INTRA- AND TRANS - DISCIPLINARY SEMANTIC TECHNOLOGIES

M.N. VAFOPOULOS[1], G.A. GRAVVANIS[2], and A.N. PLATIS[3]

[1] Department of Cultural Technology and Communication, University of the Aegean, Faonos St., GR 81100 Mytilini, Greece; Email: vaf@aegean.gr

[2] Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, 12, Vas. Sofias street, GR 671 00 Xanthi, Greece; Email: ggravvan@ee.duth.gr

[3] Department of Information and Communication Systems Engineering, University of the Aegean, GR 83200 Karlovasi, Samos, Greece; Email: platis@aegean.gr

Abstract: The production of scientific knowledge is basically devised in disciplines. The emergence of Information and Communication Technologies (ICTs) has not yet reduced the dichotomy between hard and soft sciences. Contradicting meanings for the same concept or/and dissimilar terms for identical scientific contents often exist within a discipline and among disciplines. We initiate the discussion about meaningful categorization, search and exploitation of scientific knowledge by introducing the concept of intra-disciplinary semantic technologies. G-work analytical framework is extended to trans-disciplinary scientific research in order to open a new path towards unity of scientific knowledge. In this direction, a structural improvement in Journal of Economic Literature (JEL) classification system is demonstrated. Interconnection to ACM Computing Classification System (1998) introduces trans-disciplinary semantic technologies.

Keywords: intra-disciplinary analysis, trans-disciplinary analysis, ontology, g-work, semantic web, XML, RDF, OWL.

1. INTRODUCTION

In many occasions, scientific knowledge production resembles the story of the Babel's tower. Overlapping terms and concepts among different disciplines accelerates the contestation among hard and soft sciences. Contradicting meanings for the same concept or/and dissimilar terms for identical scientific contents often exist. The monolithic way of scientific knowledge accumulation and exploitation is mainly caused by inborn propensity of scientists to over-specialize and determine the essential quality of their subject of study. The more committed a scientist is to one discipline, the less wishes to work on another discipline; simply because he believes that he will loose his comparative advantage and expertise. Fair enough, but the solution path to unity of scientific knowledge should not be driving to a personal super multi-disciplinary expertise, but to an interoperable and compatible way of communication among sciences and scientists, a common language. The emergence of Semantic Web [2,3,4] and the supporting technologies such as XML [7], Simple

Knowledge Organisation System (SKOS) [16], Ontology Web Language (OWL) [9,12] and Resource Description Framework (RDF) [11] offers a promise to facilitate organization of scientific knowledge.
In the present paper, we introduce intra-disciplinary research as the "identification and systematic representation of interoperable semantic relations for the multiple and contested regimes in a certain discipline" (Figure 1), in order to propose semantic technologies for efficient scientific knowledge re-use and dissemination. In section 2, a short description of related work in non-disciplinary research, – namely multi-disciplinary, inter-disciplinary and trans-disciplinary – is provided. The aspects of our analytical framework are examined in section 3. Apart a general discussion, we develop an application of our methodology – based on OWL/RDF – for the Journal of Economic Literature (JEL) classification hierarchy and ACM Computing Classification System (1998), which are supposed to represent the basic thematic fields in economics and computer science, respectively. Section 4 concludes.

2. RELATED WORK

The production of scientific knowledge is basically devised in disciplines. Notions as development, information, uncertainty, planning, appear in papers of different authors from several scientific fields. From economics to sociology, from epistemology to knowledge management, from informatics to mathematics, from natural scientists to literary intellectuals [18], the debates have been promoted during the last decades; reflect the emergence of new forms of scientific analysis.

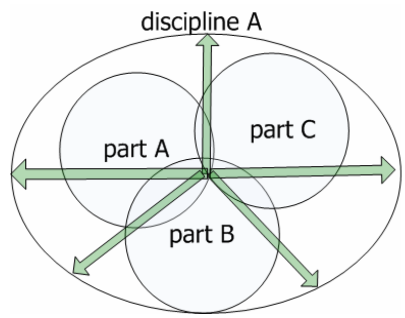

Figure 1: Intra-disciplinary research.

