

HyperClustering: from digital divide to GRID e-workspace

Michalis Vafopoulos¹, Vasilios Aggelis², Agapios Platis³

¹*Department of Business Administration, University of the Aegean, Greece.*

²*Department of Business Administration, University of the Aegean, Greece.*

³*Department of [Information and Communication Systems Engineering](#), University of the Aegean, Greece.*

Abstract

Digital divide is defined as the inequality in access to information and communication technologies (ICT) between industrialized and developing countries, and between urban and rural populations. In contrast to the fact that technological gap is narrowing in terms of access and training, content providing is still problematic. The so-called information retrieval problem is considered to be a second generation digital divide. Information retrieval problem refers to the general problem of transforming information to usable knowledge. Basic aspects of this fundamental problem is anticipated in W3C research by solving information interoperability and discovery issues, but every local society has to specialize policy interventions in order to stimulate entrepreneurship in a knowledge-based economy. Developing a knowledge-based economy is equivalent to eliminating the digital divide, promoting business environment and offering more opportunities to all citizens. In this context, we propose the HyperClustering framework (HC), a general operational web-based structure for a local economy which semantically analyses, clusters, integrates and boosts personal and social activities. The final stage of HC constitutes the creation of GRID e-workspace for every citizen and company.

Keywords: virtual communities, digital divide, HyperClustering, interoperability of web services, virtual organization, semantic GRID, local development, topic map, knowledge management, ICT clusters.

1 Introduction

Digital divide is composed of three interrelated factors: (a) access, (b) training and (c) content. During the last decade, significant efforts have been made in developed countries to built ICT infrastructure causing a fast decline in access cost. In addition, the academic and professional studies syllabi have been enriched by “new economy” courses. In contrast to the fact that technological gap is narrowing in terms of access and training, content providing is still problematic. The so-called information retrieval problem is considered to be a second generation digital divide. Namely, people having access to ICT (PC, broadband internet, intranet) can not access or operate certain web services. Information retrieval problem refers to the general problem of transforming information to usable knowledge. Basic aspects of this fundamental problem is anticipated in W3C research by solving information interoperability and discovery issues, but every local society has to specialize policy interventions in order to stimulate entrepreneurship in a knowledge-based economy. Developing a knowledge-based economy is equivalent to eliminating the digital divide, promoting business environment and offering more opportunities to all citizens. In this context, we propose the HC framework, a general operational web-based structure for a local economy which semantically analyses, clusters, integrates and boosts personal and social activities. At the first stage, we develop synergies among human activities by mapping implementation paths for the most popular of them. Based on this structured information standard, a web-based Virtual Organization (VO) is constructed which integrates all the major activities of a local economy. The added value of HC focuses on upgrading business environment by creating and organizing workflows between community members and exploiting the network effect ICT offer. In section 2 established theories, technologies, and policies for knowledge-based development are analyzed including economic policy, business, knowledge management and informatics issues. The HC framework is presented in section 3. Methodological principles and technologies are reported in subsections whilst a life-cycle topic mapping model is introduced. The last subsection refers to the Web and GRID services convergence since the final stage of HC constitutes the creation of GRID e-workspace for every citizen and company.

2 Established theories, technologies, and policies for knowledge-based development

2.1 Economic theory and policy

The knowledge-based development (KBD) should prove to be the most important determinant of growth in living standards and new job creation in the next century has a tremendous degree of attention and support from business, government and academics during the last decades. Knowledge as an engine of economic growth introduced by the endogenous growth literature pointed to a

link between knowledge and economic growth through mechanisms such as learning-by-doing (Romer [1]); human capital (Lucas [2]) public R&D and infrastructure (Jones [3]). This basic hypothesis has been tested for a link running from some measure of knowledge to growth in productivity. There are two basic visions of Knowledge-based economy. The “weightless economy” (Quah [4]) and the “information economy”. The first notes that economic value seems to be increasingly concentrated in the fundamental role of digital goods. The second focuses on the important role that Information and Communication Technologies (ICT) have come to play in modern economy. European Union policy in line with the previous analysis defines basic directions for ICT research and application topics: *“With myriads of interconnected devices, we need to explore how to tap into all this computing and networking power in a way that can be adapted to different needs. To build service-oriented infrastructures that autonomously shares and manages multiple resources across customers, business units and applications is a challenge that requires further research in areas like software, Grid and knowledge technologies.”* Technology and economy are changing very fast. The way businesses and individuals operate should change in order to avoid threats and capitalize on opportunities and that appropriate government policies will need to acknowledge those changes. Since knowledge and ICT fits poorly into the economist’s toolkit an interdisciplinary analysis –including Business and Social Studies, Knowledge Management and Informatics – is needed.

