Unleashing the Power of Data Through Organization: Structure and Connections for Meaning, Learning and Discovery†

Dagobert Soergel

University at Buffalo, Department of Library and Information Studies,
Graduate School of Education, 534 Baldy Hall, Buffalo, NY 14260-1020, USA
<dsoergel@buffalo.edu www.dsoergel.com>

Dagobert Soergel is Professor, Department of Library and Information Studies, University at Buffalo 2009- and Professor emeritus, College of Information Studies, University of Maryland (1970-2010). He has been working in the area of classification and thesauri for 55 years. He authored the still-standard text and handbook *Indexing Languages and Thesauri. Construction and Maintenance* (Wiley 1974) and *Organizing Information* (Academic Press 1985, American Society of Information Science Best Book Award), and 100+ papers and presentations in information retrieval, classification/ontologies, digital libraries, sense making, and relevance. He received the American Society for Information Science Award of Merit in 1997.

Soergel, Dagobert. **Unleashing the Power of Data Through Organization: Structure and Connections for Meaning, Learning and Discovery.** *Knowledge Organization. 42(6)*, 401-427. 65 references.

Abstract::

The Future of Knowledge Organization

Knowledge organization is needed everywhere

Create the future of KO
Think BIG. Think answers not pointers.
Focus on substantive data
Many areas, tasks, and functions that could profit from KO principles
Engage with ontologies, AI, data modeling

Knowledge organization is needed everywhere. Its importance is marked by its pervasiveness. This paper will show many areas, tasks, and functions where proper use of knowledge organization, construed as broadly as the term implies, provides support for learning and understanding, for sense making and meaning making, for inference, and for discovery by people and computer programs and thereby will make the world a better place. The paper focuses not on metadata but rather on structuring and representing the actual data or knowledge itself and argues for more communication between the largely separated KO, ontology, data modeling, and semantic web communities to address the many problems that need better solutions. In particular, the paper discusses the application of knowledge organization in knowledge bases for question answering and cognitive systems, knowledge bases for information extraction from text or multimedia, linked data, big data and data analytics, electronic health records as one example, influence diagrams (causal maps), dynamic system models, process diagrams, concept maps, and other node-link diagrams, information systems in organizations, knowledge organization for understanding and learning, and knowledge transfer between domains. The paper argues for moving beyond triples to a more powerful representation using entities and multi-way relationships but not attributes.

Received: 19 August 2015; Accepted 18 August 2015

Keywords: data, knowledge organization, information, entity, types

† I thank Stella Dextre Clark for many useful editing suggestions and Olivia Helfer for assistance in finding references.

Cab driver: Where to?
Chancellor's Chief of Staff: It doesn't matter. I am needed everywhere.
Apocryphal

1.0 Introduction

Knowledge organization is needed everywhere, it is pervasive. There are many areas, tasks, and functions where knowledge organization can make the world a better place through support for daily operations (treating patients, making a loan), for learning and understanding, for creating meaning, for sense making, for decision making, for inference, and for discovery by people and computer programs.

This paper is based on a keynote address presented at the Fourth Biennial ISKO UK Conference on 14 July 2015: "Knowledge organization–Making a difference: the impact of knowledge organization on society, scholarship and progress." The keynote was "to address the role that KO should have in the future," "the opportunities that lie ahead for KO, and what difference it could really make for economic, scientific and/or cultural development" (http://www.isko2014.confer.uj.edu.pl/en_GB/o-konferencji) To create the future of knowledge organization we should 1) broaden the application areas where the KO community contributes its much-needed expertise and 2) expand KO theory in the process. Section 2 discusses such applications and illustrates how KO can contribute. Section 3 discusses general issues in data modeling and KO that need to be addressed to fully reach the potential of these applications, challenging some long-standing, widespread, and firmly ingrained ways of thinking, particularly in entity-relationship (E-R) modeling. Section 4 presents conclusions for the role of KO professionals and implications for education. This is not a detailed review but rather a broad overview and framework, putting known detail into a new perspective.

1.1 Knowledge and its organization

Knowledge is a representation—in the mind, in computer systems, or on paper—of the "real world" (physical, social) and of assumptions, beliefs, thoughts, feelings, emotions of others or our own. Knowledge is a basis for action, an influence on one's mental state, a source of enjoyment, and more. Knowledge can be represented through a distribution of node weights and connection strengths in a neural network or by a set of propositions. Often a distinction is made between data, information, knowledge, and wisdom. While superficially this distinction may seem helpful, it oversimplifies important distinctions between pieces of knowledge based on several characteristics. From the perspective of knowledge organization the distinction is irrelevant. Data from empirical observations, conclusions drawn from such data, and sage advice on how to conduct one's life—all can be formally represented by a set of propositions.

Knowledge organization, broadly construed, deals with methods for representing and structuring a set of propositions. This requires knowledge about knowledge in two layers:

1. Entity-Relationship (E-R) schema layer: a model of a domain specifying the "types of entities" the domain covers (such as person, place, date, chemical substance, disease, concept/topic) and the "types of relationships" that can be used to connect two or more entities into propositions (such as person <"livesIn"> (place <"from"> Date <"to"> Date) or chemical substance <"treats"> disease. Such a model is called an entity-relationship (E-R) "schema" or "data model" or "ontology" (in one meaning of that word) or, in the Web context, "vocabulary." The schema may also include integrity rules, which specify for each relationship the types of entities it connects (for binary relationships: domain and range).
2. Authority layer: Lists of individual "entity values" and their names or identifiers so that propositions about these individual entities can be formulated. For example, we need to know the persons and places to be covered with their names and universally consistent URIs (universal resource identifiers). Providing these entity values is the function of "authority lists," including "name authorities" (such as maintained by the Library of Congress), "gazetteers," "thesauri," "classifications" (also called ontologies), such as the Universal Decimal Classification (UDC), SNOMED CT (systematized nomenclature of medicine-clinical terms), and the harmonized commodity description and coding system. Such systems often contain also world knowledge.

To sum up (Soergel 2009, 3), "knowledge organization systems (KOS[s]) cover a wide range of systems at both layers serving a wide range of purposes. They are known under names such as ontology, metadata schema, taxonomy, classification, Web directory structure, filing plan, thesaurus, dictionary, folksonomy, and more."

1.2 The wide scope of application of knowledge organization (KO)

In the KO community, emphasis has been on metadata and on KO to support searching; there are good reviews on this (for example, with a focus on the bibliographic universe, Dunsire et al. 2012). There is also much discussion on whether manual indexing/subject cataloging is still needed (either way, we need KO to assist users in search) and the role of KO in social tagging (KO can assist taggers to be more effective, thereby increasing ROI (return on investment) on the enormous investment, and KO is needed to assist searchers in navigating the unordered mess of tags). This is all very useful, but the scope is too narrow. For the future we need to focus also on structuring and representing the actual data or knowledge itself, issues we have left to the ontology, artificial intelligence, and data modeling communities. These are the issues addressed in this paper.

There is a wide range of KO applications and concomitant opportunities for KO professionals:

1. Knowledge bases for question answering and cognitive computing.
2. Linked data.
3. Information systems in the enterprise.
4. Interoperability of data across operational information systems. The example of patient data (Electronic Health Records, EHR).
5. Big data and data analytics. Data interoperability and re-use.
6. Knowledge bases for information extraction from text and multimedia.
7. Influence diagrams (causal maps), system dynamics models, process diagrams, concept maps, and other node-link diagrams.
8. Knowledge organization for understanding and learning.
9. Knowledge transfer between domains.

Knowledge organization systems (KOSs) are used in these applications, but in many cases these KOSs could be improved, using the expertise present in the KO community.

2.0 Advanced applications of knowledge organization

Key idea: data interoperability-systems must be able to use data from another system, and data from several sources must be mergeable and linkable. Organizing knowledge well is key.

2.1 Knowledge bases for question answering and cognitive computing

Powered by its knowledge graph, Google now gives not just Web pages where answers can be found, but answers, biographical synopses for people, or basic data on cities. The push by information companies goes beyond giving just sources to providing answers, the semantic web, and the quest for practical applications of artificial intelligence have spurred the development of many knowledge bases (Figure 1). They all use knowledge organization systems, and all their KOSs could be improved in their structure, made more interoperable, and presented better to support searchers in exploring the conceptual space and formulating queries.

CYC's very large ontology includes entity and relationship types and entity values, for example:

Knowledge base	URL	Some KOSs used
CYC Common sense knowledge	http://www.cyc.com/ http://sw.opencyc.org/	CYC Ontology, including entity types, relationship types, and entity values
IBM Watson Custom KBs for applications	http://www.ibm.com/smarterplanet/us/en/ibmwatson/	an extensible inventory of relationship types
Google Knowledge Graph Huge database of varied kinds of data (Starr 2014)	http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html	schema.org for entity types and relationship types
dbpedia Large database of statements extracted from Wikipedia	http://wiki.dbpedia.org/ http://mappings.dbpedia.org/index.php/Exploring_the_Ontology	DBpedia Ontology (E-R schema) Authority lists for individual entity values (instances), each identified by a URI.
GDELT Sentiment data, Event reports	http://gdeltproject.org/	CAMEO Coding Scheme for events 2,300 emotions and themes) World Bank Group Topic Taxonomy
Many interoperable data sets on gene products		The Gene Ontology (GO) www.geneontology.org/index.shtml

Figure 1. Some universal knowledge bases.