A new form of knowledge production – cited Mode 2 – is identified as emerging alongside the traditional and familiar Mode 1 [6]. This new mode does not only affects what kind of knowledge is produced, but also how and in what context it is produced. The fundamental characteristic is that the new mode operates within a context of application where problems not are set within a disciplinary framework – it is trans-disciplinary rather than mono- or multi-disciplinary. While the traditional, Mode 1 knowledge is basically generated by disciplinary university-based research, characterized by homogeneity and organized in a simple hierarchical way, and tends to preserve its form. Mode 2 knowledge is created in broader, trans-disciplinary social and economic contexts; in non-hierarchical, heterogeneously organized forms, which are essentially flexible and transient. In that direction, non-disciplinary research deals with the ways of combining elements form various disciplines, in order to solve theoretical and practical problems. Main variations of non-disciplinary research are considered to be (a) multi-disciplinarity, (b) inter-disciplinarity, and (c) trans-disciplinarity. Specifically, in multi-disciplinary research (Figure 2), the subject under study is examined by using different disciplinary perspectives. Examples of multidisciplinary subjects are: environment, economy, earth, health, nature, society, and mind.

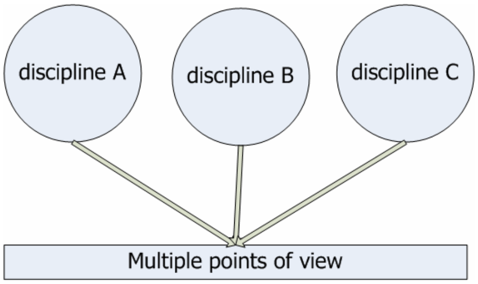

Figure 2: Multi-disciplinary research.

Contrastingly, an inter-disciplinary approach (Figure 3), builds its own theoretical and methodological tools. Examples of inter-disciplinary subjects are: bioethics, econometrics, cultural economics and biochemistry.

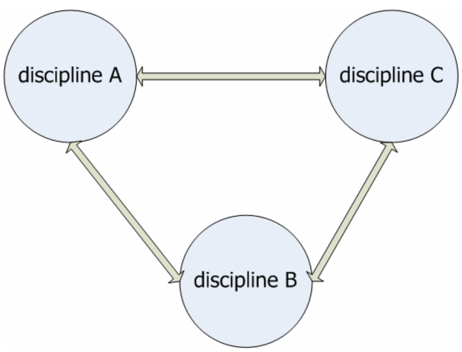

Figure 3: Inter-disciplinary research.

A trans-disciplinary approach (Figure 4) goes one step further, as it is based upon a common analytical framework, and it is followed by an integrated view of different disciplinary epistemologies. In trans-disciplinary research, the point is not just application of given methodologies, but also implication—a result of imagining entirely new possibilities for what disciplines can do. The 'trans' in trans-disciplinarity is about recognizing the holistic approach of this process of investigation which trans-forms mainstream definitions of research. A more general definition of transdisciplinarity was given by Basarab Nicolescu [13]: *Transdisciplinarity is the new "in vivo" knowledge, founded on the following three postulates:*

1. There are, in Nature and in our knowledge of Nature, different levels of Reality and, correspondingly, different levels of perception;

2. The passage from one level of Reality to another in insured by the logic of the included middle;

3. The structure of the totality of levels of Reality and perception is a complex structure: every level is what it is because all the levels exist at the same time.

Trans-disciplinary research is not counteractive, but complementary to multi-disciplinary and inter-disciplinary research.

3. ANALYTICAL FRAMEWORK

3.1 General

A comprehend analytical framework in knowledge formation has been described by Vafopoulos et al [22]: "*A major driving force in knowledge creation is human interaction. For analysis purposes, the stylized facts of knowledge creation evolution have been divided into three eras:*

1. physical era

Knowledge creation is totally based on human interaction by physical means (face-to-face, messages in paper, wood etc).