2.2 Business and Social Studies

2.2.1 Clusters

Porter (Porter [12]) defines clusters as “geographically proximate groups of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities”. Both practically and conceptually, ICT clusters have become a prerequisite of economic development policy in many parts of the world (e.g. Schmitz [5], Das [6]). It is now a universally accepted fact that successful regional economies are, to varying degrees, specialized. Even the most diversified regions are home to industries that, because of historical accident, targeted recruitment, or geographic peculiarities, are found in higher concentrations than in other places. Clusters produce externalities, lower costs of supplies and components, specialized and customized services, skilled labor, and potential partners, and the soft externalities that produce access to tacit knowledge of technologies, markets, and opportunities to network, and to aggregate interests and needs.

2.2.2 Collaborative Working, Learning Networks and Virtual Communities of Practice

Over the past two decades, key policy institutions, such as the OECD, and many national governments, have come to believe that the global economy is increasingly a knowledge-based one. Since competitive success depends on the ability to produce knowledge and utilize it effectively, there is a pressing need

for firms, communities, regions and nations to invest a greater share of resources in education and training than they have in the past. Recent work indicates that innovation is a *social process* triggered by “users” who engage in a mutually beneficial dialogue and interaction with producers. In this way, users and producers actively *learn* from each other, by “learning-through-interacting” (Lundval [7]). Learning in this sense refers to the building of new competencies and the acquisition of new skills, not just gaining access to information. The easier and cheaper access to information reduces the economic value of more codified forms of knowledge and information. Virtual communities date back before the appearance of the Web. According to Preece, an online community consists of (Preece [8]): (a) *people*: who interact socially as they make an effort to satisfy their own needs or perform special roles, such as leadership or community’s moderation; (b) a *shared purpose*: such as an interest, need, information exchange, or service that provides a reason for the community; (c) *policies*: in the form of tacit assumptions, rituals, protocols, rules and laws that guide people’s interaction; (d) *computer systems*: to support and mediate social interaction and facilitate a sense of togetherness.

2.3 Knowledge Management (KM)

Despite the fact that scientific and technological knowledge is of key importance, knowledge about how to organize and manage everyday economic activities, especially those involving the application of new technological insights, is also a fundamental determinant of economic performance. KM transforms structured and unstructured information, selects and combines the important information for a user in a specific context, so that decisions and the actions of individuals, companies and public institutions are supported. The Semantic Web (SW) (Berners-Lee [9]) can be a very promising platform for developing knowledge management systems. However, the problem is how to represent knowledge in the machine-understandable form, so that relevant knowledge can be found by machine agents. The technological framework for building up knowledge management systems is based on existing SW tools (Ontobroker) [10].

2.4 Informatics

By its very nature, information about on the Web tends to be distributed, heterogeneous, volatile, interrelated in limited and chaotic forms, and focused around topics, persons, projects, and organizations. HTML and the WWW have had enormous impact on the process of distributing human-readable data to even casual computer users. Yet these technologies are actually quite limited in scope: Web data lacks machine-understandable semantics, so it is generally not possible to automatically extract concepts or relationships from this data or to relate items from different sources. Specifically, the world of Web Services (WS) may be characterized as a world of heterogeneous and loosely-coupled distributed systems where adaptivity to ad-hoc changes in the services offered by the components of the systems is considered advantageous. By loosely-coupled, we

mean that the interactions between system components are not rigidly specified at design time, but that system components may opportunistically make use of new services that become available during their lifetime without having been explicitly told of their existence from the outset. In order to be anticipated this inconsistency the following blocks of research and practice have emerged.

2.4.1 Standardization

- The World Wide Web Consortium (W3C) develops interoperable technologies specifications, guidelines, software, and tools to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding.
- WS-I is an open industry organization chartered to promote WS interoperability across platforms, operating systems and programming languages.
- OASIS (Organization for the Advancement of Structured Information Standards) is a not-for-profit, international consortium that drives the development, convergence, and adoption of e-business standards.

