# Keynote

# The Knowledge Grid and Its Methodology [1]

Hai Zhuge
*China Knowledge Grid Research Group, Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, 100080, Beijing, China*
*E-mail: zhuge@ict.ac.cn*

## Abstract

*In the human, society, interconnection environment and systems methodology perspectives, this paper answers the following questions: What are the Knowledge Grid and its distinguished features? What are its methodology and major research issues? These answers are important to the development of this promising area.*

## 1. Introduction

Knowledge, a product of society in nature, evolves and endures throughout the life of a culture rather than that of an individual [5]. Modern communication facilities like the Internet and various wireless networks provide with unprecedented social and technical conditions for worldwide knowledge sharing. Using and improving social networks is an important way to enrich knowledge and make efficient knowledge sharing.

*The Knowledge Grid is a sustainable human-machine interconnection environment that enables people or agents to effectively generate, capture, publish, share, manage and promote knowledge, to process any type of resource through machines, and to transform resources from one form to another. It provides appropriate on-demand services to support innovation, teamwork, simulation, problem solving, and decision making by using sharable knowledge. It incorporates epistemology and ontology to reflect human cognition, exploits social, biological, ecological and economic principles, and adopts the techniques for the future interconnection environment* [11].

In human civilization, centralized control and self-organization are two social organization models. The two models coexist and sometime conflict with each other under culture. The Grid and Peer-to-Peer (P2P) are two computing models using control and self-organization respectively [2, 7]. Inspiring and sharing of human knowledge involves in both the tightly coupled coordination process and the self-organization process.

The evolving Grid, World Wide Web, P2P and social networking technologies help realize the ideal of the Knowledge Grid. However, the implementation of the Knowledge Grid requires new organization and operation models of resources and their environments.

Knowledge, society and systems are the functions of time. Harmonious evolution of this human-machine environment can ensure the sustainable development of the Knowledge Grid [10, 12].

As shown in Figure 1, the technologies of the Internet, Web, Semantic Web [4], Grid [7], P2P [2, 13], Semantic Grid [15], and various advanced networks [3] can be used to construct the underlying infrastructure of the Knowledge Grid environment. That is, *the Grid is not the unique underlying infrastructure of the Knowledge Grid*.

Different from previous centralized or decentralized databases and knowledge bases, human activities and relevant social networks are the important parts of the Knowledge Grid. In addition to previous methodologies such as software engineering and knowledge engineering, we need a new methodology — the Knowledge Grid methodology.

*The Knowledge Grid Methodology is a multi-disciplinary systems methodology for establishing and maintaining a knowledge world that obeys the principles and laws of economics, nature, society, culture, psychology, and information technology. It is based on the open interconnection semantics,*

[1] Keynote speech at 1st International Conference on Semantics, Knowledge and Grid, Beijing, China, Nov.27-29, 2005.
This work is supported by the National Basic Research Program of China (Project No.2003CB317000) and the National Science Foundation of China (Grant No. 70271007).

*epistemology, ontology, systems methodology, and knowledge management.*

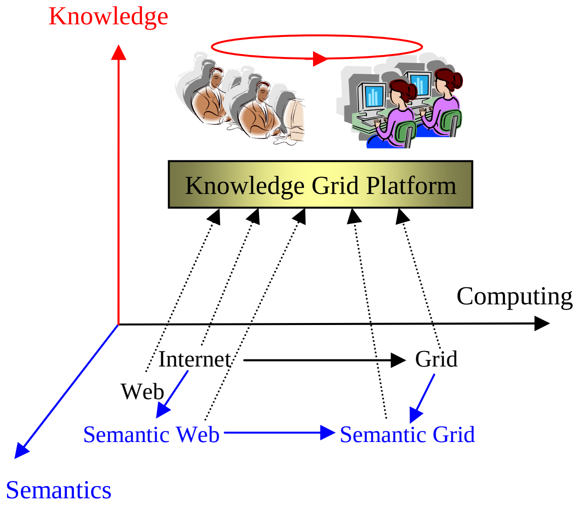


Figure 1. The Knowledge Grid and relevant areas.

