From Body of Knowledge to Base-Map: Managing Domain Knowledge through Collaboration and Computation

Sean C. Ahearn [1] and André Skupin [2]

1 Center for Advanced Research of Spatial Information, Department of Geography, Hunter College, New York, NY 10065 USA
2 Center for Information Convergence and Strategy, Department of Geography, San Diego State University, San Diego, CA 92182, USA

Abstract: This paper describes a technological implementation of a body of knowledge for domain knowledge management. The system enables collaboration, exploration, and exploitation of domain knowledge through a front-end visual wiki, an ontology-driven knowledge base and Web services. The service-based architecture enables applications that allow knowledge artifacts to be related to domain concepts through inference, and such relationships to be visualized. Essential to the design and operation of the system are such notions as reference system and base map borrowed from traditional mapping. This includes the idea of a domain base map onto which any domain artifact can be projected. We hypothesize that this type of system may break down traditional boundaries, such as between educators, students, researchers, and professionals.

Keywords: Ontology, knowledge domains, reference systems, base maps, visualization, wiki, collaboration, body of knowledge.

1 Introduction

There is a critical need to improve the construction and dissemination of knowledge within and across the many communities that constitute a knowledge domain. Scientific research, academic education, technical training, and professional practice are driven by seemingly disparate concerns, and there tends to be little interaction between these sub-communities. As a result, cutting-edge research is often slow to transition into professional practice, while educational activities and materials often do not represent the current state of research and practice. Similar inefficiencies characterize the matching of software tools to related research insights, students' aspirations to curricular options, and educational programs to workplace needs. Alternatively, what if undergraduate curricula and textbooks co-existed in the same knowledge ecosystem with research publications, software documentation, job advertisements, and grant proposals? What if a student learning an analytical software tool would have ready access to a set of research studies in which similar tools were recently used? What if she would also be shown a list of current job openings requiring mastery of associated skills, reinforcing the real-world relevance of curricular content? What if she could then compare her skill levels in different areas of the knowledge domain with those required by the jobs of interest to her to see how well she matched the various positions available? And, finally, what if that knowledge ecosystem was organically changing as the field evolved? In the discussion below we will demonstrate how the *BigKnowledge™ Body of Knowledge* system (BK-BoK) affords the vision proposed by this set of questions and in essence provides the framework for creation of a knowledge ecosystem to do so.

The metaphor of an ecosystem is gaining acceptance as a way of modeling how a community uses knowledge [6, 8]. An ecosystem consists of a multitude of participants that interact synergistically (through the creation, movement, and consumption of resources) for the success of the system and themselves. In the knowledge ecosystem, the participants include *actors* (people and institutions), *activities* (the tasks that actors perform, such as employment, education, and research), and *artifacts* (the products of actors' activities, especially published and unpublished documents; see Figure 1).

These *domain elements* interact in the ecosystem by producing, sharing, and using knowledge about the domain, including theoretical concepts and practical *skills*. A form of knowledge ecosystem already exists for any given discipline, consisting of the hundreds of thousands of relevant people, institutions, products, and publications. The problem is that the inefficient state of current knowledge ecosystems, as we suggest above, impedes the diffusion of knowledge: new research ideas, algorithms, and best practices; and the integration of disparate knowledge artifacts into more effective wholes.

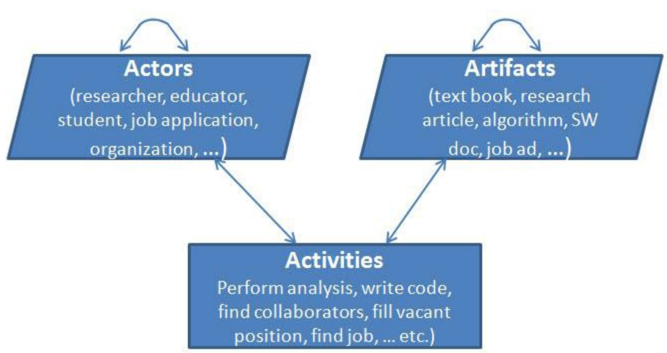

Figure 1. Knowledge Ecosystem

A number of initiatives have addressed aspects of this problem with the creation of formalized "bodies of knowledge." For example, the Geographic Information Science and Technology Body of Knowledge (GIS&T BoK) [7] was developed to provide: (i) a resource for course and curriculum planning; (ii) a basis for comparison of education programs; (iii) a foundation for professional certification, program accreditation, and articulation agreements; and (iv) a resource for HR professionals. Published in 2006, it was a landmark accomplishment in the field [12], but as a framework for actual implementation it has been somewhat limited in use (Foote et al. 2012, p8) because of a lack of formalization.

Ontologies provide a computational approach to formally capturing concepts and their relationships and enable quantitative analysis of a *knowledge ontology* (Protégé, 2013). Some recent ontologies include the Cognitive Atlas (http://www.cognitiveatlas.org) for cognitive psychology, the Indiana Philosophy Ontology project (http://inpho.cogs.indiana.edu), the Computing Ontology from the NSF-funded CPATH project (http://www.distributedexpertise.org/), the Software Engineering Body of Knowledge (SWEBOK, http://swebok.org), the 2012 ACM Computing Classification System (CCS, http://www.acm.org/about/class/2012) and the Semantic WebApplication in Neuromedicine Ontology (SWAN) [6]. Although each of these initiatives is ambitious, they tend to have a fairly narrow scope of application. For example, there seems to be very little interaction between the ACM Computing Curricula bodies of knowledge (designed for education), SWEBOK (designed for professional practice), and the ACM CCS (designed for research publications), even though they have a great deal in common, including their sponsoring organizations.