2.4.2 Semantic Web Portals

SW portals are defined as follows: (a) It is a web portal. A web portal is a web site that collects information for a group of users that have common interests (Heflin [11]), (b) it is a web portal for a community to share and exchange information and (c) it is a web portal developed based on semantic web technologies.

2.4.3 Software

During the last decade, main IT investment for an SME includes Enterprise Resource Planning (ERP) which is an [information system](#) that integrates all manufacturing and related applications for an entire [enterprise](#). Furthermore, a year before Oracle released a software package called Database10g which includes an enterprise-class, cluster-capable, grid-ready database that supports all new features equally across all platforms.

2.4.4 Semantic GRID (SGRID)

As the SW is to the Web, so is the SGRID to the GRID. By analogy with the SW, the SGRID is an extension of the current Grid in which information and services are given well-defined meaning, better enabling computers and people to work in cooperation. Both have a need for computationally accessible and sharable *metadata* to support automated information discovery, integration and aggregation. Both operate in a global, distributed and changeable environment. The SW base services can be GRID Base Services. The SW fabric is the means by which the GRID could represent metadata: both for GRID *infrastructure*, driving the machinery of the GRID fabric, and its base and high level services, and for GRID *applications*, representing the knowledge and operational know-how of the application domain.

3 HyperClustering (HC): combine them all

The primary challenge confronted by policy-makers of the European Union today is to design policies targeting very diverse technologies, areas, groups and individuals. How could we integrate such different characteristics? What could be the solution to such a complex issue? The reply was already given: building a knowledge-based society, in which the socioeconomic life and policy-making process is supported by facts and structured information. Developing a knowledge-based economy is equivalent to eliminating the digital divide, promoting business environment and offering more opportunities to all citizens. In this context, we propose the HC framework, a general operational web-based structure for a local economy which semantically analyses, clusters, integrates and boosts personal and social activities. At the first stage, we develop synergies among human activities by mapping implementation paths for the most popular of them. Based on this structured information standard, a web-based VO [13] is constructed which integrates all the major activities of a local economy. The added value of HC focuses on upgrading business environment by creating and organizing workflows between community members and exploiting the network externalities and spillovers ICT offer. Our research focuses on less favored regions at E.U., where there is locality awareness and substantial digital divide. We argue that access to structured information and computing power has to be public good in order to boost regional development. The final stage of HC constitutes the creation of GRID (Foster [15]) e-workspace for every citizen and company. HC is defined to be top-level clustering of all fields in business life. Concurrently, operates on a semantic web portal basis as the *unique electronic gate* for a specific geographical region promoting:

- Established WS like e-mail, yellow pages, maps, tour guides.
- Innovative WS including semantic e-commerce and auctioning services for local goods, human resources, raw materials.
- 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.
- eInclusion and direct democracy schemes in practice.
- An innovative environment where new ideas and individual creation can emerge and diffuse in less cost.

In order to achieve an all-encompassing task like HC, beside the fact that social participation and consensus is needed, interdisciplinary analysis is required including Informatics, Knowledge Management and Economics presented in Figure 1.

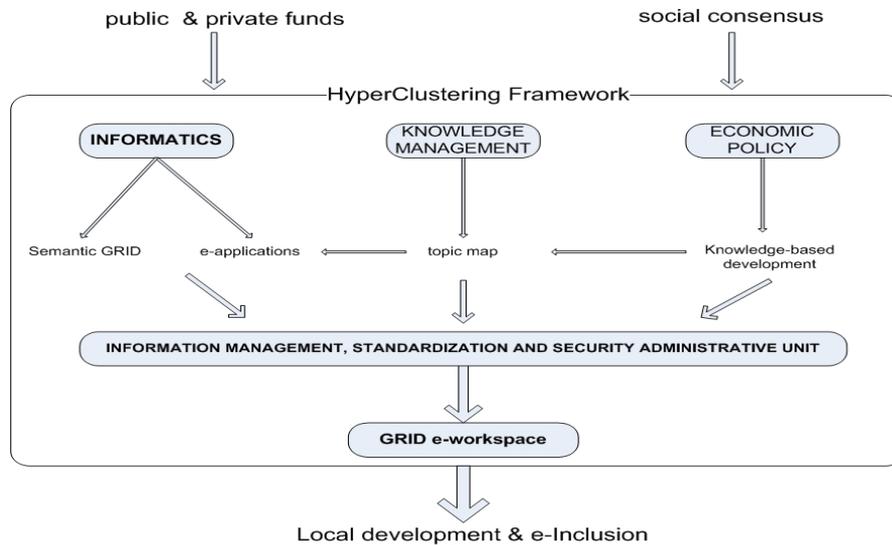


Figure 1: The interdisciplinary nature of HC.

