platform offers the advantage of a SPARQL endpoint (a page where the content of the wiki can be queried). There is no single method or approach to migrate a large body of knowledge to a collaborative semantic platform, and it is a non trivial task. It is envisaged that at a minimum requirements are a staging strategy, an editorial process and a governance model capable of accommodating a highly distributed nature of the team.

4.1 Challenges to Wikification

The first challenge when moving to a wiki is often to provide justification for what is essentially a cultural change from a 'closed world' perspective, to the open dynamic world,' which comes with many implications. This not only reflects in the logic of the technical system being implemented, but also in the organisational and management setup of development practices and documentation itself. Such a justification is already widely provided in the benefits of adoption of social media as knowledge networks as a whole [8]. More specifically 'wikis' as distributed platforms flatten an organisational /editorial hierarchy by allowing multiple simultaneous users to create and edit knowledge entries.

The requirement for a highly distributed platform capable of supporting a network of contributors to co-create and contribute to the development and maintenance of a corpus, can conflict with the requirement for minimal validation and version control over the same. A diverse range of additional challenges lie ahead, falling into one or more of the following categories:

- What's the best way to migrate the content from hard coded documents to a SMWiki environment
- What kind of structure should the wiki provide to enable meaningful representation of the content
- What model of governance is suitable for a distributed multiauthors team that would allow both max flexibility to encourage contributions, and that would also support sufficient validation for the quality of the contributions?

5. ENTANGLEMENT

Loading the corpus to a wiki requires some decision in relation to the 'structure' which means, how to best organise the corpus depends on how the corpus structure is defined. Subsequent ongoing edits and editorial changes to body of knowledge after it has been opened up on the wiki, will be determined by what editorial process is in place. The editorial process in turn depends on what governance model is chose .These interdependencies between various steps are what we refer to as 'entanglement' (Image 1)

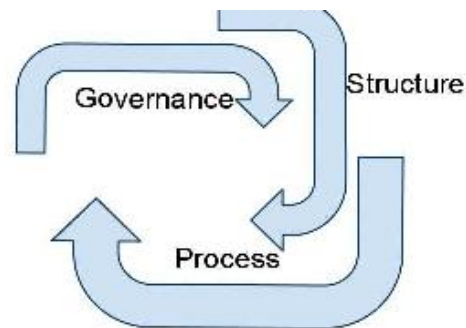


Image 1: Entanglement of governance, process, structure

Entanglement can be approached by taking small iterations to each aspect of the solution design making sure that each output (decision) of one aspect is related to its corresponding input from

the factors where the dependency arises. For example, what platform to choose may depend on the functionality required, and some costing estimates, but the costing is affected by how task is organized and managed, therefore some decisions that fall in between the editorial process and its governance. Solving the entanglement can be done by

- a) Identifying and mapping all the key factors and their dependencies
- b) Constructing a project schedule and a plan build around them, in what is essentially a very 'agile' style.