2 Relationships and Knowledge Pathways

In [1] a visual wiki was implemented that enabled the capture of concepts and their relationships as well as related knowledge artifacts. Since then work on the ontology has been expanded to include a richer set of relationships among *concepts*. Those relationships include: broader, narrower, pre-requisite, post-requisite, and similar (Figure 2). Note that pre-requisite, post-requisite and similar can all be weighted as to the strength of their relationship. Expansion of the types of relationship affords different "views" or "pathways" through the knowledge base. Those pathways are navigated through related pairs of concepts. A *prerequisite pair* is a relationship that reflects the fact that one concept depends on another; that is, one must understand and be competent in Concept A (to some degree) before one can understand and be competent in Concept B. A *taxonomic pair* represents a broader-narrower relationship and reflects the fact that one concept is a superset of the other concept. A third pair-wise relationship is *similar pair*, in which the concepts share a similar meaning.. These pair-wise relationships are a feature of the knowledge ontology and once discovered, should be encoded in the domain knowledge base. A geographic analogy would be network topology, which represents how individual elements of a network (say, road segments) connect to each other. A *knowledge pathway* is an ordered sequence of any number of concepts. For example, a course may cover Concept X, then Concept A, then Concept D, then concept F. This is a feature of the domain ontology, as each domain element (such as a book, a course, or a student) may find its own unique pathway through the Body of Knowledge. Knowledge pathways are shaped by the network of prerequisites, in the same way that a route between two cities depends on network topology. In this “learning pathway” we might follow the sequence of pre and post requisite concepts through the knowledge space to understand the best sequence for understanding a certain concept or becoming competent in a skill. We are in essence creating a "back path" through the knowledge space that would be required for understanding the concept in question.

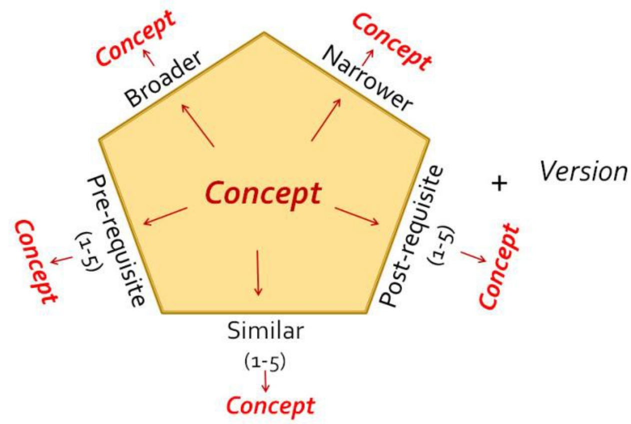

Figure 2. Concept Relationships

3 Collaborative Input and Bottom-up Design

Capturing knowledge is a central challenge for any system that aspires to develop and maintain a formalized Body of Knowledge. The knowledge captured is intended to reflect the collective knowledge of the community. The community therefore needs to be able to participate in improving and updating the standard. This is in direct contrast to the “top-down” approach of committees of experts that have created the standard bodies of knowledge in most domains (e.g., [7]). Since a "top-down" approach cannot always accurately reflect the collective knowledge of the entire community, a more inclusive approach is to use a wiki service that relies on crowd sourcing for the revision of the domain ontology. This allows anyone in the community to propose changes, debate those proposals, and collaboratively come to consensus. An implementation of this type of approach was discussed in [1] with additional editorial capabilities added since. The new system involves four different types of users: viewer, contributor, editor and super-user. Viewers see only approved concepts in the authoritative version of the Body of Knowledge. Contributors make proposals for new concepts, relationships and knowledge artifacts or changes to the existing concepts (e.g. its definition, its existence, etc.). Editors are assigned to a part of the Body of Knowledge relevant to their expertise and it is their job to determine which additions/changes that were made by the contributors should be added to the knowledge base. Determination of acceptance is made based on the contributions related to a given concept by the various "contributors" that have weighed in and requires a consensus of the editors responsible for and have approved access to, that part of the BoK. Super-users manage accounts (Figure 3).

The visual Wiki enables the user to have two views of the data: a *synoptic view* in which the entire knowledge space is visible and a *focal view* that shows a particular concept in its relational context (Figure 4). Using either view, changes can be made to the system by contributors and editors can approve or reject proposed changes.