- social being: A specialization of intelligent agent. Each instance of social being is an intelligent agent whose status as an agent is acknowledged within some social system and who is capable of playing certain social roles within that system.
- controlling: (controls AGENT THING) means that AGENT has effective control over THING.

CYC was created mainly by people from the ontology community; its organization could profit from people from the KO community.

IBM's Watson is a platform for developing custom knowledge bases for cognitive computing. It was demonstrated and made known when it competed successfully in the trivia game *Jeopardy* using a large KB assembled through a combination of methods, including information extraction (Section 2.6). (See Ferucci 2012 and the other papers in the same journal issue.)

DBpedia has a plethora of data, some from the formatted Wikipedia infoboxes, some from information extraction; DBpedia is available and searchable as linked data. It uses its own ontology, an E-R schema with a proliferation of entity types (classes) and relationship types. Entity types include "place," "species," "person," "organization," "work," and "topical concepts," each with many subdivisions. For each entity type, DBpedia has many instances (entity values) (e.g., ~250,000 individual organizations. DBpedia is one source of the Google knowledge graph.

GDELT (global database of events, language, and tone) is a huge repository of formatted data extracted from newspapers across the world. It includes for each news item data about sentiments expressed and multiple event reports, a summary of data about an event as reported in this news item. It would be nice to have data on the same event integrated. GDELT uses GNIS and GNS, geographic place authorities produced by the US Geological Survey and the US National Geospatial-Intelligence Agency, respectively; the GDELT list of emotions pooled from 24 sentiment analysis packages is a great candidate for semantic unification (Section 3.4.2).

The Gene Ontology is a classification of "cells and cell parts," "molecular functions," and "biological processes" widely used to record data on gene products found in biological research so that many data sets are interoperable. It would benefit from facet analysis.

2.2 Linked data

Linked data form a knowledge base for question answering and cognitive computing in a (so far) limited format. The vision of linked data is that many data sets will be posted openly accessible on the Web so they can be searched together, linking across data sets, enabling more complex searches with more complete answers. Linked data are a key implementation component of the semantic web. Many KOSs are now available as linked data, promoting their use.

A linked data set consists of very simple propositions, such as

Acyclovir <"treats"> VirusInfection

a simple sentence consisting of a verb (a relationship, also called property) connecting two nouns (entities). These propositions are called (RDF) triples. RDF is an E-R model restricted to two-way relationships (with work-arounds for representing multi-way relationships). This simplicity is at once the reason for the success and the bane of linked data (see Section 3.1). Linked data (from one or more data sets) can be searched with the query language SPARQL. For a more general look at the underlying data model and a vision of expanded search possibilities see Soergel (2011).

To reach their full potential, linked data need standardization through properly structured and often very large KOSs, presenting an enormous opportunity for KO. Ideally, all linked data sets would use the same universal extensible E-R schema and universal authority lists:

1. Use the same universal extensible entity-relationship schema. Some partial schemas (vocabularies) are widely used, such as "foaf" (Brickley and Miller nd) and "vcard" for data on people; the *W3C Organization Ontology* (WW3C 2014); the *Dublin Core*, *BIBO*, and *FaBiO* for bibliographic data (Peroni 2014); *SKOS* (simple knowledge organization systems) (W3C 2009, DuCharme 2011). But despite re-use of schema pieces and faint attempts at integration, such as schema.org, there is a wilderness (see Figure 2).
2. Use universal authority lists for "places," "chemical substances," "species," "persons," "organizations," "works," "topical concepts," etc. that give entity values as URIs to be used consistently across the Web. Thus triples on the same entity value from different linked data sets can be easily combined. Multiple linked data sets can be traversed as one big graph. Such large authority lists must be created and agreed on (See Section 3.4).

Many linked data sets do not document what entity types and relationship types and what, if any, authorities for entity values they use There is much room for improvement For more information on linked data and their uses, see Alexander et al. (2009), Davies and Edwards (2012), Jingyuan et al. (2014), Sören et al. (2013), and Baierer et al. (2014).

www4.wiwiss.fu-berlin.de/dailymed		**www4.wiwiss.fu-berlin.de/drugbank**	
Drug <*hasName*>	Text	DBDrug <*hasName*>	Text
Drug <*hasGenericVersion*>	Drug	DBDrug <*hasGenericName*>	Text
		DBDrug <*hasCASRegistryNumber*>	URI
Drug <*hasActiveIngredient*>	ChemicalSubstance		
Drug <*hasClinicalPharmacologyDescr*>	Text	DBDrug <*hasAbsorptionDescr*>	Text
		DBDrug <*hasBioTransformDescr*>	Text
		DBDrug <*hasPharmacolDescr*>	Text
		DBDrug <*hasProteinBindRate*>	Pct
Drug <*hasIndicationDescr*>	Text	DBDrug <*hasIndicationDescr*>	Text
		DBDrug <*hasPossibleDiseaseTarget*>	Disease
Drug <*hasContraIndicationDescription*>	Text	DBDrug <*hasContraIndicationInsert*>	Document
Drug <*administeredVia*>	RouteOfAdministration	DBDrug <*hasDosageForm*>	DosageForm

Figure 2. Different relationship types used in two databases about drugs. Some relationship types correspond and are named the same, some correspond but are named differently, and others do not correspond at all.

2.3 Information systems in the enterprise

2.3.1 Enterprise data registry through an enterprise ontology (E-R data schema)

Many companies, government agencies, and other enterprises lack a registry of all their data systems and data files, the individual data items in those files, and the standards used for entity types, relationship types, and entity values. This makes it nigh impossible to avoid duplication, to find data, and to combine data from multiple sources to solve a problem. Chief information officers work on addressing this problem. In the database world, the idea of a data dictionary was developed to solve it. Today the best approach uses two steps:

1. Develop a comprehensive ontology (an E-R conceptual data schema, a data model) for the entire enterprise, re-using pieces or at least re-using ideas, from standard schemas, such as foaf, SKOS, and the W3C Organization Ontology Recommendation. This schema will be quite large to present a view of all the data the enterprise keeps in minute detail.
2. Use the schema to organize a registry of data systems and the types of information each includes.

Both the ontology and the registry must be continuously updated.

2.3.2 KOS support for merging and linking data in the enterprise

The World Bank Group (WBG) has a number of databases that include connections to organizations in their data, as shown in Figure 3.

Each of these databases uses its own list of organizations and organization identifiers. It has been proposed to build instead a central "WBG Organization Authority Database" (OAD) as follows:

1. Import organization information from all WBG sources and from external sources (such as DBpedia; Library of Congress Name Authority, corporate headings; Dun & Bradstreet). Merge information on the same organization from several sources using automatic mapping based on organization name with manual edit. Give each organization a URI, using what exists if possible. Maintain the source of each piece of data.
2. Determine for each organization a preferred and alternate names in multiple languages.
3. Link organization URIs to identifiers used in WBG systems and to Dun & Bradstreet ID (which is used in WBG systems and heavily in external systems). This enables linking all organization data in WBG and external systems, such as DBpedia.

Figure 4 shows a few sample relationship types from the E-R schema or ontology envisioned for OAD, to be constructed drawing on widely used schemes. This schema deliberately eschews constraints imposed by predominant practice to better represent reality (see Section 3.1).

2.4 Interoperability of data across operational information systems. The example of patient data (electronic health records, EHR)

Interoperability of data is the ability to use data produced by one system in another system (for example, using pa-