5.1 Top Down from the Bottom up

The tradeoff between “Top-down versus bottom-up” information structures represents a classical dilemma in knowledge organization efforts. Top-down refers to the development of an ‘authoritative view of an information model, that models the viewpoint of an individual, a team or an organisation. Bottom-up approaches are associated with “folksonomies” [9] where information structure is not derived from preconceived categories but rather emerges from the organic association of user-generated content and, where available, their metadata (tags). The lack of an intrinsic logical structure can make folksonomies difficult to apply and manage systematically at an organizational level, although it offers different perspectives and conceptual paths to navigate an information structure.. Denise Bedford, thesaurus manager and senior information officer at the World Bank, advocates a method that she calls ‘*top down from the bottom up*’ where an information structure is first “extracted” through inventorying and systematically analyzing sources scattered around heterogeneous and unstructured data (for example, a wiki used as a knowledge repository by a community of practice).[10] The output is then integrated using “domain experts” inputs and then matched to a more canonical, preexisting view of the domain in question. The ultimate goal of a knowledge mining and representation exercise is to devise a global map of a given knowledge domain, the most formal representation of which is known as an “ontology”; But it can start as a simple concept extraction, which results in a vocabulary, and mapping exercise. The First step of Bedford’s method consists of analyzing the content objects therein contained and “automatically” (or semi automatically, as the process will need to be vetted by a human expert in most cases) generate metadata for each individual object. The intent is that the metadata will be used to enrich search — as well as become the spinal chord of a domain’s vocabularies (or glossaries) and the basis for a future more refined conceptual structures. Using appropriate text analysis tools, it is possible to semantically “deconstruct” each content object — for example, open up the document and find keywords — and generate XML-formatted metadata for each object. Such metadata can then be embedded or associated back into the original documents. It can also be used to generate a “metadata store,” a catalog of all metadata that can be used as the backbone of an information structure. The resulting “lists” — likely a vocabulary or thesaurus — can be navigated according to different points of view and can be rendered in XML-RDF or OWL files. These can then be used as a first-draft ontology. The “bottom-up” part of the process consists of deriving the top-level concept clusters from analyzing the relevant concepts and keywords in a pool of user-generated heterogeneous resources.

This approach is then balanced and weighted against “top-down” domain expertise and professional opinion and then finalized into a final structure that can be quite comfortably used to structure and classify previously unstructured content [11].

6. PARSING THE CORPUS

Language Engineering (LE) may be defined as: ...the discipline or act of engineering software systems that perform tasks involving processing human language. Both the construction process and its outputs are measurable and predictable. The literature of the field relates to both application of relevant scientific results and a body of practice. [Cunningham] There are different approaches and techniques, platforms and environments. We take the ‘most economical’ route, and prototype a text extraction solution as follows [12].

6.1 Identifying the Body of Knowledge

In our initial experimentation, we select as sample ‘sources of knowledge’ from the INCOSE/SEBOK working draft. We ‘chunk’ the text into chapters, and perform text analysis on selected ‘chunks’. This is to keep the effort contained and manageable. We then compare the output of the extraction (Tables 5 and 6) with the glossary of terms, where available, compiled manually by authors and editors (Table 4), and also compare the output of different tools and environments, for example Alchemy and Open Calais. Further planned text extraction experiments include using various configurations of tools such as GATE, Google Refine and Topbraid Composer.

Table 4: Manual Vocabulary

Behavior
Emergence
Environment
Hard System
Hierarchy
Network
Paradox
Respondent System
Situation System
Soft System
Structure
System
System Asset
System of Interest (SOI)
Systems Thinking
Thematic System
Togetherness

6.2 Output

On a selected portion of corpus, the first set of reference is the glossary compiled manually by authors and editors contains initially 16 terms, which are unstructured, and no metadata is associated with them. The same portion of corpus, parsed using semantic extraction tools, yielded consistency more granular terminology (longer and more detailed list of glossary entries than what extracted manually by authors), and some categories, pointing to the emergence of a metadata structure.

Table 5: OpenCalais

Entities:
 Company
 Facility
 Industry Term
 Organization
 Person
 Position
 Product
 Programming Language
 URL
 Events & Facts:
 Company Product
 Natural Disaster
 Person Career

Table 6: Alchemy

FieldTerminology (11)
 Person (38)
 Company (4)
 Organization (6)
 NaturalDisaster (1)
 Automobile (1)
 Technology (2)
 Concept Tags (8)
 Tags (30)
 Category: Science & Technology
 (confidence: 0.85)

This comparative analysis of different tools detail is documented online [13] and was achieved with contained effort (approximately one hour's work, and limited expertise). In this demo we sample automatic tools to demonstrate the benefits of semantic content extraction, and envisage a much larger task ahead before a full vocabulary can be finalized. Extracting the vocabulary from the corpus is per se a rather trivial task, especially when using automated text extraction tools, but decisions as to what text extraction tool and environment offer the most suitable output, and what metadata schema to adopt are non trivial, and are the result of a strategy. Consultation are currently underway with the systems engineering community as to what approach is best to follow, and it is likely to remain an open task for sometime to come. To our knowledge the first public effort to be carried out in the public domain to publish an online a systems engineering vocabulary, and the main contribution to knowledge of this research at this stage.