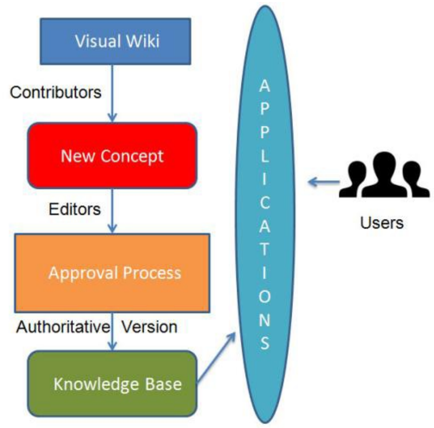

Figure 3. Approval Process for Changes to the BoK

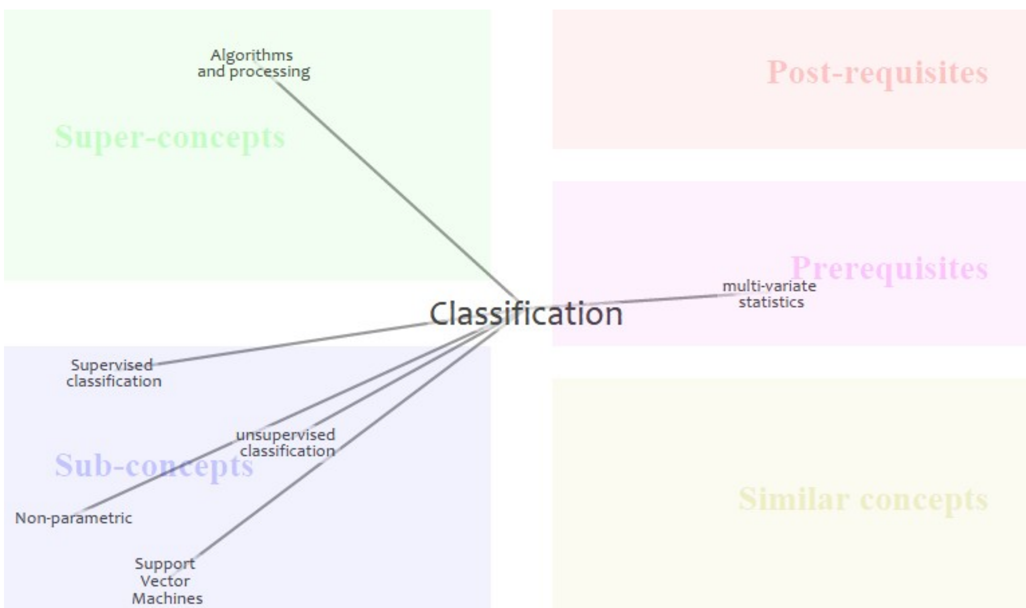

Figure 4. Focal View in Visual Wiki

4 System Architecture

The ***BK-BoK*** system is composed of these main components: a visual Wiki (***BoKWiki***), an ontology store (***BoKOnto***), and a series of BoK web services that can be ingested by other applications, as demonstrated by a visualization application (***BoKVis***) (Figure 5).

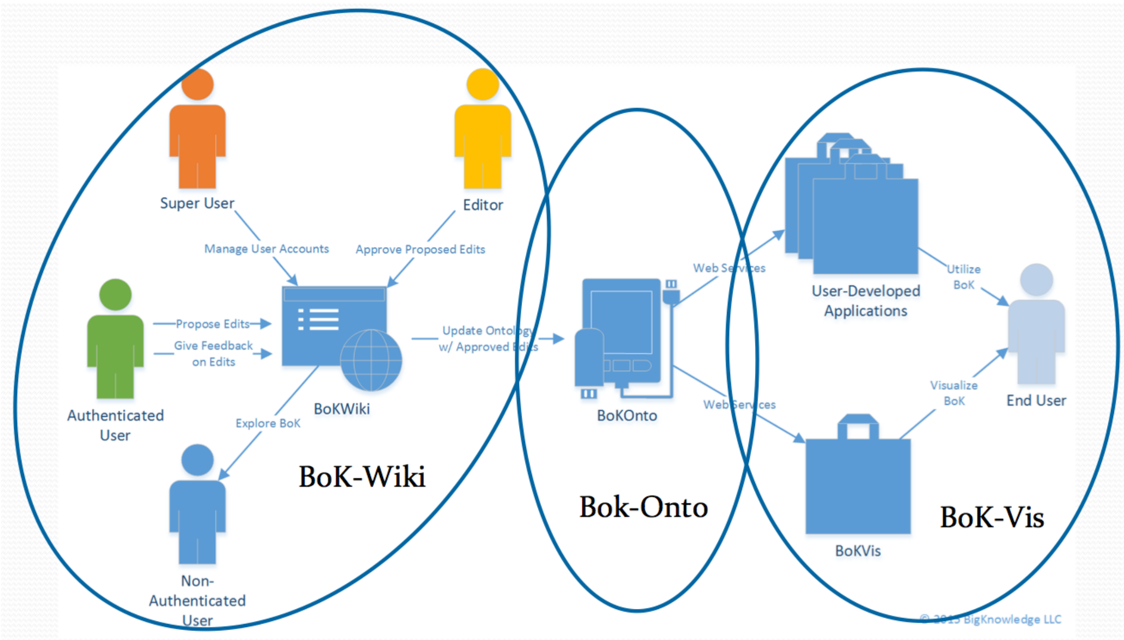

Figure 5. BK-BoK System Architecture

BoKWiki is the environment in which the knowledge space can be explored by a viewer of the system, in which a contributor can make suggestions for new concepts, new relationships or to edit existing ones, and in which editors can approve changes made by the contributors. Upon approval, which is triggered by a click of a button, a service is called for moving modifications over to ***BoKOnto***, which holds the authoritative version of the BoK. If such modifications are additions than the initial versions of corresponding concepts and associated triples are created. If changes are made to existing concepts than a new version number is assigned to the triples and the older version is saved. In this manner, all changes are version-managed and changes to the BoK over time can be tracked and analyzed. This permits the creation of a view of the BoK for any date in time. ***BoKWiki*** is then updated with the new changes.