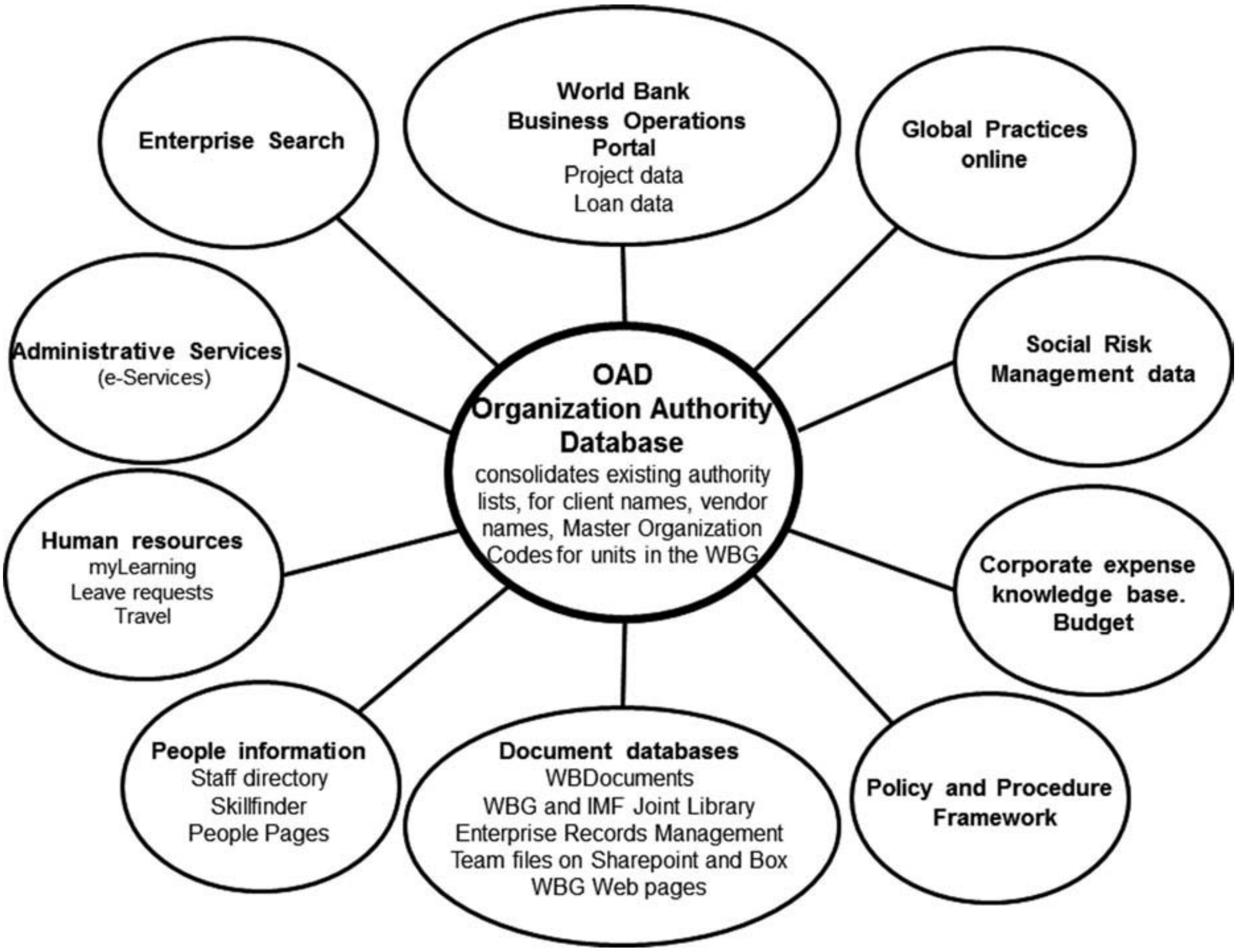


Figure 3. A central organization authority database (Soergel and Popescu 2015).

tient data produced by one hospital by another hospital in the treatment of the patient) or the ability to integrate or combine two data sets to use in a third system, for example (Fridsma 2013), to integrate patient data from several hospitals into one large data set for medical research. Interoperability of data is a pervasive problem. The solution depends crucially on KO.

EHR interoperability (Ceusters et al. 2009; Moreno-Conde et al. 2015, Smith et al. nd) is a huge problem in health care and biomedical research. An EHR uses many KOSs, for example:

– race/ethnicity, age, sex
– bodily or mental functions or conditions (measured through patient perception, observation, or a measuring device or test and possibly used in diagnosis)
– diseases
– medical procedures
– drugs

Using different (sometimes proprietary) KOSs creates problems when EHR data are transferred from one health care provider to another. Example: You require emergency treatment while traveling. The hospital cannot easily interpret the EHR from your hometown hospital; the physician misses a condition you have and treats you with a commonly prescribed drug that has only one drawback: for people with your condition it is fatal. Lack of interoperability can kill.

2.5 Big data and data analytics. Data interoperability and reuse

Big data means managing very large and often varied data sets mostly for use in data analytics. Data analytic refers to sophisticated analysis 1) to find patterns that can be detected only in large data sets, or 2) to find cases similar to a given rare situation (a patient with a rare genetic make-up, a student with unusual characteristics, an unusual business problem) to discover what worked in these known cases to solve a problem.

There are three characteristics that contribute to making data "big:"

Entity	<*hasID*>	(ID, System, TimeSpan)
Entity	<*isPartOf*>	(Entity, TimeSpan)
Organization	<isS*ubsidiaryOf*>	(Organization, TimeSpan)
Organization	<isS*uccessorOf*>	(Organization, TimeSpan)
LegalEntity	<*hasAddress*>	(Location, Work/Home, TimeSpan)
LegalEntity	<*hasPhoneNumber*>	(PhoneNumber, Voice/Fax, Home/Work, Land/Mobile, TimeSpan)
Organization	<*hasName*>	(Text, Language, NameStatus. TimeSpan)
Organization	<*hasMember*>	(LegalEntity, TimeSpan)
Organization	<*hasStaffMember*>	(Person, InOrganizationRole, TimeSpan)
Entity	<*isAbout*>	(Entity, TimeSpan)
LegalEntity	<*isResponsibleFor*>	(Entity, TimeSpan)
Event	<*hasParticipant*>	(LegalEntity, EventRole, TimeSpan)

Some observations on the schema: "LegalEntity" includes Person and Organization; this is often called Agent, but agent implies a role. LegalEntity is a Thing that is capable of playing the agent role. The relationship type <"hasphonenumber"> allows for storing any kind of phone number, rather than defining a separate relationship type for each kind. The TimeSpan argument is included for every relationship type for clarity. This could also be stated as a general rule.

Figure 4. Example elements of an organization data base ontology.

– large size of a data set – hundreds of megabytes, terabytes, petabytes;
– variety-data of different kinds (for example, from medical records, from social work records, from school records;
– velocity-speed and size of updates.

The big data and data analytics "movement" arose due to two main developments:

– There are many data sets that can be shared. Many funders require sharing raw research data. Hospitals increasingly share de-identified or anonymized patient data. Linked data is all about sharing data. Instruments, including wearable gadgets (*Washington Post* 2015) produce an avalanche of data that under the right arrangements can be shared.
– Increased capabilities of computer systems. Advances in both hardware and software allow storing and processing very large amounts of data, perhaps "in the cloud." Advances in algorithms for data analysis, logical inference over large sets of propositions, and machine learning, coupled with the availability of large knowledge bases, support a combination of statistical and knowledge-based processing.

Most big data endeavors depend on data interoperability and reuse. See Section 3.5 on merging like data sets to increase sample size and linking data sets that have data on different variables.

2.6 *Knowledge base to support information extraction from text and multimedia*

Where do all the data in the knowledge bases discussed in the previous sections, especially Section 2.5 on big data, come from? Some come from instrument data collection. Some come from large structured databases, such as the *Library of Congress Subject Headings* or a drug database or a patient database. Some come from crowdsourcing. An ever-growing volume (Maybury 2012) comes from automatic information extraction from multimedia: text, images, video, and voice recordings. Information extraction is already used heavily and its use will only increase. Information extraction both uses and feeds knowledge bases for question answering. Information extraction has two components:

1. Entity extraction (named-entity recognition)—Locating references to named entities (people, organizations, places) and to concepts in text, graphs, images, audio (e.g., locating people through face recognition or voice signature). Large KOS with many entity values, such as geographic places, many variant names (in multiple languages), and information useful for disambiguation are important for this component.
2. Relationship extraction—Identifying relationships that link entities into propositions, statements that are made in or derivable from the text and multimedia content.

Information extraction uses any combination of statistical methods, including machine learning, and syntactic / structural /semantic analysis. Either way, it needs a lot of information, much of it organized in KOSs.

– Linguistic knowledge (morphological, part-of-speech, and lexical), including knowledge about lexicalized phrases, knowledge of synonyms / variant names or labels for the same entity, and knowledge of the multiple senses of words and how to determine which sense (which entity value) applies in the context (word sense disambiguation, WSD).
– Hierarchical and other relationships among concepts and knowledge about the entity type to which a concept belongs.
– World knowledge. For example, a large database about people, their basic demographic data, their organizational affiliations, roles, subject interests, etc., helps in correctly identifying a person mentioned in a text when the name could refer to any of many different persons. Such knowledge can be found in knowledge bases (Section 2.1), including searching linked data (Section 2.2) or in the Web at large, using the documents found for question answering or information extraction on the fly.

A particularly thorny problem in information extraction is word sense disambiguation (WSD), not just for words and phrases referring to concepts but also to names of places and of people. Place names in particular are highly ambiguous; Paris can refer to any of 30 places, to people, and to other things. WSD can use semantic knowledge available in KOSs, world knowledge, and statistical association between words or phrases in text (which may reflect semantic information or contiguity in the world). For an example, consider the meaning of the word "white." When "white" modifies a noun that refers to a person, it means race/ethnicity (unless the person is frightened or about to faint); when "white" modifies any other noun, it usually refers to color. Even this rule is not sufficient to disambiguate "white dishwasher," since "dishwasher" could refer to a person or a machine, so the sense of "dishwasher" in the given context needs to be determined first. For a more complex example, consider the meaning of the phrase "white drinking fountain." In the context of a tourist guide to Rome, Italy, "white drinking fountain" refers to a fountain whose color is white. In the context of talking about a segregated society, "white drinking fountain" means a drinking fountain reserved for white people; it tells nothing about the color of the fountain. It would be very hard for a computer system to interpret this phrase correctly; it would interpret "white" as the color. Even a human reader who does not know about segregated fountains may get it wrong. (See Kaye and Aung 2013 for a brief survey of WSD.)

With the large and increasing amount of digital text and multimedia documents, especially audio, and the increasing speed of computers, statistical methods have taken a front seat in information extraction. But KO also has a place; fast computers and large knowledge bases enable complex reasoning and inference.

High-quality information extraction in complex domains, such as medicine, profits from a complete knowledge base. In turn, information extraction augments the knowledge, making it even more useful for subsequent information extraction. Information extraction can be used for the construction of KOS; see, for example, Aussenac and Soergel (2005). Even linguistic information can be learned from text and multimedia; Sonnenberger (1995), Winiwarter (2011).