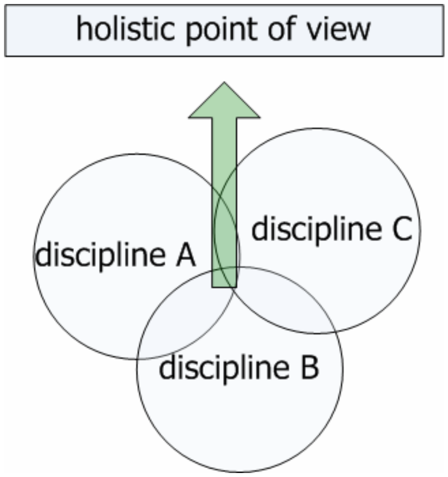

Figure 4: Trans-disciplinary research.

2. computer era

Computer machines introduced implicitly or explicitly in human interaction, information storage and exploitation.

3. ambient era

Completely enveloping technologies emerge in human-machine interaction and knowledge creation.

Today, we face the first phases of the computer era, which is dominated by Internet's penetration in many aspects of social life, accelerating transformation to the ambient era".

In ambient era's realm, the major part of human trans-action will be g-work [19,20,21,22]. G-work was initiated as a personal Grid e-workspace for every citizen [19], and is defined to have four interconnected parts:

- Digital Storage.
- Network Traffic.
- Processing Power.
- One-stop Web Services.

As Vafopoulos [19] argues "*The first three aspects are related to technological infrastructure investments. The fourth aspect, one-stop web services, is the fundamental one for ICT exploitation. Specifically, operates on a semantic web portal basis as the unique electronic gate for a specific geographical region promoting:*

- *Established web services like e-mail, yellow pages, maps, tour guides.*
- *Innovative web services including semantic e-commerce and auctioning services for local goods, human resources, and raw materials based on grid computing technology*.
- *Advantageous mega-marketing features by aggregating marketing expenses under a single umbrella achieving economies of scale.*
- *Personal and entrepreneurial productivity upgrade.*
- *A structured, no disposable, comprehensive and expandable social knowledge base available to all citizens.*
- *E-Inclusion and direct democracy schemes in practice.*
- *An innovative environment where new ideas and individual creation can emerge and diffuse in less cost."*

G-work is a reality for major companies across the globe, through semantic integration technologies like ERP and XML software and hardware. In addition, primitive forms of g-work framework are considered to be present in NASA's Information Power Grid (IPG) [10], in various Grid projects [5,8] and in the Amazon's web service marketplace, announced during the last months. But the stylized fact of g-work, which is considered to be the availability of user-centric interoperable and vertical Grid-web services, is not employed in contemporary scientific research. The proposed holistic g-work model in scientific research (Figure 5) is based on a functional combination of funds, infrastructure, trans-disciplinary technologies and tacit knowledge in problem-solving. Analytically: (a) funds include public and private funding, (b) infrastructure refers to hardware platforms for ubiquitous and grid computing which are combined and cooperate with (c) trans-disciplinary technologies (including intra-disciplinary technologies) stand for semantic integration technologies across all disciplines, and (d) tacit knowledge - the concept comes from philosopher Michael Polanyi [14] - models knowledge that has not yet been codified, but remains embodied in researchers, and includes beliefs and preferences. Tacit knowledge is considered valuable because it provides context for people, places, ideas, and experiences. Effective transfer of tacit knowledge generally requires extensive personal contact and trust. In few words, the g-work problem-solving methodology in scientific research, suggests that in order to solve a specific problem, we first search what has been done until now in all disciplinary fields, then we might operate a software program and after that, search and empirical results are filtered through personal opinions in order to answer the initial problem. Finally, our scientific outcome should be feed-backed to the system in a semantic and re-usable way.