BoKOnto has a set of services that enable the connection to any applications that a user may wish to develop. At its simplest level a service can provide a complete XML file of the entire Body of Knowledge or any other format that might be desired.

5 Bok to Base Map

The notion of a base map has long been central to the practice of cartography and GIS. It is meant to provide a stable, foundational spatial platform, onto which thematic layers can be projected. Such projections are effectively locational inferences derived from entities represented in certain foundational reference systems. For example, a geographic phenomenon can be mapped onto the base map by first determining its location in latitude and longitude (i.e., the foundational reference system) and then determining its location in the projected space of the base map (i.e., locational inference). Analogously, when a base map is created for a knowledge domain, then one can, for any given knowledge artifact, actor, or activity [1], first determine its location in the knowledge reference system and then infer the phenomenon's location in the base map. How can such knowledge reference systems and base maps be created? The following outlines two approaches, one directly based on a formally defined domain BoK and its prescribed ontological structure, the other based on mining of domain artifacts. Hybrid approaches are possible as well, such as when Bodies of Knowledge themselves are subjected to content analysis [1].

5.1 From BoK to Base Map

Ahearn et al. 2013 created a service-oriented architecture for a domain knowledge ontology based on content from the 2006 Geographic Information Science and Technology Body of Knowledge [7]. This approach expanded the scope of not only how one could *update* and *maintain* any BoK on a on ongoing basis, but also how and for which purposes a BoK could be *operationalized*. A key contribution of that effort was the transfer of geographic and cartographic principles to the *contextualization*, mapping and *visualization* of a knowledge domain ontology. By conceptualizing a formally defined body of knowledge (an ontology) as the foundation of a "spatial reference system," arbitrary knowledge artifacts can be readily related to this domain reference system, and thereby to each other.

In terms of visualization, this reference system approach links up with the notion of a base map that in a visual context is kept relatively simple (Figure 6), but provides the ability to perform inference and overlays derived from user queries (Figure 7) or even whole knowledge artifacts.

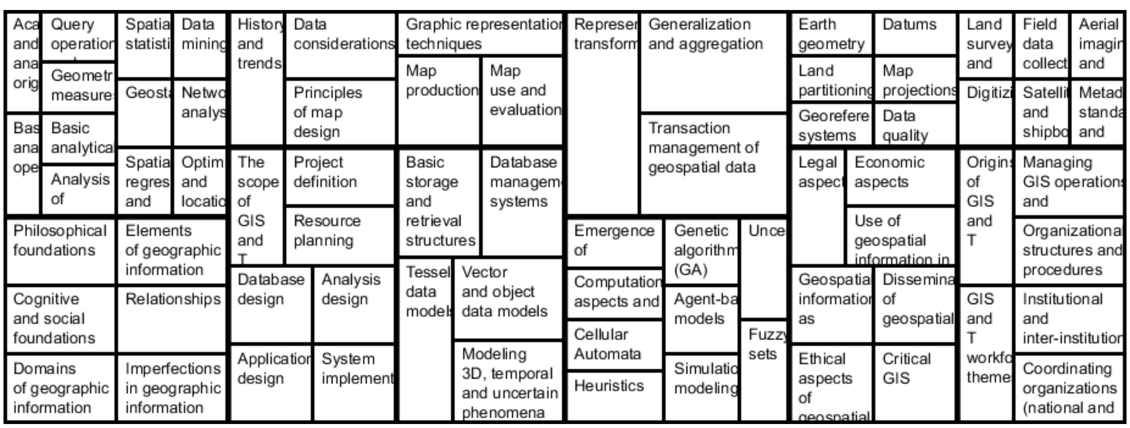

Figure 6. Simple base map derived from a BoK and implemented as treemap. BoK hierarchy is ingested from BoK Web service.

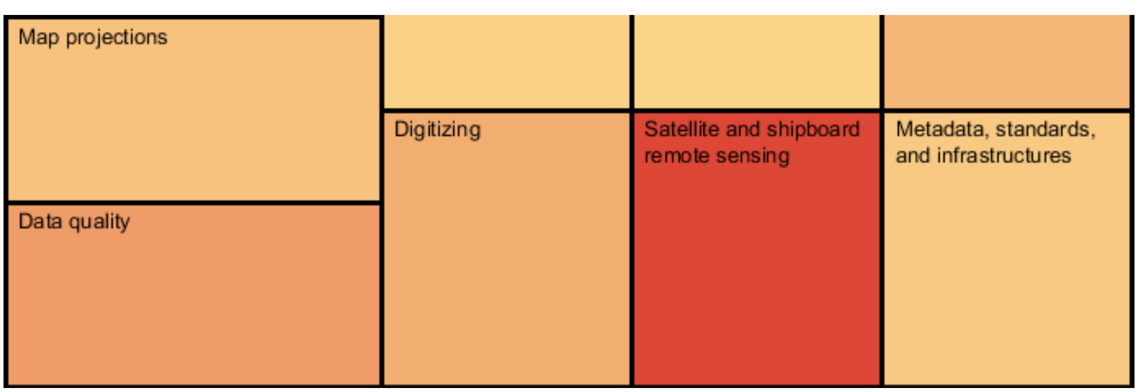

Figure 7. Visualization derived from two different BoK Web services: BoK concept tree service and topic inference service. Overlaid is a query for "remote sensing", with

higher red color value indicating a stronger similarity match of the concept to the query. Lower red color values indicate a weaker match.