2.7 *Causal maps (influence diagrams), system dynamics models, process diagrams, concept maps, and other node-link diagrams. Models and simulations*

A causal map is a node-link diagram that includes nodes for all the variables that directly or indirectly influence a phenomenon of interest and links to show the influence of one variable on another. Causal maps are very useful to represent synthesized information for quick assimilation, for playing out what-if scenarios, and ultimately for guiding research and policy (Roberts 2013) and evidence-based decisions. Follow some of the links in Figures 5-7. System dynamics modeling is a closely related area; see, for example, Borshchev and Filippov (2004).

Causal maps are closely intertwined with KOSs. They present variables, often arranged in groups to form a hierarchy. The complete shiftN map (Figure 7 shows a piece) contains 102 variables arranged in 8 groups (Figure 8). Arranging these variables in meaningful order as done in Figure 8 is a KO issue. Causal maps also need a schema of relationship types to distinguish different types of influences; these are not always made explicit.

Comparing and perhaps integrating (Graudina et al. 2012; Hao et al. 2007 and Marshall et al. 2006) several causal maps on the same topic may lead to new insights. For this we need to ascertain which relationship types (link types) and which of the variables connected are the same or can be mapped, a classic case of KOS mapping (see Section 3.4; Figure 9 shows a manually produced mapping of four lists of variables). Then we can determine which individual link instances are the same and which are different and produce a pooled map that includes all variables and all influence links from the maps being merged (horizontal integration). We can see which influence relationships occur in multiple maps and which only in one.

It is also often useful to integrate two or more causal maps dealing with different phenomena into one more comprehensive map that shows a wider range of influences (vertical integration). This requires that one can

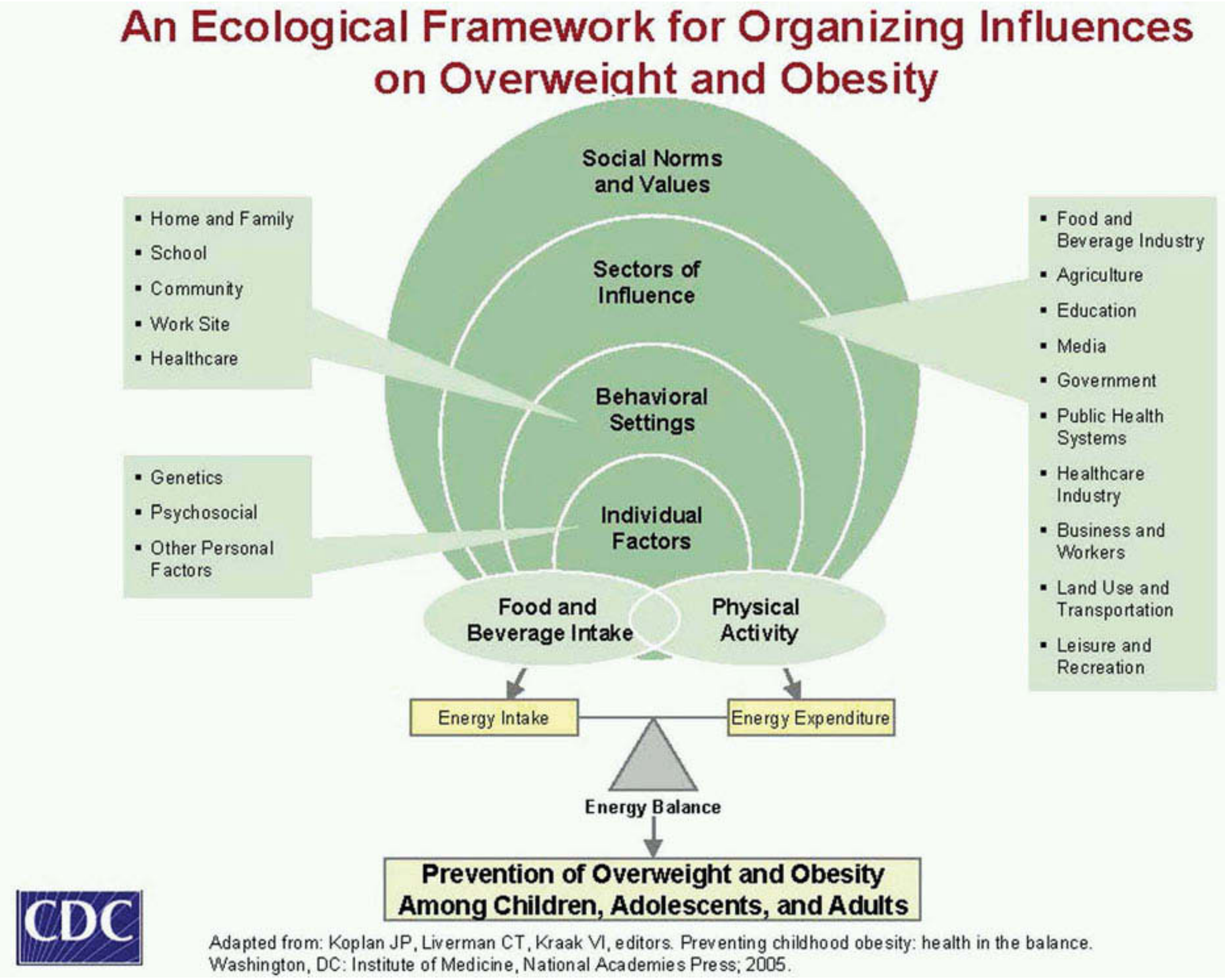


Figure 5. An overview diagram of influences on obesity (Milstein and Homer 2009).

identify one or more connecting nodes, nodes that represent the same variable in both diagrams to be connected. As an example, consider a causal map of healthy living and a causal map for environmental sustainability. There are at least two connecting points: 1) urban planning for walkability of the environment contributes both to healthy living and to lower energy use and lower pollution; and 2) the proportion of plant-based food in the diet affects both health and resources for food production; more fruits, vegetables, and grains is healthier, and plant-based foods use fewer resources with a lower carbon footprint. So what is good for your health is good for the environment, especially climate change. For research on these interactions, see www.nsf.gov/news/ "food, energy, and water."

Causal maps are a form of representing KOSs. Consider a system that lets the user explore a topic in a well-presented dynamic causal map, highlight one or more variables and/or one or more links, and receive documents discussing these variables and links. The system might also let the user type in a term for a concept or a link and see this concept in the context of a causal map.

Figures 10 and 11 show a different type of node-link diagram, diagrams of sequential and interrelated processes that lead to some outcome or state, which are important in biology and in industrial engineering. Biology specifically uses diagrams (Le Novère et al. 2009, Sander et al. 2009, Chaouiyal et al. 2013) of signaling pathways, metabolic networks, and gene regulatory networks. For standards in this area see COmputational Modeling in BIology NEtwork (COMBINE) (Le Novère et al. 2013). Soergel (2004) describes MONOD, a prototype system that assists researchers with extracting data from the literature and synthesizing them into a knowledge base, of which a graph is one representation. Each data item is linked its source(s), so the graph can serve as a search interface, as suggested above.

A more generic form of node-link diagrams, concept maps, have been used as thesaurus displays since 1950 (Soergel 1974, Section D3.2, 249, and the references given there, including the EURATOM Thesaurus) and have resurfaced forcefully in education (see Section 2.8). It would be interesting to find earlier uses.