3.2 Technologies

One of the thwarting and practical challenges of contemporary science is the diversity of concepts, terms, notations and identifiers for the same content. Contradicting meanings for the same concept or dissimilar terms for identical scientific contents often exist across the scientific spectrum, causing a communication and dissemination deficit to wider public. Syntactical, structural, and semantic conflict issues are becoming increasingly evident within a certain discipline and among different disciplines. With XML becoming a basic building block for exchanging data, – mainly in information and life sciences – it is obvious that this progress only partially resolves the problem. Additional technologies are needed in order to effectively rationalize processes and information sets between and among disciplines – without requiring point-to-point data and terminology mappings, processes that are both time- and labor-intensive.

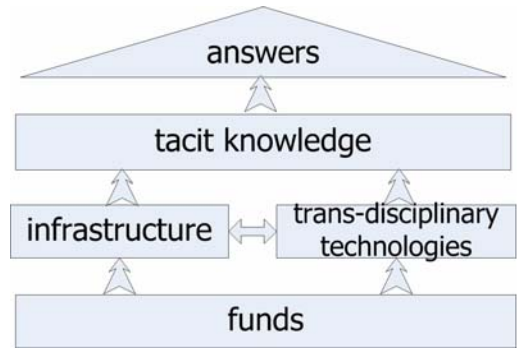

Figure 5: The g-work problem-solving methodology in scientific research.

The basic part of our toolbox is metadata. Metadata or “data about data” are considered to be one of the first forms of supplemental data description. There is an increasing interest among researchers and practitioners to develop semantic metadata models that can efficiently describe large arrays of relationships within a knowledge domain space.

The simplest semantic model is taxonomies [4]. Taxonomies should be considered as a way of categorizing information within a relatively well-defined associative structure. The form of association between two items is inherent in the structure, and in the connections between items. Taxonomical models include connections between terms, but do not define their nature. All the relationships become hierarchical “parent-child” links. Alternatively, this hierarchical structure is called a “tree,” with the root at the top and branching downward. A thesaurus is a higher order form of semantic metadata model than taxonomy, because its associations contain additional inherent meaning. Specifically, a thesaurus is a taxonomical model with some additional semantic relations in the form of a controlled vocabulary. The nodes in a thesaurus are “terms,” meaning they are words or phrases. These terms have “narrower than” or “broader than” relationships to each other. A thesaurus also includes other semantic relationships between terms, such as synonyms. Taxonomies and thesauri are limited in their semantic expressiveness and scope, because they offer a single dimensional axis on which to define relationships. As such, they are typically used to create a classification system, but they fall flat when trying to represent multi-dimensional and/or varied conceptual domains. Concepts are the bearers of meaning, as opposed to the agents of meaning. They are largely abstract, and therefore more complex to model. Concepts and their relationships to other concepts, their properties, attributes, and the rules among them cannot be modeled using taxonomies. Other more sophisticated forms of models, like ontologies, can represent this content. A semantic model in which relationships are explicitly named and distinguished is called “ontology”. Because the relationships are specified, there is no longer a need for a strict structure that encompasses or defines the relationships. The model essentially becomes a network of connections with each connection having an association independent of any other connection. Unlike a taxonomy, which is commonly shown as a “tree,” ontology typically takes the form of a “graph,” i.e., a network with branches across nodes and with some child nodes having links from multiple parents. This connective variability provides tremendous flexibility in dealing with concepts, because many conceptual domains can not be expressed adequately with either a taxonomy or a thesaurus. Just as improvements in languages of model-based programming increased the ability to move from conceptual models to programmatic models without the need for human coding steps, similar advancements have taken place within ontological development. Whereas once ontologies were created primarily for human consumption, – epistemological issues – the development of well-defined protocols for expressing ontologies along with a growing infrastructure that support such models (i.e. semantic grid computing infrastructure), provides increased capabilities for models to deduce the underlying context and draw logical conclusions. Ontologies, in our point of view, should be considered as the connector between epistemology and semantic technologies. The Semantic Web Wedding Cake (or “layer cake”) is also consisting of programming language tools like XML, RDF and OWL. XML is an abbreviation of eXtensible Markup Language, and is a standard way of describing, transporting, and exchanging data that was pioneered by the W3C in the 1990s. XML makes use of metadata by establishing a protocol for using descriptive terms for data fields, and enabling the creation of logical schemas and namespaces around associated data elements. RDF [11] is an abbreviation of Resource Description Framework, and offers ways to make data richer and more flexible, and therefore able to exist in environments outside those explicitly defined by system programmers and data modelers. RDF encodes information in sets of triples, each triple being rather like the subject, verb, and object of an elementary sentence. RDF provides an infrastructure for linking distributed metadata and also serves in conjunction with OWL as a core language for describing and representing ontologies [12]. OWL is an abbreviation of Web Ontology Language. Whereas RDF's fundamental value focuses in enabling association and integration of distributed data, OWL's added value is in facilitating reasoning over distributed data. OWL is a highly expressive modeling language that is compatible with existing data stores and modeling constructs including XML, rational, and object-oriented approaches. OWL also provides loosely-coupled "views" of data which makes federated knowledge bases easy to build and evolve. Significantly, OWL has machine-actionable semantics. Run-time and design-time software tools interact with models, data, metadata, rules, and logic without human assistance or highly specific application code.