A domain ontology based approach allows extending domain knowledge projection beyond domain *concepts*. For example, one can link up with domain-specific skills, as illustrated in the sorted listing of GIS&T skills in response to a query for "remote sensing" (Figure 8). Note that the query phrase does not have to be verbatim contained in the skills description. Instead of performing client-side text matching, an inference service is invoked whose results are then displayed.

The combination of a knowledge reference system approach with base mapping and inference services also enables more complex knowledge-algebraic operations, such as the explicit comparison of two courses (Figure 9) or of two persons or of an individual's expertise vis- à-vis the stated requirements of a job. These kinds of projection and overlay operations enables user applications that operate on the knowledge spaces that we intellectually and productively "inhabit" [1).

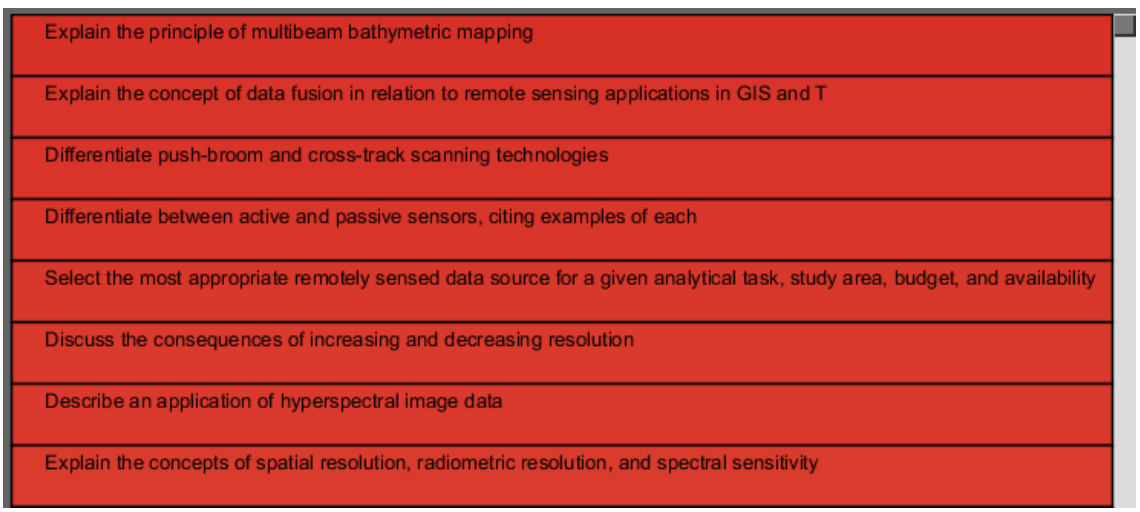

Figure 8. Snapshot of an interactive visualization of the BoK concept tree service and topic inference service combined in a scrollable list display. Shown are not concepts, but skills retrieved and sorted in response to query for "remote sensing". Only the top of the ranking is shown in figure, listing the best matching learning objectives. Applying a red value color scheme to several hundred learning objectives results in the top-ranked objectives being displayed in full red color.