Major role is assigned to a “third intermediary entity” (Gouscos [16]) expanded to the information management, standardization and security unit consisting of three basic components: (1) *Information management, workflow and expert’s opinion* component, including day-by-day management and long run operational framework setup, (2) *Standardization* component, promoting localization of international standards and practices and (3) *Security and personal data* component, responsible for the enforcement of laws and technologies for data security and privacy.

3.1 Basic methodology

HC is introducing an innovative, complete and direct method to exploit ICT for local development. The following challenges could be specified to the architectural domain: (i) diverse groups of players and processes producing informational output, each of these crucial for a particular context of service (users and processes heterogeneity), (ii) growing concerns for privacy and security related to the ownership and management of personal data and (iii) implementation of integrated and personalized services for citizens including distributed processing power and software sharing. HC must be distinguished from e-government since is designed to be a superior level of WS incorporating and interconnecting the building blocks of existing e-applications (e-government, e-education, e-commerce, e-culture). Nowadays, advanced e-government portals include for instance, business startup services. In contrast, HC methodology considers the fact that when we start a business we follow a series of obligatory and optional actions involving government (various applications and approvals), goods (furniture, PC, raw materials, etc.), labor (consultants, accountants,

lawyers, workers and executives) and real estate (rent or buy a building) markets which follow a time path and are interdependent in some ways. In the context of HC time paths and interdependencies are modeled and described in machine-readable language based on Topic Maps and XML schemas.

3.1.1 Topic Maps

In classical software development the first task is to define the scope and the requirements of the system to be developed (Kotonya [17]). In HC framework in order to tackle the complex challenges of building comprehensive user-centric model, a life-event approach is needed. A human life-cycle model is introduced in Figure 2, based on the socioeconomic needs during different phases of a human being's life. For instance, a student (school phase) asks for information related to school studies, entertainment, future studies needed for specific profession, a temporary job and he could face the option of attending university (university phase), becoming a professional (freelancer, employee, employer) or being unemployed. The life cycle model presented here accounts also for movements between occupational statuses (a) without education or training (movement across the vertical axe) or (b) by going back to university in order to change profession (i.e. employee – university – freelancer).

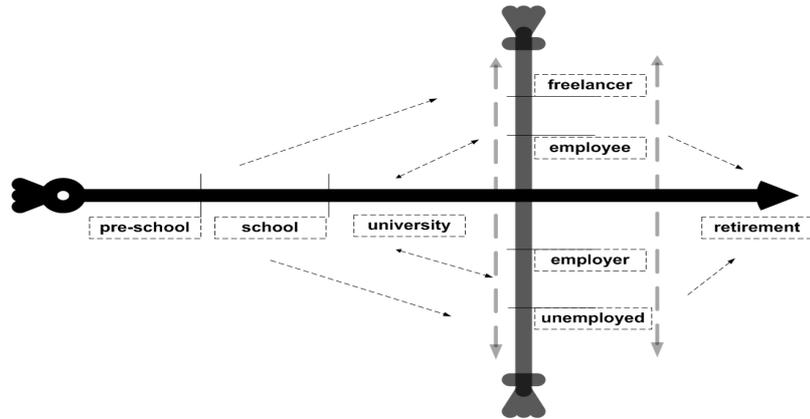


Figure 2: User needs identification is based on a life-cycle model.

3.1.2 “Ontology-driven” information system

The available technological toolkit for working out user needs identified by the life-cycle model is divided in the following groups: (a) *Markup languages*: the most popular markup languages are XML, RDF, DAML+OIL and OWL, (b) *Editorial tools*: these tools are for construction and use of ontologies and (c) *Inference engines*: their purpose is to generate new semantic expressions from

available semantic-based data representations. Ontologies are regarded as a key to solving interoperability problems. The standardization of ontologies used within a network provides a common frame of reference for cross-organizational applications. Service ontologies and service repositories in the HC framework arise from the life-cycle model presented above and are based on the layered architecture for one-stop service provision model (OSP) built by Gouscos et al [16].