7. DECONSTRUCTING ENTANGLEMENT

Having established that the vocabulary extraction is feasible, and can be performed with contained effort in terms of people hours, it must be synchronized to other aspects of the task, which in turn correspond to tackling the challenges outline in sections 3. and 4. above.

7.1 Process

In answering the question: *What's the best way to migrate content from a hardcoded document to a SMWiki environment, and then to maintain the future updates?*

At least two different activities are envisaged:

1. to upload the working version of a Body of Knowledge to a wiki after it has been finalized. This can be done as a single step, or a series of steps, for example one chapter at the time.
2. to allow users to make ongoing edits and changes throughout the life of the BOK In our description, a process consists of a series of dynamic steps (as opposed to a model, which is a static representation'). The following processes are required to deploy a body of knowledge to a wiki:
 1. A content migration process. This can be carried out as a phased activity, but would typically involve one or very few material executors of the migration (transform the content of a large document, say a 'systems engineering handbook', to a wiki). This process requires some decisions in relation to the structuring of the content, but the same content structure that applies to a document (say the table of content, sections, subsections) can easily be transferred to a wiki. Its a no brainer.

2. An ongoing editorial process. This is required to make sure that the content in the wiki is maintained /updated. A typical editorial process would consist of the following steps [16]

- Submission
- Review
- Editing
- Publication

In a highly distributed, networked environment, such as a wiki, an editorial process will also have to be 'distributed', and should reflect the governance structure of the task and team, briefly discussed below.

7.2 Structure

"What kind of content structure should the wiki provide to enable meaningful representation of the content?"

It may be useful to define further 'structure'. It is envisaged that a minimum of two levels of structure can be easily inferred from the hard version of the corpus

- 1) **content structure**, for example a TOK (Table of Contents), including chapters, topics and/or subtopics, can be considered as a first level of content structure,
- 2) **'semantic' structure**, essentially what a semantic wiki environment supports – refers to the allocation to each term (glossary entry) of a value and properties (is a concept, or is it a relation, either, or both, for example) as well as some lexical definitions for each term (a description of the meaning).

It is suggested that some elements of text extraction are considered in support of the wikification of a body of knowledge, as described above. However decisions on how to go about '**structuring**' the content are tightly coupled to decision about how to manage the task and how to organise the whole project, so essentially, needs to be aligned to the other processes in the project and governance. For example semantic structuring decisions in relation to each chapter could be initially assigned to each chapter editorial team. Experts (chapter authors) evaluate if the extracted terms correspond to a good enough representation for each concept, and if not make the relevant adjustments To create the semantic backbone of the wiki resource, Key terms identified by the authors/experts for each concept are first loaded to the wiki (either bulk uploaded, or manually entered one by one) as special pages [14] before the chapters or the rest of the body of knowledge is uploaded, which will result in these key terms to become automatically hyperlinked.

7.3 Governance

A governance model represents 'who' does 'what' (possibly also 'when' and 'how'). The characteristics of the technological environments adopted (say a wiki) influence and shape directly the governance model. The governance model required for a large distributed project would necessarily have to take into account the 'virtual' and 'distributed' nature of the team. A governance model can be built by establishing

- what tasks constitute the editorial process
- what roles (team members and other users) have what permissions (what tasks in the editorial process can carry out).
- what sequence/priorities/rules should be applied to the execution of the tasks that. A workflow can then be built, and tags (metadata) can be derived/associated with the content at different stages of the process.

The underlying constant principle is that: whatever in an offline editorial environment (especially in C2 culture) was traditionally 'centralized', in adopting new networked information structures it must become 'distributed', as tentatively illustrated in Image 2.