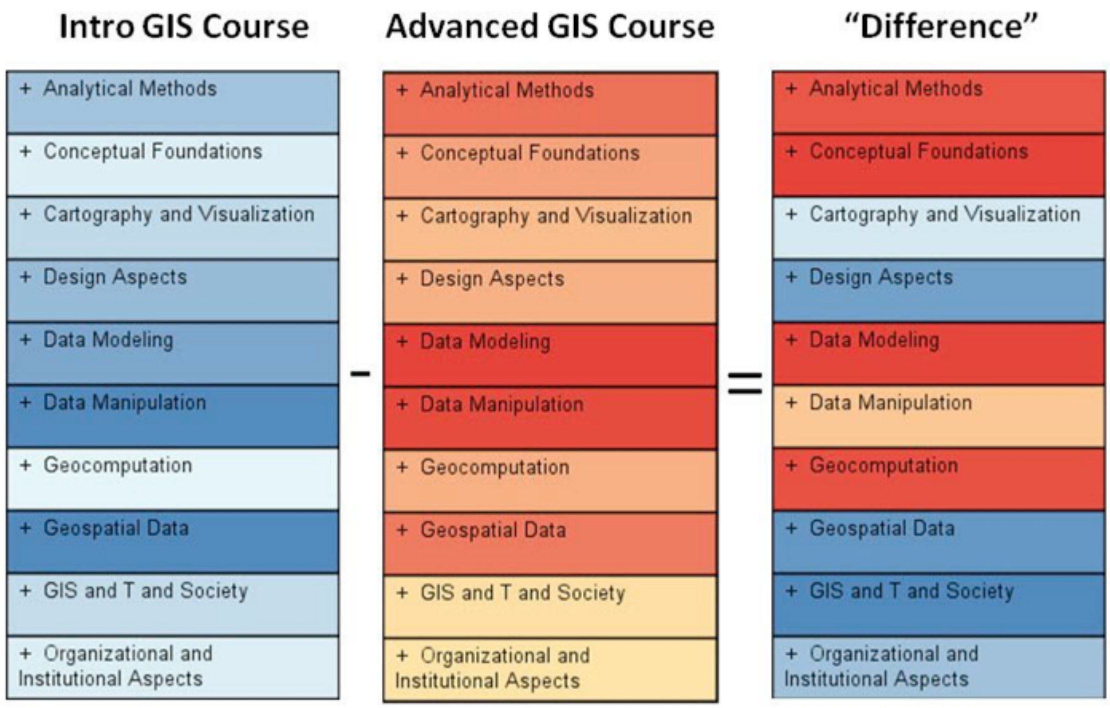

Figure 9. Knowledge algebra operation applied to two GIS courses in a reference system derived from the GIS&T BoK and projected onto a coarse domain base map through natural language inference. For individual courses, darker colors indicate a better match to a GIS&T BoK knowledge area. In the course "difference" overlay, a divergent color scheme indicates that dark blue topics (e.g., "Geospatial Data") are more strongly associated with the Intro GIS class, while the dark red topics (e.g. "Geocomputation") are associated with the Advanced GIS class.

5.2 From Book to Base Map

The reference system and base map notions introduced by [1] are applicable even when no formal domain ontology exists or if its current content does not reflect the evolution of the domain. In that case, content mining comes to the fore and could be applied to any collection of domain artifacts, including knowledge canonized in text book form (figure 10a) and formal (Figure 10b) and informal collections (Figure 10c) of domain writings.

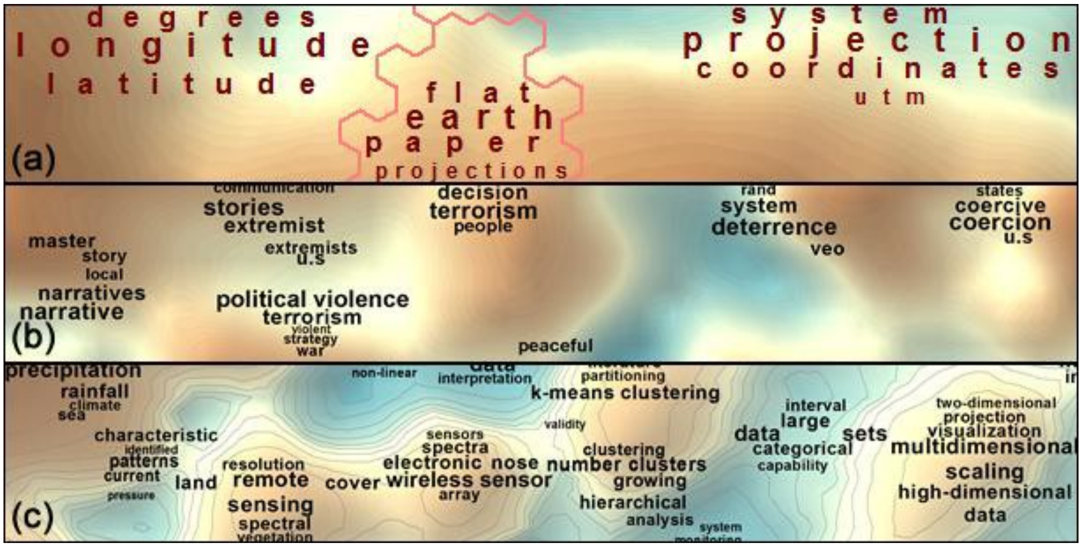

Figure 10. Snippets of base maps derived from domain artifacts for three different domains and artifact types: (a) GIS textbook, (b) an edited collection of research papers on violent extremism, (c) corpus of independently authored research papers on self-organizing maps. Contours, coloring, and hill shading derived using a term dominance landscape approach [14, 16].

These base maps and underlying reference systems can reflect recent advances in domain knowledge that have not yet been captured in a formal manner in a BoK. By projecting an existing BoK into such an artifact-derived reference system, it becomes possible to either confirm the continued relevance of topics already contained in a BoK or detect new domain structures (indicated by significant base map / reference system structures void of coverage by the existing BoK) or even help identify concepts as being outdated or deprecated within the domain. Intelligent use of natural language processing – as opposed to simple string matching – can ensure that this works even in the presence of common domain language issues, like synonyms (e.g., "conformal" and "angle-preserving") and homonyms (e.g., "map" in different domains). Another, mostly unexplored, possibility consists of using this "book to base map" approach to identify key concepts and structures as a starting point in the creation of a completely new BoK, especially for novel domains, in which there is a lack of canonized knowledge resources.