3.3 An application

The number of institutions involved in knowledge production has expanded significantly. In the past, universities had an unquestionable oligopoly in this area. Today, however, knowledge is produced by many different kinds of public and private institutions. ICTs are accelerating the dissemination of digital goods [15], and globalize knowledge production process among various actors and across different disciplines. Altogether, scientific journals have failed to take into account these changes that have occurred in the production of scientific knowledge. In our analysis, intra-disciplinary research is a prerequisite for non-disciplinary analytical frameworks, because it offers efficient "connectors" among different disciplines. For instance, the Journal of Economic Literature (JEL) classification hierarchy for economics incorporates in A12 node the "Relation of Economics to Other Disciplines" as the unique connecting point to economics-based inter-disciplinary analysis. Intra-disciplinary semantic technologies – initiated in the present paper – include an ontological conceptualization of economic terminology and theory, which facilitate a meaningful and re-usable knowledge function within economics, and among other scientific fields. Our application is concerning macroeconomic policy formation. For demonstration reasons, it is assumed that the starting point is a scholar who wants to receive economic education. Subsequently, utilizes various resources (namely human, hardware and software) and forms macroeconomic policy proposals (Figure 6).

Actually, macroeconomic policy formation is a more complicated phenomenon, but is beyond present paper's scope to analyze it. Firstly, new intra-disciplinary dynamics are introduced in the classical JEL classification hierarchy by originating causal and semantic relationships among existing JEL categories (Figure 6). Another crucial advantage of this approach (except reusability of abandoned scientific knowledge) is considered to be the maintenance of a successful classification system, which is popular among the majority of economists and compatible to past journals issues. Trans-disciplinary semantics are introduced by analyzing hardware and software resources, according to the ACM Computing Classification System (1998). For instance, category B stands for

HARDWARE (HW_Resources class) and B.1 stands for CONTROL STRUCTURES AND MICRO-PROGRAMMING (Figure 6).
Authors argue that current taxonomical systems in scientific knowledge representation are myopic barriers to scientific knowledge accumulation and exploitation. The proposed paradigm shift in research and science policy facilitates the e-science efforts [8], towards the vision of Ambient Intelligence (AmI) [1] and unity of scientific knowledge.

4. CONCLUSIONS AND FUTURE WORK

In the current paper, we provide practical content to trans-disciplinary research by introducing intra-disciplinary dynamic analysis and integrating to both of them semantic technologies. The major effects of trans-disciplinary semantic technologies are the following: (a) the effects on research practices are concentrated in the production of more comprehend results through heterogeneous networks of scientists and practitioners, (b) the effects on knowledge accumulation, exploitation,