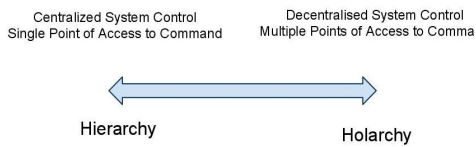


Image 2: Hierarchy vs Holarchy

8. INTEGRATION

The three activities described above (process, structure and governance) ideally should be 'integrated' and synchronized, so that synergy can be created. A possible model of such integration is presented in the diagram below, and it suggests the adoption of different kinds of metadata in support of such integration

8.1 Metadata Architectures

An ideal combination of metadata sets would synthesize and integrate different dimension. metadata can be categorized with different methods. Technical (internal) metadata, Business (external) metadata. Process metadata (Kimball et al) or for example Structural metadata, intended to describe structure of a systems vs Guide metadata general keywords in a natural language. (Bretheron & Singley (1994) [15]

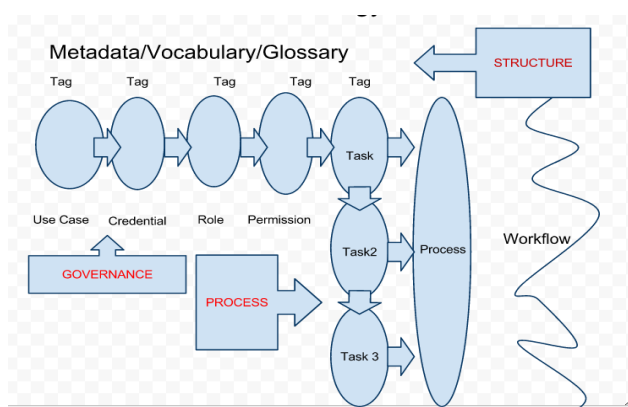


Image 3: Integrating structure, process and governance

After the initial migration of the corpus using a structured approach (entering each term in the glossary as a special page and

then uploading the rest of the corpus to the wiki), it is suggested that (Image 3)

- a distributed editorial process is defined
- each new contribution to the corpus from user is associated upon using existing metadata
- new metadata (tags) are added by users (integrating a 'folksonomy' approach) provided they can be associated to existing categories
- new categories and key terms can be entered by users, and are subject to the editorial process
- the editorial process appropriately stag will result in a collaborative workflow

9. EVALUATION

The first part of the study consists of evaluating the vocabulary obtained by different semantic extraction tools against the corresponding one compiled by hand by chapter authors. In literature this approach is known as 'Gold standard' evaluation (Grefenstette). To our knowledge, no such standardized hand compiled vocabulary exists in the systems engineering domain, and the aim of the current effort is to generate a consistent SEBOK for the systems engineering community of practice at large, which would include a set of lexical terms (glossary entries). The automatically extracted terms using semantic text analysis tools generated consistently greater level of granularity and semantics. A more detailed evaluation of the automatically extracted results was presented at a SEBOK workshop in Tolouse in 2010 [22]. A heuristic evaluation of the quality of the glossary can be performed answering questions, example listed in the Table 7 below [20]:

Table 7: Evaluation of Vocabulary

Coverage/Scope	Is the vocabulary capable of representing all of the concepts used in the chapter?
	Does the vocabulary have the terms necessary to represent the full range of issues?
	Does the vocabulary encompass the terminology used to describe the various procedures?
	Does the vocabulary use terms that are commonly used by SE?
	Does the vocabulary include related terms, as well as synonyms and variant forms of terms?
	Does the vocabulary include modifiers or qualifiers that express the certainty, degree, or severity of a process?
Specificity	Is the vocabulary specific enough to accurately represent the many aspects of SE reality?
	Does the vocabulary capture information in sufficient detail?
Structure	Are the vocabulary hierarchies logical and complete?
	Are the meanings of terms clearly defined?
	Does the vocabulary contain redundant terms?
	Are there explicit rules for combining terms, or for combining terms and qualifiers?
	Does the vocabulary allow for multiple classification of terms, that is, can terms appear in more than one hierarchy?
Maintenance	Does the vocabulary have ongoing institutional support?
	Does the institution or body that developed the vocabulary have stable funding?