6 Discussion and Conclusion

Catalysts for creative thought and the emergence of new ideas are still something of a mystery. As Salman Rushdie once said “a little bit of this and a little bit of that is how new ideas come into the world”. More and more we see that the old paradigm of disciplinary and sub-disciplinary silos that don’t interact, is fading. In fact we had an early sense of the power of interdisciplinary research when such scientists like Amar Bose and Noam Chomsky were crowded out of their space at MIT and placed in a barely serviceable edifice called Building 20. A heterogeneous group of researchers from different disciplines were working in a cramped space that "forced solitary scientists to mix and mingle" [9]. This resulted in some of the most innovative research in decades. What does the new Building 20 look like? Is it an abstract space in which diverse constituencies can act based on shared knowledge? What would it take to conceptualize, structure, and populate a single space (Figure 11) in which in domain actors, from students to educators, researchers and professionals, can act and interact in an manner that is more efficient and, yes, catalytic, than allowed by current knowledge ecosystems?

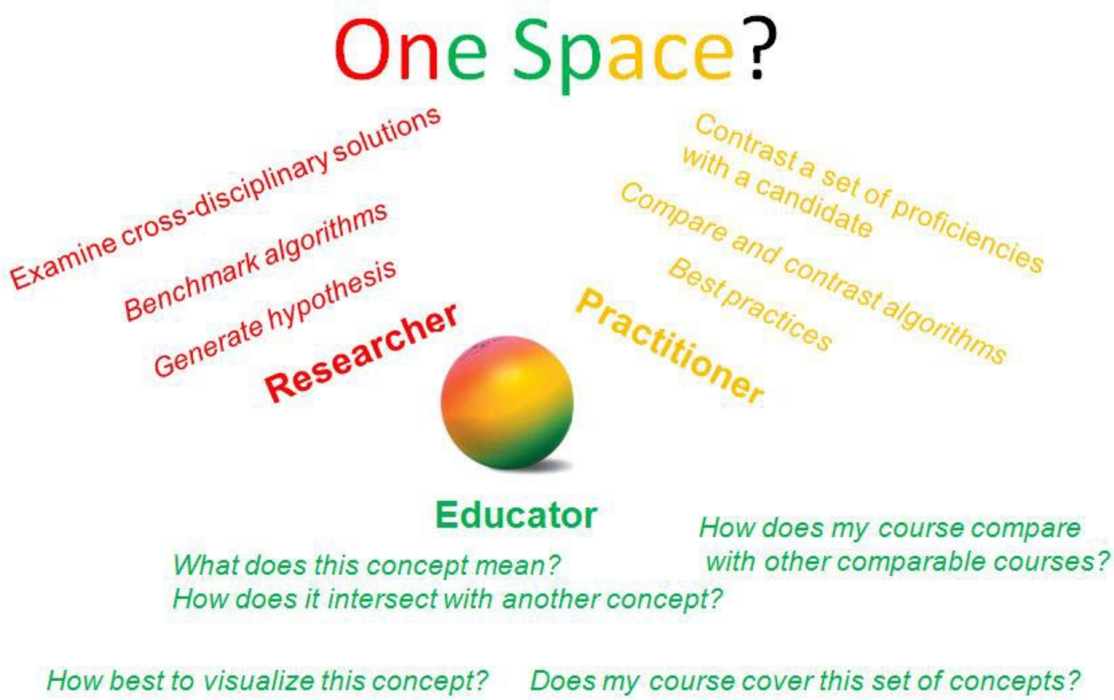

Figure 11. One Space for Education, Research, and Practice

6.1 Making Space for Knowledge

We believe that underlying such a shared space must be a knowledge reference system that can simultaneously encapsulate, canonized and uncover domain knowledge; support computational inference (e.g., the representation of any artifact); in an environment whose spatiality is made tangible through visualization.

Cartography, geography, and geographic information science have a key role to play in conceptualizing and implementing this vision, but they can also catalyze the injection of key ideas into efforts of computer scientists, information scientists, and non-geographic domain specialists. For example, the *base map* notion is being increasingly adopted in information visualization ([3-5][13]), in a marked departure from common practices in that domain. That the *base map* and *reference system* concepts are sometimes conflated raises the need for GIScientists to clarify how measurement, projection, distortion, and standardization can be usefully addressed in any domain that is attempting to be spatialized in the broadest sense. That effort can itself result in new techniques [15] and new interdisciplinary collaborations [16].

Canonization of domain knowledge through formal processes of concept elicitation and structuring – the creation of a Body of Knowledge – is a useful step towards operationalizing these ideas. This requires involving broad constituencies, in order to foster a *sense of place* within the resulting knowledge spaces, lest those spaces should remain "foreign lands" even for domain insiders. Wikification is a key strategy for capturing the breadth of domain concepts and for generation buy-in from the community. Another requirement is for the resulting knowledge structures to be made accessible to domain services, from human resource management to coordination of research activities. If a BoK is to form the sustainable heart of a domain knowledge ecosystem, then it has to be accessible through a variety of means and for many purposes, from knowledge management to exploration and analytics. The ***BK-BoK*** system discussed in the article provides such a framework, since it provides a front-end visual wiki (***BokWiki***), an ontology-driven knowledge base (**BoKOnto**), and a service-based architecture for enabling applications.