## 2. Features of the Knowledge Grid

The Knowledge Grid has the following features:

**(1)** ***Virtual Cyberspace.*** The Knowledge Grid consists of requirements, roles, resources and regulations. With machine-understandable semantics, a resource can actively and dynamically gather relevant resources and fuse them to provide appropriate on-demand services for applications by understanding requirements and functions and relating them to each other. A resource can intelligently cooperate with others to accomplish complex tasks and to solve problems by participating various flow cycles through virtual roles.

**(2)** ***Social Grid.*** People live and work in a *social grid* obeying social and economical rules and laws. The Knowledge Grid is a *virtual social grid*, where people enjoy and provide services through versatile flow cycles like control flows, material flows, energy flows, information flows and knowledge flows. People can communicate and gain knowledge from each other through mutually understandable semantics. Any artificial interconnection environment can be effective only when it works harmoniously with its social grids. The appropriate semantic representation supports mutual understanding between the social grid and the artificial interconnection environment. In the future, different artificial interconnection environments will co-exist and compete with each other for survival, rights and reputation, and will harmoniously evolve with the social grid [1, 10].

**(3)** ***Economic and Adaptive System.*** The Knowledge Grid's three major roles (*producers*, *consumers* and *market mechanism*) adapt the behavior of different participants. Adopting economical and ecological principles to balance the interests of knowledge producers and knowledge consumers can adapt to appropriate evolution and expansion of resources. Avoiding complex computation, the market mechanism automatically and reasonably adjusts the decisions and behaviors of market participants. This ecological system balances species by various flows, which in turn influence the social system. Different species evolve and interact with each other to push the evolution of the whole system.

**(4)** ***Semantic Networking***. Current Web search engines are based on rough semantics (keywords). A huge gap exists between the rough and fine semantics that involves in complex cognitive and psychological processes. Semantics has been studied in different areas such as natural language processing, programming languages and the Semantic Web. The Knowledge Grid needs an open semantic system that establishes the understanding between machines as well as between machine and human. Figure 2 suggests an infrastructure of the interconnection semantics. Resources are identified in the *name space*, and then get meaning in *ontologies*. Various *basic data structures* reflecting natural and social phenomenon like tree and queue form a *computable semantic overlay* on the name space. The *primitive semantic relations* are commonsense relations, like "cause-effect" and "is-part-of", which can derive new semantics. *Semantic patterns* are large granularity semantic units that participate semantic computing. Patterns and semantic relations can be clustered and used for reasoning and explanation. Various *markup languages* such as RDF and XML can be the media of semantic interaction. The *semantic computing model* operates the semantic patterns, primitive relations, ontologies and data structures according to rules. The semantic computing model can infer new semantics according to predefined semantics. Semantic issues such as automatic acquisition, expression, normalization, processing and maintenance are discussed in [11].

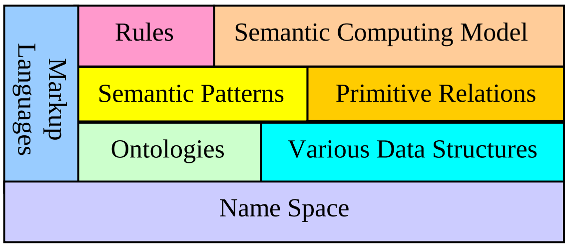


Figure 2. Architecture of the interconnection semantics.

## 3. Synergy Ontology and Epistemology

Ontology reflects people's consensus on semantics in name spaces or symbolic spaces. But it is limited in ability to cope with human cognitive processes.

Epistemology concerns the nature, scope and source of knowledge. People could improve their understanding by studying the development of epistemology.

Knowledge in nature is active, can be transmitted from one individual to another, and thereby belongs to a community or the society. So communication and social processes play an important role in the development of knowledge. Knowledge and society are inseparable. Knowledge is enriched by social activities, and social systems are constructed and developed with knowledge processes.

*To keep evolution sustainable, knowledge should have the characteristics of diversity.*

## 4. Synergy Normalization and Autonomy

Accuracy and correctness require normalization. An example of normalization is the relational data model, which uses a set of normal forms to specify and manage data to ensure the accuracy and correctness of operations on data.

Autonomy requires equality and scalability. P2P computing model has the features of equality and scalability in large-scale networking environment [2]. Synergy normalization and autonomy can have the advantages of both. Figure 3 is a scenario of such synergy. The RSM (Resource Space Model [11]) is a solution to normalize the semantic space, and the P2P semantic link network is a solution to establish the semantic overlay over a P2P network [11, 13].