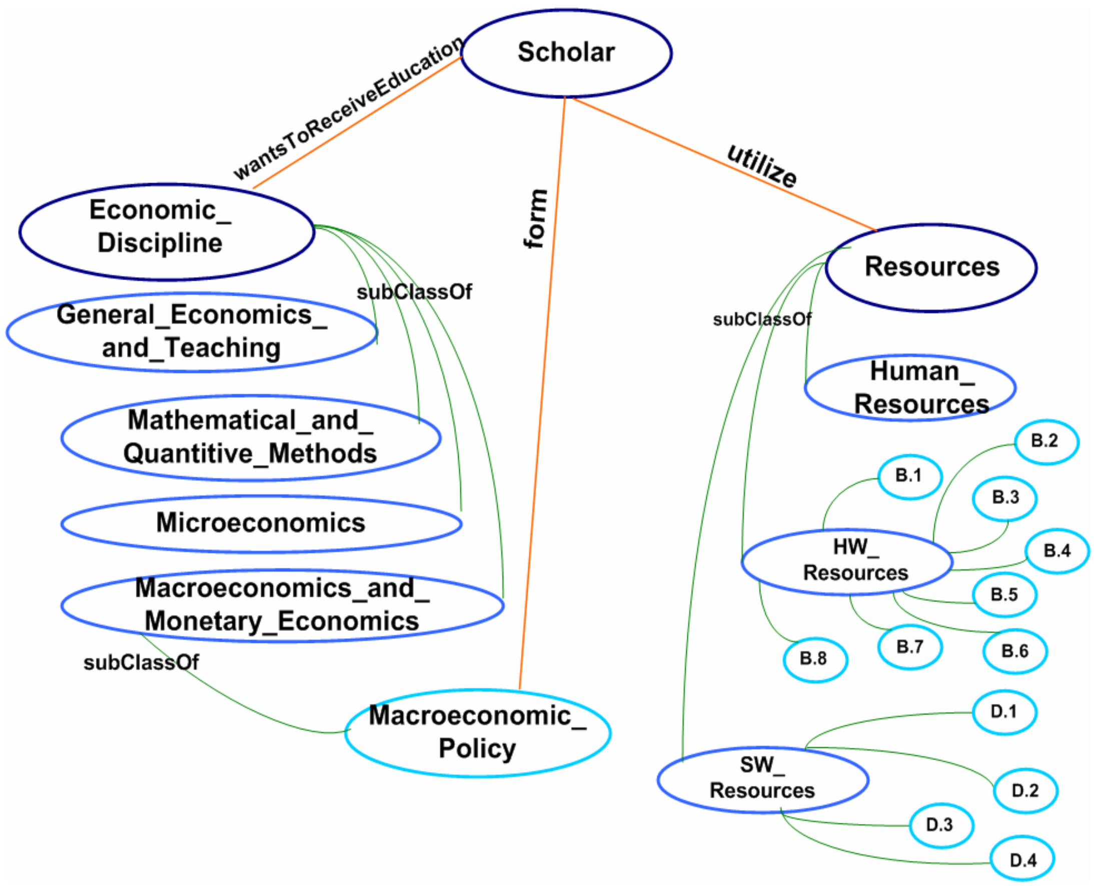

Figure 6: Graph for the macroeconomic policy formation ontology.

diffusion and re-use, occurred from conceptual codification and first-best querying technologies – opposed to second-based google-like capabilities. Some of others positive effects concern the process of diffusion and accumulation and the elaboration of innovative teaching activities and practice transfers through patents. (c) The effects on scientific publishing practices by anticipating an increased inflation of redundant articles and difficulties to have some panoramic point of view in different research fields. Relative projects and institutions until now are stick in a theoretical level (i.e. International Center for Transdisciplinary Research (CIRET) and Coherent Organization of a Navigable Problem-Solution-Learning Space). Working groups in W3C - style should be initiated, in order to address the development of intra – and trans-disciplinarity standards, based on the established Semantic Web's technologies.

ACKNOWLEDGEMENT

We would like to thank Mr. Michalis Klonaras for his valuable assistance in developing the RDF/OWL application.

REFERENCES

[1] **AARTS, E., HARWIG, R., and SCHUURMANS, M. (2001):** Ambient intelligence in The Invisible Future: The Seamless Integration of Technology into Everyday Life, McGraw- Hill Professional.