6.2 Putting Domain Knowledge to Work for Education

Almost without exception, most BoKs have typically been created with the explicit and exclusive goal of supporting tasks in education, especially curricular planning. This has included efforts in the GIS&T domain [10-12]. Though [1] presented a vision significantly expanded beyond this, education and training is an arena where an overtly knowledge-centric approach to integration of actors, artifacts, and activities, including an elaboration of canonized and emergent practices, shows particular promise.

A range of novel education applications can be envisioned, involving numerous elements of higher and vocational education, from faculty and administrators to students and the learning infrastructure, such as courseware. An operationalized domain knowledge reference system allows breaking down traditional barriers in the knowledge ecosystem. What if the theory-laden, long-term approaches of *education* and the hands-on, short-term view of *training* could be represented in a single space? What if core concepts could be explicitly linked to hands-on tools and real-world *practices*? For example, how could students quickly find mappings between the concept of "functional distance," a GIS software tool for computing "cost distance," and a map of hospital service areas published by a county health and human services agency? Other types of mappings supported by a reference system and base map approach include course articulation and course equivalency, which can now be represented in overtly spatial terms, by identifying overlaps and gaps.

The explicit linking of concepts, skills, tools, and literature through ontological relationships allows tracking students' progress in a number of ways. For example, instructors could project individual performance results into the knowledge space to identify which tasks students found especially challenging and take note of the associated prerequisite concepts. This would create a whole new dynamic in pedagogy: interactive, expansive, and comprehensive. The ability to project learning outcomes into a common space also allows comparison of different instructional modes, such as traditional lecture-lab approaches versus the various forms of flipped learning [2].

References

[1] AHEARN, S.C., ICKE, I., DATTA, R., DEMERS, M.N. PLEWE, B., AND SKUPIN, A. (2013) Re-engineering the GIS&T Body of Knowledge. International Journal of Geographical Information Science 27(11), 2227-2245.

[2] BISHOP, J.L. AND VERLERGER, M.A. (2013) The Flipped Classrom: A Survey of the Research. ASEE Annual Conference and Exposition, June 23-26, 2013, Atlanta, Georgia.

[3] BÖRNER, K. (2010) Atlas of Science: Visualizing What We Know. MIT Press.

[4] BÖRNER, K. (2015) Atlas of Knowledge: Anyone Can Map. MIT Press.

[5] CHEN, C., CRIBBIN, T., MACREDIE, R., AND MORAR, S. (2002) Visualizing and Tracking the Growth of Competing Paradigms. Journal of the American Society for Information Science and Technology 53(8): 678-689.

[6] CICCARESE, WU, E., WONG, G., OCANA, M., KINOSHITA, J., RUTTENBERG, A., AND CLARK, T. (2008). The SWAN biomedical discourse ontology, Journal of Biomedical Informatics 41:739–751.

[7] DIBIASE, D., DEMERS, M., JOHNSON, A., KEMP, K., TAYLOR LUCK, A. PLEWE, B. AND WENTZ, E., eds. (2006) Geographic Information Science and Technology Body of Knowledge. Association of American Geographers.

[8] FOOTE, K. E., et al. "GIS&T in higher education: Challenges for educators, opportunities for education." Teaching geographic information science and technology in higher education (2012): 3-15.

[9] GAO, Y., KINOSHITA, J. , WU, E., MILLER, E., LEE, R., SEABORNE, A., CAYZER, S., CLARK, T. (2006) SWAN: A distributed knowledge infrastructure for Alzheimer disease research, ScienceDirect, Web Semantics: Science, Services and Agents on the World Wide Web, 4:222-228.

[10] LEHRER, J. (2012) GROUPTHINK: The Brainstorming Myth. The New Yorker 87(46): 22-27.

[11] PRAGER, S., AND PLEWE, B. (2009) Assessment and Evaluation of GIScience Curriculum using the Geographic Information Science and Technology Body of Knowledge, Journal of Geography in Higher Education, 33: 46-69.

[12] PRAGER, S. (2011) Using the GIS&T Body of Knowledge for Curriculum Design: Different Design for Different Contexts. In: Unwin, D.J., Foote, K.E., Tate, N.T., and DiBiase, D. (Eds.) Teaching Geographic Information Science and Technology in Higher Education. Chichester, UK: John Wiley & Sons, Ltd. 63-80.

[13] RIP, F. (2008) GIS&T Body of Knowledge: basis for e-learning, certification and curriculum planning? Proceedings of the Eleventh Annual AGILE Conference.

[14] SKUPIN, A. (2000). From metaphor to method: Cartographic perspectives on information visualization. In IEEE Symposium on Information Visualization (InfoVis 2000), IEEE: 91-97.

[15] SKUPIN, A. (2004) The World of Geography: Visualizing a Knowledge Domain with Cartographic Means. Proceedings of the National Academy of Sciences. 101 (Suppl. 1): 5274-5278.

[16] SKUPIN, A. (2009). Discrete and continuous conceptualizations of science: Implications for knowledge domain visualization. Journal of Informetrics, 3(3): 233-245.

[17] SKUPIN, A., BIBERSTINE, J. R., AND BÖRNER, K. (2013). Visualizing the topical structure of the medical sciences: a self-organizing map approach. PloS ONE, 8(3), e58779.