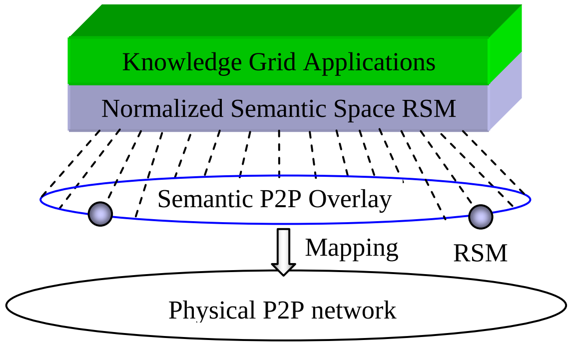


Figure 3. A solution to synergy normalization and autonomy.

## 5. Synergy Domination and Autonomy

Domination and democracy organize human society in different facets and carry different social functions. In current effort to large-scale networking, some technologies like the Grid represent domination, optimization and explicit order, which potentially lead to efficiency. Other technologies like P2P represent democracy, autonomy and implicit order, which meet human aspiration of freedom. The synergy of the technologies of domination and autonomy organizes the interconnection environment and balances the interests of individuals and communities under different culture. The diversity of culture provides more possibilities for such synergy.

## 6. Synergy Trust Space, Strategy Space, Reputation Space and Semantic Space

Knowledge in Knowledge Grid is not statically stored like that in traditional knowledge bases. It flows through a social network of people/agents to self-organize temporal virtual teams. It inspires new knowledge during flowing. Knowledge has higher probability to flow through semantic links such as "co-author", "cite" and "is-supervisor-of". Trust between people/agents influences the effectiveness of knowledge flow.

Selfish and unselfish strategies can be adopted to operate a knowledge flow. Adopting selfish strategy, a node in the network tends to input knowledge from "rich" knowledge nodes (high-energy nodes) by establishing some semantic relationships. This strategy leads to a quick rise of its knowledge energy.

Since unselfish help could lead to indirect benefit [9], a node could adopt the strategy that outputs knowledge to many other nodes to help them become rich. This can be explained in science: scientist *A* published paper *P* that helps a group of scientists *S*, who in-turn can publish papers *PS* that can help a group of scientists including *A*. Knowledge flows through semantic space and trust space with strategies as shown in Figure 4.

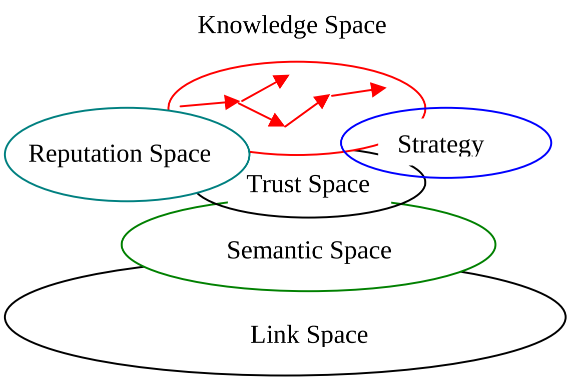


Figure 4. Knowledge flows through trust space, strategy space and semantic space. Strategies can be selfish and unselfish.

## 7. Synergy Social Selection and Social Regulation

An individual has natural attribute, social attribute and economical attribute. Each attribute needs multiple criteria to evaluate. A selfish strategy is successful only when it can balance all these attributes.

In the Knowledge Grid environment, the following social rules interact with each other to adapt individual attributes:

(1) *Individuals always want more rather than less* [14].
(2) *Interpersonal comparisons are meaningful*.
(3) *The rich gets richer in a self*-organization and selfish world.
(4) *No attribute can unlimitedly grow* [12]. For example, individual property will be limited by tax policy.

The failure of social regulation would lead to the emergence of new system and all individuals' social properties will be reassigned under new social regulations. In recent years, P2P users grow rapidly, relevant information flow accounts for 40% of the whole information flow on the Internet. Currently, the P2P model is challenging the central control model.