	Does this institution or body regularly evaluate and update the vocabulary?
	Does this agency regularly consult with users of the vocabulary on a formal or informal basis in order to obtain feedback?
Useability	Is the vocabulary mapped to other vocabularies used in the practice?
	Does the vocabulary meet the needs of a range of end users?
	Does the user interface facilitate optimal use of the vocabulary with minimal training?

However since the proposed approach integrates heterogeneous techniques and processes, no single standard evaluation method by itself is adequate. In addition, the integrated approach presented in this paper is framed within a socio-technical context: while in a theoretical setting, there are many known mathematical and computational techniques for the evaluation of distributed processes, these are mainly directed to evaluate machine performance ref.Pi

Already some decades ago, pioneers in electronic publishing knew that "Each new medium has started by emulating the medium it replaced. Eventually the capabilities added by the new medium allow it to evolve in innovative, and often surprising, ways.

Alexander Graham Bell thought that the telephone would be used to broadcast music into homes. Thomas Edison thought that recordings would be mostly of speech rather than music. Marconi thought that radio's most common use would be two-way communication rather than broadcast" [21]

An integrated distributed editorial-governance process will require the training of authors team to the principles of semantic vocabularies and wikis. The evaluation of such integrated process will have to combine different measurements, in terms of benefits versus costs of adoption of the proposed approach, and should include parameters such as:

Hardware/software/system resources/cost
Cognitive load (mental effort)
People hours

The combination parameters can be weighted heuristically against some nominal comparative parameter that loss of utility (cost), such as the cost of not having a collaborative semantic vocabulary at all. Developing adequate evaluation methods for socio-technical systems falls under the 'future work'. One of the outstanding benefits of a shared, collaborative semantic vocabulary is that it will serve as a gravitational hub for practitioners and academics to start communicating and exchanging knowledge views. Organisationally, this is a cost, since such argumentation can require consider time investment and effort, in the long term however the quality of the knowledge understanding can only reflect positively and can only broaden the mindset and enhance the overall experience of being part of the systems engineering community.

10. CONCLUSIONS AND RELATED WORK

This paper consists of an analysis and discussion of different factors, conceptual and pragmatic dimensions pertaining to the publication of a high level publicly available and collaboratively developed vocabulary for the systems engineering domain. It identifies the challenge as a 'wicked problem' and proposes an integrated systemic approach to tackling it that attempts to balance the manual effort of authors with the output of automatic text extraction techniques. It outlines an integrated metadata architecture to support the synchronisation of decisions in the various dimensions of the task, such as structure, processes and governance. It introduces a novel integrated continuum to shared representation of knowledge in systems engineering, and it represents, to the best of our knowledge, the first effort of its kind to be carried out in the public domain, aimed at contributing to the development of an open, publicly accessible standard vocabulary for the systems engineering domain and practice.

11. ACKNOWLEDGMENTS

Thanks to Nasa JPL, INCOSE for making their knowledge resources available online, to all researchers and systems engineers who are working to maximise knowledge sharing and reuse in the practice. Part of this research is supported by EPSRC Grant Nr EP/D505461/1 (Networked Enabled Capabilities Research).