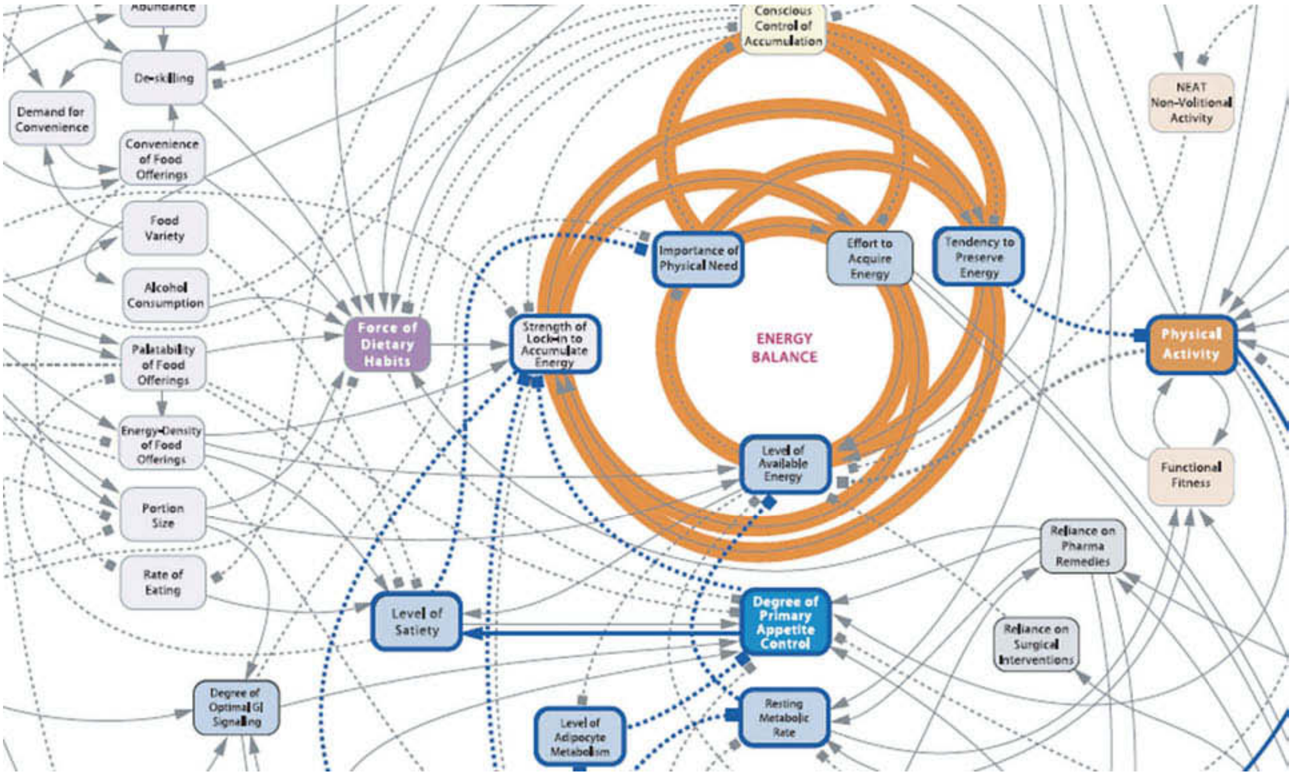


Figure 6. Segment of the large and detailed shiftN causal map for obesity
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/296290/obesity-map-full-hi-res.pdf
(Vandenbroeck et al., *Building the obesity system map* 2007; Vandenbroeck et al., *Obesity system atlas* 2007).

2.8 Knowledge organization for understanding and learning (for more see Soergel 2013 [2015])

> "[M]eaningful or deep learning can be supported through giving learners
> – well-structured presentation of material,
> – schemas they can use to organize knowledge specific topics in their minds,
> – general knowledge organization principles they can use to construct their own schemas." (Soergel 2013 [2015], modified; Zhang and Soergel 2014).

This theme is relevant not just for K-20 education and professional training, but well beyond that for any information transfer and interaction with information of any kind (including search, information display, and games) in any format (print, online, live).

The need for knowledge organization to support comprehension is well recognized in education, including instructional design, and information architecture, but implementation in the design of learning materials and other informative documents leaves much to be desired. The KO community has paid only scant attention to the use of KO in learning; that must change. Educators, instructional designers, and user interface designers need help from KO in creating meaningful presentations and in helping students to learn principles for their own knowledge organization as an essential part of thinking skills-an important role for KO professionals. It requires deep engagement with the subject matter and sufficient knowledge of cognitive psychology to understand how information must be presented and learning activities must be structured to assist different types of learners in discerning meaning and making sense. Figures 12-16 (Soergel 2013 [2015]) give examples. Figures 15 and 16 emphasize meaningful arrangement.

Causal maps, influence diagrams, system dynamics models, and similar representations of complex relationships (Section 2.7) can help students understand complex systems. On the use of simulations 1) in business courses (Papenhausen and Parayitam 2015), and 2) in helping students understand complex science systems (Nuhoğlu 2010 and Eseryel and Law 2012).

2.9 Knowledge transfer between domains

Subject domains or disciplines often share substantive or formal structures that, if recognized explicitly, can facilitate knowledge transfer between domains. A KO specialist working on a KOS spanning multiple domains is in a

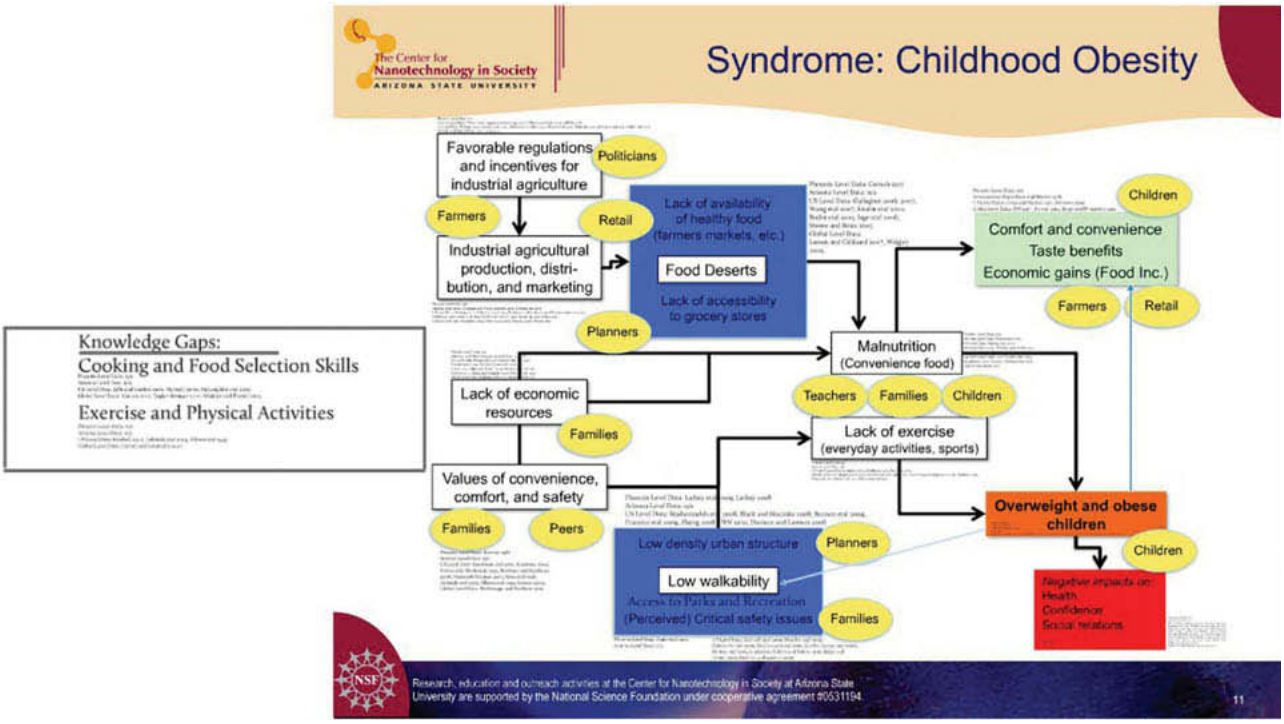


Figure 7. Childhood obesity causal map (Foley 2012; Wiek et al. 2012).

good position to discern such commonalities. For example, "management styles and educational styles" align to a large extent, so they can be unified under a heading "style of social interaction" (Figure 17). Any of the alignments are potential places for knowledge transfer, but non-alignments (a blank in one of the original columns) are especially promising. "Montessori-style" education helps children to become self-reliant and to behave properly in society based on their own volition, their own internalized norms and values. There is little or no reliance on teacher supervision or external rewards such as handing out stars for good behavior. What would a "Montessori-style organization" look like? It would hire people who are self-reliant and have inner discipline and motivation, with much less need for supervision and extrinsic motivation, such as constant praise or money incentives. Conversely, what would "holacracy" in education look like? As a management style applied to schools, holacracy would give responsibility for how to teach to the teachers.

Holacracy in the classroom means giving students more responsibility for their own learning. Rather than have the teacher lecture, task a team of students with learning about a topic, giving them materials and clear learning objectives. Or put students in charge of a course: Teams of students prepare classes, with readings, presentations, and discussion questions. Or students work on projects in independent teams, using holacracy ideas for making teamwork efficient. Unification of two classifications enables transfer of ideas between education and management.

There are a number of examples of technologies that have been inspired by observations of structures and mechanisms in living organisms. Such transfer might be further facilitated by developing a KOS of system structure and functions that captures commonalities between the structure and function of mechanical and electronic devices on the one hand and the anatomy and physiology of living organisms on the other. Breton (1991) argued for indexing documents to support transfer. These are all examples of flexible unification as discussed in Section 3.4.

3.0 Improving KO tools and techniques to meet the requirements of advanced applications

After discussing advanced applications of KO, we now turn to KO improvements that would benefit these applications.

3.1 Better data modeling

The entity-relationship (E-R) approach with its close relationship to predicate logic is arguably the most basic and general approach to data modeling. However, its application suffers from limitations; most serious the heavy preference for binary relationships, as in RDF triples and in E-R diagrams. Second is the complication of data modeling by using "attributes." Third is the unfortunate choice of calling relationship types *properties*, which just sows confusion.

Individual	Environment
Engine Energy balance Conscious control of accumulation Effort to acquire energy Importance of physical need Level of available energy Strength of lock-in to accumulate energy Tendency to preserve energy	
Physiology Degree of primary appetite control by brain Extent of digestion and absorption Genetic and/or epigenetic predisposition to obesity	
Food consumption Force of dietary habits Tendency to graze Demand for convenience Food exposure	**Food production** Societal pressure to consume Demand for health Pressure to improve access to food offerings Cost of ingredients
Individual physical activity Level of transport activity	**Physical activity environment** Dominance of motorised transport Opportunity for unmotorised transport
Individual psychology Food literacy Stress	**Social psychology** Exposure to food advertising Peer pressure

Figure 8. shiftN causal map variables. Top level with example detail (arranged by DS). Observe the two columns for individual factors and environmental factors and the juxtaposition of related factors, such as food consumption and food production.

shiftN Figure 6	**Kaplan** Figure 5	**Nano-technology** F .7	**Downey' list**
Engine			
Energy balance	Energy balance 1)		
	Energy intake1)		
	Energy expenditure1)		
Conscious control of accumulation			lack of self-control
Effort to acquire energy			
Physiology			
Degree of primary appetite control by brain			hunger response to food cues
Digestion and absorption			
Genetic and/or epigenetic predisposition to obesity			genetics epigenetic factors

Figure 9a. Variable matches between four causal maps.