## 8. Methods and Principles

The dissipative structure, synergetics and hypercycle theory can help us explore the intrinsic self-organization principle of the future interconnection environment and its resources [6]. We can imagine the future interconnection environment as a live system or environment, which consists of species in the form of live resources and versatile flow cycles. Resources could be dynamically organized into diverse flows such as knowledge flows, information flows, and service flows to provide users or applications with on-demand services. Once a requirement is confirmed, all relevant flows could be formed automatically [11, 12].

As a complex system, the Knowledge Grid obeys the principles of systems methodology:

(1) *Integrity and uniformity principles* — require the correctness and simplicity.
(2) *The hierarchical principle* — constructs a Knowledge Grid as a hierarchical system.
(3) *The open principle* — keeps the Knowledge Grid away from the equilibrium state.
(4) *The self-organization principle* — realizes autonomous cooperation under social regulations.
(5) *The principle of competition and cooperation* — enables systems including resources to evolve through competition and cooperation so that competitive resources or systems could play a more important role [1].
(6) *The optimization principle* — makes a system more effective. Information flow, knowledge flow and service flow can be optimized to achieve efficiency in logistic processes.
(7) *The principle of sustainable development* — enables individuals and communities, the inter-connection environment and its human-machine interfaces, the human-machine society, and the natural environment to harmoniously co-evolve.

## 9. Major Research Issues

(1) *Knowledge capture and representation*. Capturing knowledge herein means two folds: (a) people learn from each other directly, or from the resources published by others, and then publish new knowledge; and, (b) gets knowledge from resources by mining, induction, analogy, deduction, and synthesizing. An open set of semantic primitives is needed to represent multi-granular knowledge.
(2) *Knowledge visualization and creation*. How to enable people to share knowledge in a visual way. The semantic link network and the cognitive map

are two ways to display knowledge. A local semantic link network dynamically changing with user focus can better express and focus knowledge. The ideal interface should implement the distinctive characteristics of the Knowledge Grid and be able to inspire people's discovery of knowledge through analogy and induction.

(3) *Propagation and management*. This could eliminate redundant communication between team members to achieve effective knowledge management in a cooperative virtual team. Knowledge flow management is a way to achieve knowledge sharing in a virtual team.

(4) *Organization, evaluation, refinement and derivation*. Knowledge should be organized normally to obtain high retrieval efficiency and ensure the correctness of operations. The Knowledge Grid should be able to eliminate redundant knowledge and refine knowledge so that useful knowledge can be increased. It can also derive new knowledge from existing well-represented knowledge, from case histories, and from raw knowledge material like text.

(5) *Knowledge integration*. Integrating knowledge resources at different levels and in different domains could support cross-domain analogies, problem solving, and scientific discovery.

(6) *Abstraction*. It is a challenge to automatically capture semantics from a variety of resources, to make abstractions, and to reason and explain in a uniform semantic space. The semantic constraints and rules of abstraction ensure the validity of resource usage at the semantic level.

(7) *Scalable and capability-aware network platform*. The Knowledge Grid should enable a user, a machine or a local network to freely join in and leave without affecting its performance and services. At the same time, it should know the changing capacity of the underlying network. It is a challenging task to organize and integrate knowledge within a dynamic network platform [13].

(8) *Knowledge flow dynamics*. A knowledge flow network is a kind of organizational knowledge. It more concerns the content of knowledge and the effectiveness of sharing in distributed cooperative teams. Epistemology plays the key role in the process of generating knowledge. The Knowledge Grid needs a kind of semantic description and generation mechanism that reflects human cognition. Different people may have a different epistemology in the same resource or event. Epistemological mechanisms help humans and agents understand, generate and describe new knowledge when they share resources. An easy way to implement an epistemological mechanism is to develop an epistemological appendage, generated and used in conjunction with the original resources.

(9) *New software Methodology*. Previous software methodology focuses on software process, domain business process and the modeling of the real world. The Knowledge Grid methodology not only concerns the software and real world, but also human behavior and cognition as well as the effective sharing method and the corresponding human-machine process.

(10) *Performance evaluation.* The Knowledge Grid needs objective criteria for evaluating the performance of a Knowledge Grid system.

(11) *Interconnection culture.* It evolves, fuses, and influences knowledge sharing and the whole Knowledge Grid environment.

(12) *Cyberspace philosophy.* It studies the ultimate reality, causes and principles underlying being and thinking within the cyberspace of the Knowledge Grid environment.