3.1.3 Web Services (WS)

Loosely-coupled WS described in section 4 emerged in order to help or/and replace services occur in the traditional physical space. Due to heterogeneity and evolving nature characterizing human needs, combined to scarce resources available, emphasis must be given to KM of WS. Namely, an ex ante topic mapping for WS interactions and interconnections could be crucial in comprehensive one-stop services development. Despite the fact that popular standards for WS interoperability, information retrieval and knowledge discovery technologies analyze methods applied in existing and new datasets, have never being used explicitly and systematically in motivating novel knowledge creation. The HC framework offers a creative and functional environment which encourages, structures and diffuses personal and social knowledge instauration.

3.1.4 GRID e-workspace

The final fundamental stage of HC is designed to be the introduction of the GRID e-workspace. The GRID e-workspace provides a web-based integrated and collaborative hardware and software resources for an individual or an enterprise. Concentrates all services in a single domain for all citizens and companies in a specific geographical region in a collaborative working environment where it is possible to produce, post, search and exchange structured information based on Open Grid Services Infrastructure 1.0 specification [18] which integrates WS-Resource Framework with GRID infrastructure. In addition, GRID technologies provide opportunities to distribute significant amount of processing power to the public through effective distributed computing and shared web based applications.

4 Further research

Architectural centralization is criticized because the center of control becomes a performance bottleneck (Gouscos [16]), but it can be solved by an effective of GRID middleware and infrastructure. Localization of standards published from international organizations are promising since they are providing interoperable end-user services. The HC framework could function under a fully private or public funded business model depending on established social structures of the specific geographical region.

5 References

- [1] Romer, P.M., Endogenous technological change. *Journal of Political Economy*, 98(5, part 2), S71–S102, 1990.
- [2] Lucas, R.E. Jr., On the mechanics of economic development. *Journal of Monetary Economics*, 22(3), 3–42, 1988.
- [3] Jones, C.I., R&D-based models of economic growth. *Journal of Political Economy*, 103, 759–784, 1995.
- [4] Quah, D., Increasingly weightless economies. *Bank of England Quarterly Bulletin*, 37(1), 49–56, 1997.
- [5] Schmitz, H., Collective Efficiency: Growth Path for Small-Scale Industry. *Journal of Development Studies*, 31/4, pp. 529-566, 1995.
- [6] Das, K., Flexibly Together: Surviving and Growing in a Garment Cluster, Ahmedabad, India. *Journal of Entrepreneurship*, vol. 2, 1996.
- [7] Lundvall, B.-Å, (ed). National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. London: Pinter Publishers, 1992.
- [8] Preece, J., *Online Communities: Designing Usability, Supporting Sociability*. New York: John Wiley Sons, 2000.
- [9] Berners-Lee, T., Hendler, J., & Lassila, O., The Semantic Web, *Scientific American*, 2001.
- [10] On-To-Knowledge: Semantic Web Enabled Knowledge Management. <http://homepage.uibk.ac.at/~c703205/download/otk.ieee-article.pdf>
- [11] Heflin, J., Web Ontology Language (OWL) Use Cases and Requirements. W3C Working Draft, 2003.
- [12] Porter, M., *Clusters and Competition*, in: *On Competition*, Harvard Business School Press, Cambridge, 1998.
- [13] The VOSTER IST project, 2001. www.voster.org.
- [14] Shapiro, C. & Varian, H., *Information Rules: A Strategic Guide to the Network Economy*, Boston, Mass. Harvard Business School Press, 1998.
- [15] Foster, I., Kesselman, S. & Tuecke, S., The anatomy of the GRID: Enabling scalable virtual organizations. *International Journal of Super Computer Applications*, 15 (3), 2001.
- [16] Gouscos, D., Laskaridis, G., Lioulias, D., Mentzas, M. & Georgiadis, P., An approach to Offering One-Stop e-Government Services: Available Technologies and Architectural Issues. *Proc. of the EGOV 2002 Conf*, eds. R.Traunmuller & K. Lenk, Springer-Verlag, Berlin, Heidelberg, pp. 264 – 271, 2002.
- [17] Kotonya, G. & Sommerville, I., *Requirements Engineering: Processes and Techniques*, Chichester, Wiley, 1998.
- [18] From Open Grid Services Infrastructure to WSResource Framework: Refactoring & Evolution; GLOBUS. http://www.globus.org/wsrfspecs/ogsi_to_wsrf_1.0.pdf