12. REFERENCES

- [1] Honor. E Valerdi, R Advancing an Ontology for System Engineering to Allow Consistent Measurement, Proceedings of the Conference on Systems Engineering Research, Los Angeles, CA, 2006 [*.pdf 159KB*]
- [2] ISTCS.org Systems Engineering Ontology, is there anything out there? A survey - <http://tinyurl.com/3ypolu3>
- [3] OWL Specification
http://www.w3.org/standards/techs/owl#w3c_al
- [4] Gruber. T, 1993 Toward Principles for the Design of Ontologies Used for Knowledge Sharing (page 4, principle 4 Minimal encoding bias <http://tomgruber.org/writing/onto-design.pdf>
- [5] Di Maio, P, 2009 Just Enough Ontology Engineering, Cutter Consortim (presented as tutorial, EUSEC 2010)
<http://www.cutter.com/content/bia/fulltext/reports/2009/03/index.html>
- [6] Nasa Handbook of Systems Engineering, 2007
<http://education.ksc.nasa.gov/.../NASA%20SP-2007-6105%20Rev%201%20Final%2031Dec2007.pdf>
- [7] Incoase Handbook of Systems Engineering
www.incoase.org/ProductsPubs/products/sehandbook.aspx
- [8] Coenen T, et al 2006 Knowledge Sharing over Social Networking Systems: Architecture, Usage Patterns and Their Application On the Move to Meaningful Internet Systems

- OTM 2006 Workshops Lecture Notes in Computer Science
<http://www.springerlink.com/content/dlnq4k2l2p54np81/>
- [9] A folksonomy is the practice and method of collaborative categorization using freely-chosen keywords called tags"; the term is attributed to information architect Thomas Vander Wal (see <http://en.wikipedia.org/wiki/Folksonomy>).
- [10] Bedford, D. 2004 Creating a Taxonomy Blueprint to Operationalize Your Enterprise
www.asindexing.org/files/conf/Enterprise.ppt
- [11] Di Maio, P - 2007 Top-Down Information Structures from the Bottom Up Cutter. com Business Intelligence
- [12] Cunningham, H. -1999 A Definition and Short History of Language Engineering. Journal of Natural Language Engineering
- [13] Comparative analysis of text extraction output, working notes <http://tinyurl.com/2vwq5w>
- [14] Semantic Forms, Special pages extension for semanti mediawiki
http://www.mediawiki.org/wiki/Extension:Semantic_Forms
- [15] Metadata architectures, from Wikipedia
<http://en.wikipedia.org/wiki/Metadata>
- [16] Open Journal System
https://wikis.tdl.org/_data/tdl/images/0/07/OJS-in-an-Hour_220.pdfhttp
- [17] Childs, 1970 Feasibility of a set-theoretic data structure : a general structure based on a reconstituted definition of relation Childs, David
<http://deepblue.lib.umich.edu/handle/2027.42/4164>
- [18] Codd, E.F. 1970. "A Relational Model of Data for Large Shared Data Banks". [Communications of the ACM](#)
- [19] Quick Guide to Model Driven Architecture
http://www.omg.org/mda/mda_files/Cephas_MDA_Fast_Guide.pdf
- [20] Toevs, L 1995 An evaluation methodology for clinical vocabularies and evaluation of the read codes
http://www.caais-acs.ca/proceedings/1995/toews_1995.pdf
- [21] H. R. Varian The Future of Electronic Journals University of California, Berkeley
<http://people.ischool.berkeley.edu/~hal/Papers/>
- [22] SeBOK Toulouse Workshop October 2010
<http://tinyurl.com/2vwq5wh>
- [23] Rittel, Horst, and Melvin Webber; 1973 "Dilemmas in a General Theory of Planning," , Policy Sciences, Vol. 4, Elsevier Scientific Publishing Company
- [24] Zwicky, F. 1969, Discovery, Invention, Research through the Morphological Analysis, The Macmillan Company
- [25] Di Maio, P, Formal Analysis Methods for Socio-Technical Systems Workshop Online Proceedings Jun3 2010 SICSA/University of Sterling
<http://www.cs.stir.ac.uk/events/network-analysis/slides/dimaio-analysis.pdf>
- [26] De Geytere, T. (2005) "SECI model (Nonaka Takeuchi)" URL
http://www.12manage.com/methods_nonaka_seci.html