3.1.1 Limitation 1. Restricting the modeling space to binary (two-way) relationships

In much E-R diagraming and in RDF triples (the present RDF-based version of linked data), relationships are limited to binary relationships, connecting two entities. However, the world is not made up of triples. Reality and thought are much more complex than that; adequate representation requires multi-way relationships. The designers of frame-based systems understood that. While it is possible to represent multi-way relationships as a structure consisting of several binary relationships and issue SPARQL queries that consider such structures, it is not a natural way of representation and therefore difficult and not widely used.

Food consumption	Food and beverage intake		Overeating
Force of dietary habits			
		Malnutrition (convenience foods)	high fructose corn syrup
		Lack of econ. resources	
Food production	Food and beverage industry	Agricultural production	agricultural policies
Demand for health			
		Food deserts	food deserts
Cost of ingredients			
Individual physical activity	Physical activity	Exercise and physical activities	
		Lack of exercise	low level of phys. Activity
Level of transport activity			
Physical activity environment	Land use and transportation		transportation policies
Opportunity for unmotorised transport		Low walkability	
Safety of unmotorised transport		Perceived safety issue	perception of neighborhood safety
Individual psychology	Psychosocial		
Food literacy		Cooking and food selection skills	lack of nutritional education
Stress			Stress
			smoking cessation
Social psychology			
Peer pressure			
	Behavioral settings		

Figure 9b. Some (approximate) matches and non-matches between variables from four causal maps.

[1]) Note: The three basic variables "Energy balance," "Energy intake," and "Energy expenditure" are implied by all four models but explicitly mentioned only as shown. For still more variables in this domain, see French et al. (2012) and Food4Me Project (2015). Note the many empty cells-even the most comprehensive of the four maps (Col. 1) is missing factors. Merging the four maps would produce a more complete causal map. Some of the matches are tenuous. Where Col. 1 has general variables, Col. 4 often has a specific value ("lack of, low level of"). Col. 1 captures, for example, the influence of control on food intake, which could be good or bad, while Col. 4 looks only at factors that influence obesity, so it looks only at the influence of lack of control on food intake. Or Col. 3 listing "malnutrition (convenience foods)" while Col. 4 gives just one example, "high fructose corn syrup." To solve the match problem, researchers and modelers in the area would need to agree on a common set of variables and methods to measure these variables.

The following two examples show that binary relationships cannot adequately model real-life data as needed in the knowledge bases discussed in Section 2.1):

Example 1. Drug treatment

Not adequate: disease <"treatedBy"> drug

Reality, and therefore the information needed by a physician, is much more complex; it requires modeling dependencies between several entity types (or variables). Following the notation used in mathematics for multivariate functions, we can write:

Realistic: disease<"treatedBy"> (drug, dosage, effectiveness, age, gender, ethnicity)

Example 2. Impact of events

Faithfully recording information on any type of impact of an event requires he following complex relationship type:

Event <"hadImpact"> (TargetType, ImpactType, ImpactSeverity, TimeSpan, CountNumber)

where

TargetType could be a person, a building, or a crop, or …

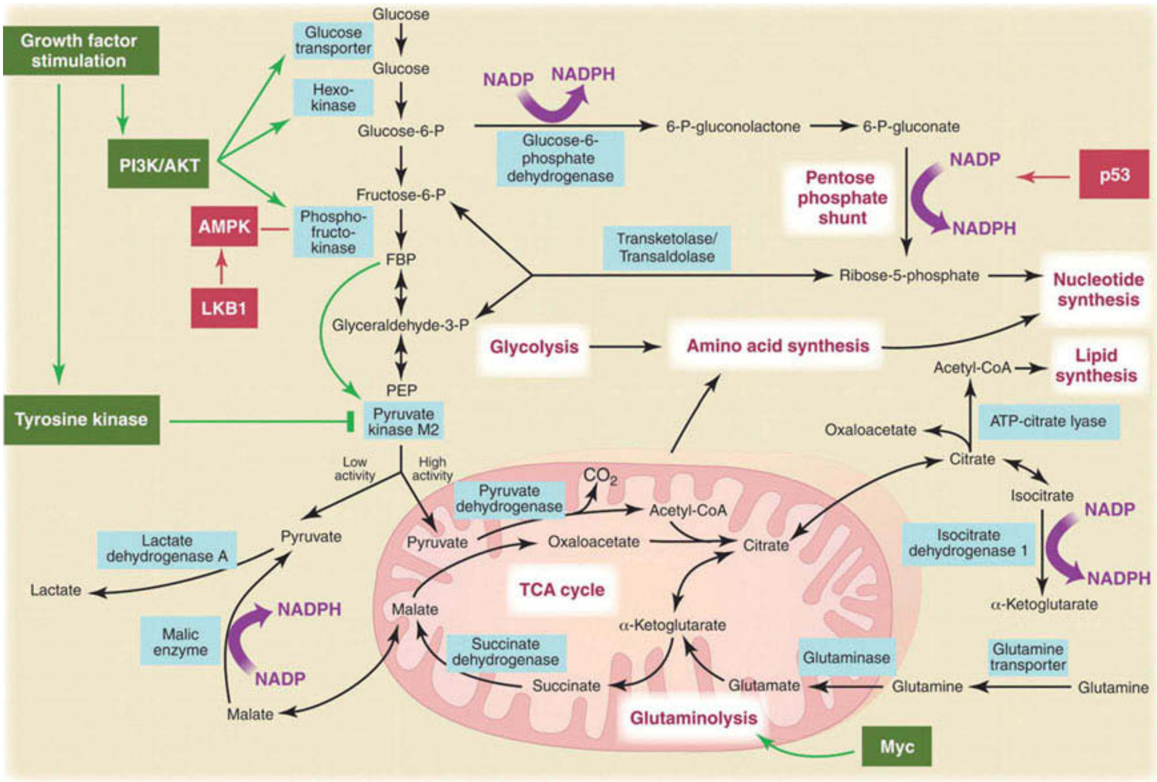


Figure 10. Cell metabolism diagram (retrieved from http://nutritionpaperideas.com/cell-metabolism-diagram/ 2015-05-10).

1420 *C. Stötzel et al. / Theriogenology 78 (2012) 1415–1428*

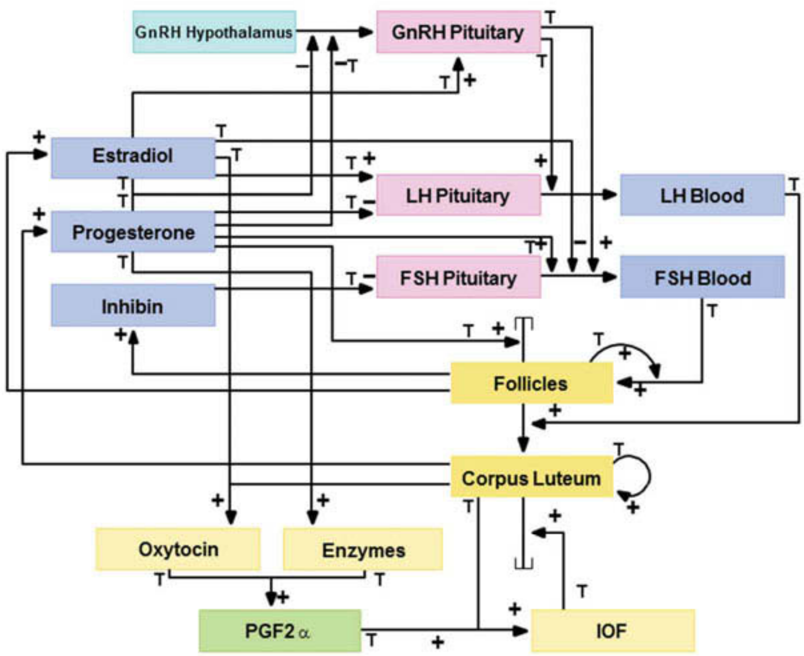


Fig. 4. Schematic representation of the components and relations in the herein presented advanced model of the bovine estrous cycle. Each box represents a substance for which a differential equation is derived, the color resp. grayscale of a box indicates the different physiological compartments where the substance mainly occurs: hypothalamus (GnRH), pituitary (GnRH, LH, FSH), blood (LH, FSH, estradiol, progesterone, inhibin), ovaries (follicles, corpus luteum, oxytocin, enzymes, IOF), uterus ($PGF_{2\alpha}$). The arrows represent the regulations between the substances in the model, where "+" marks a stimulatory effect, "T" denotes a threshold within a Hill function. No description means a transition, and 'Λ ϑ ' marks a degraded substance.

Figure 11. A process regulation diagram (Stötzel 2012).