## 10. Strategy

A worldwide Knowledge Grid is a long-term target. A preliminary stage developing a medium-sized Knowledge Grid based on a team's intranet would be an appropriate step in the long march towards the long-term target. *Team Knowledge Grids* could become components of the worldwide Knowledge Grid.

A *micro Knowledge Grid* could be the basic component of a medium-sized Knowledge Grid and thus the basic component of the worldwide Knowledge Grid. It would be useful in helping individual knowledge management — managing raw knowledge (in some forms of natural language) and codified knowledge, and transforming tacit knowledge into explicit knowledge. *However, a worldwide Knowledge Grid is more powerful than the sum of its components*.

A Knowledge Grid should support more semantic spaces than just one text space. Knowledge sharing in a Knowledge Grid depends on a correct understanding of its resources. But these semantics are not the same as traditional formal semantics. These should be a kind of informal computable semantics that supports computing, reasoning, abstraction, integration and transformation between semantic spaces. The semantics of the Knowledge Grid should be easily understood by humans and processed by machines.

## 11. Summary

The ideal of the Knowledge Grid is challenging and feasible. Gray proposed a dozen IT research goals in [8]. The development of the Knowledge Grid area will benefit from the progress not only in IT but also in other disciplines.

Progresses in physics, biology, communication, material science and life science will greatly push the development of computer and network. So research in the Knowledge Grid area will cross the IT and other disciplines.

The Knowledge Grid methodology is the study of the fundamental principles, strategies and methods for the development and maintenance of the Knowledge Grid as a human-machine environment. The methodology was proposed for the first time in [11]. This paper enriches the content and focuses on the semantics, society and dynamics, as semantics are the understanding basis of knowledge, social principles play an important role in the Knowledge Grid, and dynamicity is the nature of knowledge sharing.

Relevant research progress and practice are available at Knowledge Grid Research Center (www.knowledgegrid.net), China National Semantic Grid (www.semgrid.net), China Knowledge Grid Research (kg.ict.ac.cn), and China Culture Grid (www.culturegrid.net).

## References


[1] R.M. Axelrod and D. Dion, "The Further Evolution of Cooperation", *Science*, vol.242, no.4884, 1988, pp.1385-1390.

[2] H. Balakrishnan, et al. "Looking up Data in P2P Systems", *Communications of the ACM*, vol.46, no.2, 2003, pp.43-48.

[3] A.L. Barabási and R. Albert, "Emergence of Scaling in Random Networks", *Science*, vol. 286, 1999, pp. 509-512.

[4] T. Berners-Lee, J. Hendler and O. Lassila, "Semantic Web", *Scientific American*, vol.284, no.5, 2001, pp.34-43.

[5] V. Bush, "As We May Think", *The Atlantic Monthly*, vol.176, no.1, 1945, pp.101-108.

[6] Eigen, M. and P. Schuster, The Hypercycle: A Principle of Natural Self-Organization, Springer, Berlin, 1979.

[7] I. Foster, "Service-Oriented Science", *Science*, vol.308, no.5723, 2005, pp.814-817.

[8] J. Gray, "What Next? A Dozen Information-Technology Research Goals", *Journal of the ACM*, vol.50, no.1, 2003, pp.41-57.

[9] M.A. Nowak and K. Sigmund, "Evolution of Indirect Reciprocity", *Nature*, vol.437, no.27, 2005, pp.1291-1298.

[10] H. Zhuge and X. Shi, "Toward the Eco-Grid: A Harmoniously Evolved Interconnection Environment", *Communications of the ACM*, vol.47, no. 9, 2004, pp.79-83.

[11] Zhuge, H., *The Knowledge Grid*, World Scientific Publishing Co., Singapore, 2004. www.worldscibooks.com/compsci/5738.html.

[12] H. Zhuge, "The Future Interconnection Environment", *IEEE Computer*, vol.38, no.4, 2005, pp. 27-33.

[13] H. Zhuge, X. Sun, J. Liu, E. Yao and X. Chen, "A Scalable P2P Platform for the Knowledge Grid", *IEEE Transactions on Knowledge and Data Engineering*, vol.17, no.12, 2005, pp.1721-1736.

[14] Evolutionary Game Theory, plato.stanford.edu/entries/game-evolutionary.

[15] Semantic Grid Community Portal, www.semanticgrid.org.