[2] **BERNERS-LEE, T. (1998):** The Web Model, World Wide Web Consortium (W3C).

[3] **BERNERS-LEE, T., HENDLER, J. and LASSILA. O. (2001):** The semantic web, Scientific American.

[4] **DACONTA, M.C., OBRST, L.J. and SMITH K.T. (2003):** The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management, Indianapolis, Wiley.

[5] **FOSTER, I., KESSELMAN, S. and TUECKE, S. (2001):** The anatomy of the GRID: Enabling scalable virtual organizations, International Journal of Super Computer Applications, 15 (3).

[6] **GIBBONS, M., LIMOGES, C., NOWOTNY, H., SCHWARTZMAN, S., SCOTT, P. and TROW, M. (1994):** The New Production of Knowledge, The dynamics of science and research in contemporary societies, London, Sage Publications.

[7] **GOLDFARB, C.F. and PRESCOD, P. (2003):** XML Handbook, 5th Edition, Prentice Hall.

[8] **HEY, T. and TREFETHEN, A.E. (2002)**: The UK e-Science Core Programme and the Grid, Future Generation Computer Systems, 18, 1017–1031.

[9] **HORROCKS, I., PATEL-SCHNEIDER, P. F. and VAN HARMELEN, F. (2003):** From SHIQ and RDF to OWL: The making of a web ontology language, Journal of Web Semantics, 1(1):7–26.

[10] **JOHNSTON, W.E., GANNON, D. and NITZBERG, B. (1999):** Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid, Proceedings of the Eighth IEEE International Symposium on High Performance

Distributed Computing. Aug. 3-6, 1999, Redondo Beach, California.

[11] **KLYNE, G. and CARROLL, J.J. (eds) (2004):** Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

[12] **MCGUINNESS, D.L. and VAN HARMELEN, F. (eds) (2004):** OWL Web Ontology Language Overview, W3C Recommendation, http://www.w3.org/TR/2004/REC-owl-features-20040210

[13] **NICOLESCU, B. (2002):** Manifesto of Transdisciplinarity, SUNY Press, New York.

[14] **POLANYI, M. (1967):** The tacit dimension, New York, Doubleday.

[15] **QUAH, D. (1997):** Increasingly weightless economies, Bank of England Quarterly Bulletin, 37(1), 49–56.

[16] **SIMPLE KNOWLEDGE ORGANISATION SYSTEM (SKOS) (2005)**: World Wide Web Consortium (W3C) Activity. http://www.w3.org/2004/02/skos/

[17] **SMITH, M.K., WELTY, C. and MCGUINNESS, D.L. (2004):** Web Ontology Language (OWL), W3C Recommendation. http://www.w3.org/TR/owl-guide/

[18] **SNOW C.P (1993):** Two Cultures, Cambridge University Press, (1959).

[19] **VAFOPOULOS, M.N. (2005):** A roadmap to the GRID e-workspace. Proceedings of the Advances in Web Intelligence: Third International Atlantic Web Intelligence Conference, AWIC 2005, Lodz, Poland, June 6-9, Lecture Notes in Computer Science 3528 Springer 2005, ISBN 3-540-26219-9, 504 – 509.

[20] **VAFOPOULOS, M.N. and AGGELIS, V. (2005):** A business model for the Grid e-workspace, Journal of Applied Systems Studies, in press.

[21] **VAFOPOULOS, M.N, AGGELIS, V., and PLATIS A.N. (2005):** HyperClustering: from digital divide to the GRID e-workspace. Proceedings of the VI Data Mining: Data Mining, Text Mining and their Business Applications, WIT Transactions on Information & Communication Technologies, vol. 34, ISSN 1743-3517, 311-321.

[22] **VAFOPOULOS, M.N., GRAVVANIS, G.A. and PLATIS, A.N. (2005):** The personal grid e-workspace, In: “Grid Technologies: Emerging from Distributed Architectures to Virtual Organizations”, M.P. Bekakos, G.A. Gravvanis and H.R. Arabnia, eds., WIT Press, submitted.