User topic 'Food in Auschwitz' for illustration
Segments of Holocaust survivor interviews relate to the topic in different ways:
Direct relevance: Direct evidence for what the user asks for
Example: A Holocaust survivor talks about food available to Auschwitz inmates.
Indirect relevance: From which one can infer something about the topic
Example: A Holocaust survivor talks about seeing emaciated people in Auschwitz
Context relevance: Provides background/context for topic
Example: A Holocaust survivor talks about physical labor of Auschwitz inmates
Comparison relevance: Provides information on a similar or contrasting situation
Example: A Holocaust survivor talks about food in the Warsaw ghetto

Hypothesis 2

Students who understand these relevance relationships prepare richer descriptions and analyses of historical events using more and more types of information

Elaboration of the comparison relationship as applied to events

(1) Varying time or place of event
(1.1) Compare with similar event at a different time
(1.2) Compare with similar event at a different place

(2) Varying the participant(s) in the event
(2.1) Compare with similar event that has a different actor
(2.2) Compare with similar event that has a different person or object being acted upon

(3) Varying the act / experience
(3.1) Compare with a different act carried out by a similar actor
(3.2) Compare with a different experience experienced by a similar person or object

Hypothesis

Students who understand the relevance relationships related to comparison can think of more events that can be compared with an event in focus and prepare a richer report comparing and contrasting the event in focus with other events.

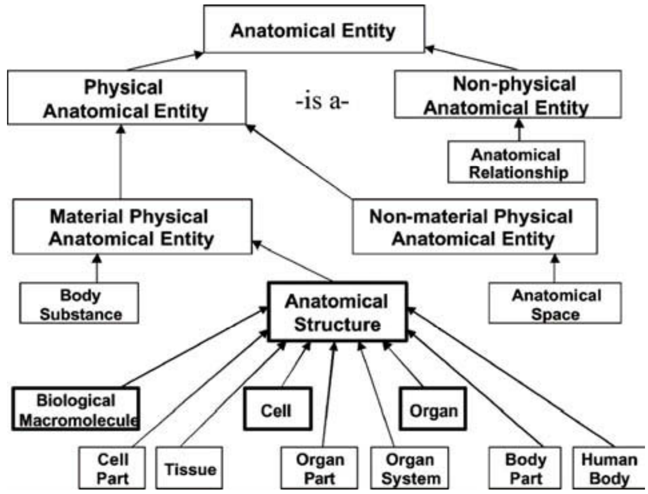
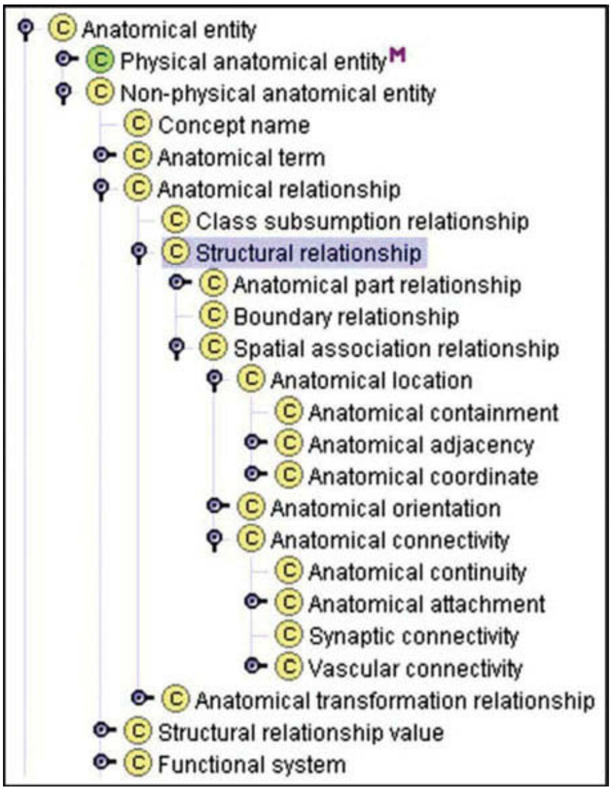
Figure 12. Relevance relationships that help people think (Huang and Soergel 2006).

ImpactType could be wounded, killed, damaged, destroyed, …
ImpactSeverity would be given on a scale, say 1-5

Following are brief observations on how multi-way relationships would benefit some of the other applications discussed in Section 2:

– "Linked data" (Section 2.2) needs to evolve beyond triples to be able to store more complex data from many sources in a way that enables integration of data from many sources in complex queries and reasoning. We should not let present, but transient, technological limitations stand in the way of proper conceptual modeling.
– "In information extraction" (Section 2.6) there is work to move beyond extracting just binary relationships to extracting complex relationships. See, for example, McDonald et al. (2005) and, applied to events, Döhling and Leser 2014.
– "Causal modeling" (Section 2.7) focuses on binary influence relationships, especially if the models are presented graphically. However, influence is often more complex. Independent variables may interact; there are statistical methods to find such interactions. Such interactions are represented by three-way or higher-order relationships as C = f(A, B). Advanced modeling and simulation techniques allow for representing such patterns of interaction through formulas. However, a node-link graph cannot represent such multi-way dependencies (an observation that also applies to E-R diagrams).
– "For learning" (Section 2.8), one needs to be able to express complex relationships that students or policy makers need to understand.

To sum up, the world is not made of triples. To be fully useful, representations must be able to capture complex relationships to allow for more accurate knowledge bases for question answering and cognitive computing, for more adequate representation of information extracted from text, for more expressive causal models, etc. Software for processing these representations—query, reason over, merge—must follow suit.

Entity types	Relationship types
	

Schematic representation of the principal classes of the Anatomy Taxonomy. (Rosse & Mejino, 2003, Fig. 4, p. 486)

(From Rosse & Mejino, 2003, Fig. 8)

Hypothesis

Students who are taught anatomy using the *Foundational Model of Anatomy* have a better grasp of the structure of the body.

Figure 13. Foundational Model of Anatomy (an entity-relationship schema).

Hypotheses

The concept map representation will allow some learners to form a better internal representation of a bird as a system.

Constructing a concept map will lead most learners to a deeper understanding of a topic

Note: Students need to think about the entities and relationships involved and need to make these explicit to themselves. Students need to first learn (at their level) about E-R modeling and concept map construction as mental tools.

Figure 14. Concept map about birds (from Cañas and Novak 2006/2008).

3.1.2 Limitation 2. Attributes as elements in entity-relationship (E-R) modeling

The proposal for E-R modeling presented by Peter Chen (1976), drawing on previous work (Williams 2012), is ingenious, except for introducing "attributes." Attributes are not needed as modeling elements. Using only entities and relationships is simpler and avoids many problems that arise from distinguishing between relationships and attributes.

Consider "birthday," which is often modeled as an attribute of person; but

person <"bornOn"> date or
person <"associatedDate"> (date, DateRelationshipType)
with DateRelationshipType = BirthDate

models this piece of data perfectly and has the advantage of making explicit the entity type date. Also, developers tend to store a piece of information they call attribute differently from what they call a relationship, possibly creating problems for processing. One might argue that "date" is not a "primary" entity in its own right, but con-

a. Britannica Elementary: Menu for *Animal Kingdom*

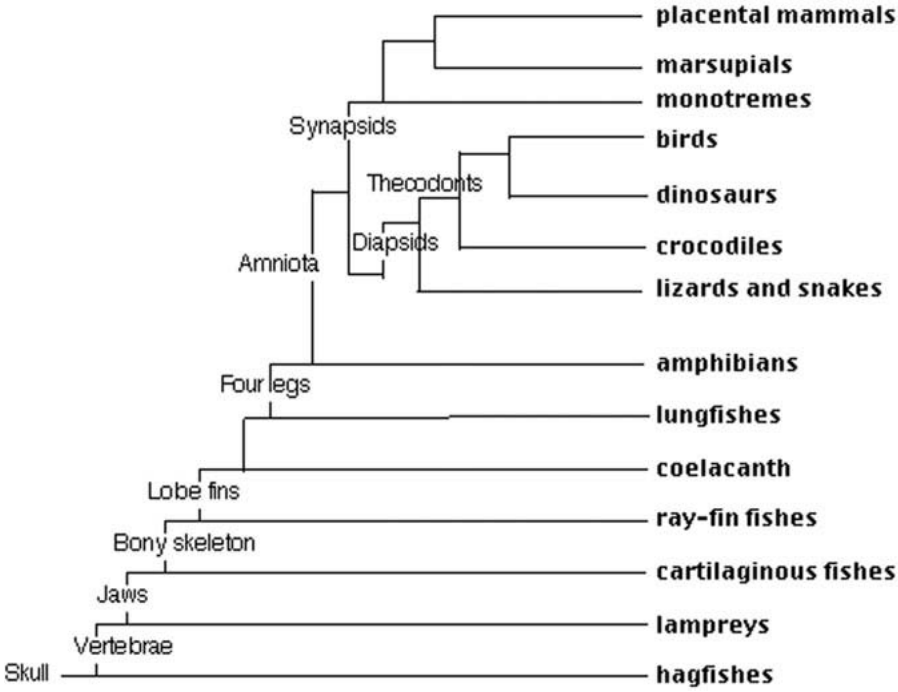
b. Meaningful arrangement *Animal Kingdom*

Invertebrates (animals without a spine**)**
 Mollusks (snails, octopus, mussels)
 Arthropods (bugs (insects), spiders, crabs)

Vertebrates (animals with a spine**)**
 Fish
 Amphibians frogs, toads, salamanders)
 Diapsids (lizards&snakes, crocodiles, dinosaurs, birds)
  Reptiles (lizards&snakes, crocodiles, dinosaurs)
  Birds
 Mammals (elephants, whales, cows, dogs, bats, mice, monkeys)

Note: Would need to be made visually more attractive for children.

c. Vertebrates cladogram (basis for b.)



http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/V/Vertebrates.html

a Encyclopedia Britannica. Thoughtless arrangement, devoid of any meaning
b Meaningful arrangement. Based on the modern cladogram shown in **c**.,.

Hypothesis

Young students who use the animal home page with the meaningful arrangement will over time absorb the sequence and perceive a progression. When much later in biology (perhaps in 10th grade) the structure of the animal kingdom and the evolution of animals are discussed, these students will understand more quickly.

For use of science categories to design Web directories, see Bilal and Wang 2005

Figure 15. Classification of animals.

XF **body system or organ** XG . musculoskeletal system XH . skin system XJ . cardiovascular system XK . respiratory system XL . mouth, larynx, vocal organ XM . digestive system XN . urogenital system XP . . urinary system XQ . . reproductive system XR . blood, immune system XS . . blood XT . . immune system XU . endocrine system XV . sensory system XW . nervous system XY . . peripheral nervous system XZ . . central nervous system XZ8 . . . brain	Arranged by increasing complexity and integrative function
Source: Alcohol and Other Drug Thesaurus https://web.archive.org/web/20111015041143/http://etoh.niaaa.nih.gov/AODVol1/aodthome.htm	
Hypothesis A learner presented with this meaningful arrangement of body system according to a germane principle (as opposed to an alphabetical arrangement) will gain a better understanding of the function of each body system and how they work together.	

Figure 16. Meaningful arrangement of body systems

Style of social interaction Result of unification	**Management style** Input to unification	**Educational style** Input to unification
Autocratic, authoritarian, directive	Autocratic, authoritarian, directive (coercive), top-down	Direct instruction, teacher-centered Teacher as formal authority, expert
Military style	Military style	Military style
Paternalistic	Paternalistic	
Authoritative (visionary)	Authoritative (visionary)	
Persuasive	Persuasive	
Coaching	Coaching	Teacher as facilitator
Individual inner discipline, motivation, agreement with norms		Montessori
Participatory, democratic	Participatory (democratic), consultative	Democratic and Free Schools
Collaborative, teamwork	Collaborative, teamwork	Cooperative Learning Teacher as facilitator, delegator
Self-directed groups	Holacracy, self-management in groups	
Laissez-faire, free-wheeling	Laissez-faire	Open Schools (and Classrooms) (Summerhill)
Chaotic	Chaotic	
People try their own thing		Inquiry-based learning, student-centered Teacher as facilitator, delegator

Figure 17. Management styles and educational styles compared. Management styles from http://en.wikipedia.org/wiki/Management_styles and http://www.slideshare.net/Meimeibarcoma/management-style-models?related=1. Educational styles from http://www.ratical.org/many_worlds/PoL.html and http://teach.com/what/teachers-teach/teaching-methods (see also Dobish 2003 and Todorović et al. 2012).

sider the *Timelines of World History* in Wikipedia and the many books with similar titles; events are arranged by date-"Date" is the primary entity.

3.1.3 *Limitation 3. Calling relationships properties, as is done in RDF*

This unfortunate choice of term, rooted in the origins of RDF, has led to many misunderstandings. In particular, it obscures the bi-directional nature of relationships. It also fixes in people's minds the limitation to binary relationships (see Section 3.1.1). Terminology does matter. Language influences the way we think. So this is a plea, especially to the Semantic Web community, to simplify terminology and use one term, "relationship" (or relationship type, to be precise) rather than confusing readers by using a second term, "property," to express the same idea.

3.1.4 *Implications of improved data modeling*

Following the principles outlined in Sections 3.1.1-3.1.3, one can construct an entity-relationship conceptual data schema (an ontology) that is simple and clear and easily understood even by non-technical people, yet represents information in the requisite complexity, laying the ground for improvements in the applications discussed in Section 2.

3.2 *Refined relationships in KOSs. Integration of KOS data with other data*

Traditional thesauri limit conceptual relationships to two: hierarchical (BT/NT) and associative (RT). One can argue through examples that more differentiated relationships would allow for better retrieval. There has long been a call to refine relationships in thesauri and other KOSs. For example, in hierarchical relationships one should distinguish "generic" ("isa" and "isSubclassOf"), "part-whole," and "topic inclusion." In associative relationships one might distinguish "similarInMeaning" and "servesPurpose," to give just two of many examples. The Getty Art and Architecture Thesaurus (AAT), FAO's AGROVOC, and GO, among others, include refined relationships.

Using refined relationships in KOSs is closely related to a broader issue: what is the boundary between 1) KOSs and 2) empirically-based data (propositions) about worlds, real and imaginary? On theoretical grounds one can argue a distinction between 1) definitions and relationships between concepts that follow from definitions and 2) statements about a world. In this view, definitions are contrasted with the results of empirical investigations. Definitions are neither true nor false (they may be more or less useful). In practice, all kinds of data are used together in showing information or answering questions. Concept maps often represent much factual data. Factual data can be useful for query expansion in bibliographic retrieval. For example, one may want to find documents on the biology of all insects that are pests of rice. So one can first find all such insects (using real-world knowledge) and then OR all the insect names to find documents. So KOS data should be stored simply with other data in one large (virtual) database where queries can freely use any and all data. This is existing practice in linked data on the Web, since all linked data may be considered to form one large knowledge base. The Getty vocabularies (AAT, ULAN, TGN, and CONA) include among them much real-world data.

Sections 3.3 and 3.4 on unification and Section 3.5 on integrating KOS and other bodies of data, merging and linking, all look at the same phenomenon from different perspectives.

3.3 *Formal unification of KOSs and other bodies of data*

There are many types of KOSs (Soergel 2009), known as ontologies, thesauri, classifications, dictionaries, etc. that give widely overlapping information but also differ in the kinds of information they emphasize. These different types of KOSs have different standards, different formats for storing their data, different software tools, and different display formats. This is not good. There should be one comprehensive standard that accommodates all types of data stored in any KOS and concomitant comprehensive software tools. The Organization Authority Database discussed in Section 2.3.2 is a case in point; in its abstract structure, it is just like a hierarchically structured thesaurus; it (Soergel and Popescu 2015) uses the same data structure and the same display formats and hierarchical navigation, all powered by the same software. In sum, KOSs (and possibly other bodies of data) should use the same principles of organization, internal structure, and display for systems with different kinds of entities.

3.4 *Flexible semantic unification of KOSs and other bodies of data*

For all applications discussed in Section 2, the principle of "unification" is very important. It is the key to unleashing the power of data through integrating bodies of data. Unification means using the same entity types and relationship types, defining concepts the same way (or clearly note differences), using the same terms (or at least clearly mapped terms) for concepts, using the same URI to identify an organization anywhere.

3.4.0 Flexible unification. General principles

Flexible unification of two or more KOSs or other bodies of data in the same or interacting areas does not require total merger and loss of identity. It can be achieved by analyzing the systems, finding commonalities and agreements where they exist, creating agreement where it is possible, and explaining differences that must be maintained while still establishing relationships of partial agreement. A single KOS database can and should be hospitable to many viewpoints. See Section 3.4.2 for examples.

Unification involves crossing boundaries; it makes boundaries more permeable to promote exchange and interoperability. Unification is often useful:

- across applications or databases, across information systems in an organization;
- across types of data (example: organization database treated like classification);
- across disciplines, supporting knowledge transfer between disciplines (Section 2.9);
- across languages (precise definitions);
- across cultures, across organizations (organizational cultures); and,
- across worldviews (important for negotiating the politics of creating a new KOS or of unifying existing KOSs when stakeholders have different views of the world, requiring KO expertise to create a system that reflects commonalities but has room for differences).

One may need to establish formal KOS structures before attempting unification, for example, when relating the kinship classification and terminology from two cultures.

At one end of the unification spectrum we can indeed establish a single scheme to be used across many applications:

- at the schema level, MARC and *Resource Description and Access* (*RDA*) for library catalogs or an entity-relationship schema or ontology for financial transactions;
- widely used library classifications (LC*C, DDC,* UDC) and subject headings (*LCSH, MeSH*);
- the same classification of medical procedures for insurance billing;
- a system of identifiers for geographic places used across the Web (with possible extensions within specific organizations); and,
- the same database of organizations with a URI assigned to each (Section 2.3.2).

At the other end we leave the source schemes the same and simply establish correspondences.

Unification involves identifying, in the two or more sources, elements that are the same or similar at the concept level. In first approximation this mapping can be done based on terms or names. Automated KOS mapping is an active research area (Euzenat and Shvaiko 2013 and OM 2015); using available large KOS as resources can help.

3.4.1 Unifying entity types and relationship types. The dream of a unified ontology

In the database world, this is known as schema integration or schema mapping. We start with an obvious case. Users of linked data now must contend with a hodgepodge of entity types, relations ship types, and data standards (Section 2.2). A comprehensive registry of entity types and relationship types would make searching and processing linked data ever so much easier.

But linked data are just the tip of the iceberg. Node-link diagrams and computer models use entity types and typed links (links that indicate the relationship type involved) (see Section 2.7). Any of the information systems discussed in Section 2 need a conceptual data schema. So it would be really useful to have, if not a unified ontology, an inventory of entity types and relationship types as defined and used in many systems with their definitions and with relationships of sameness, similarity, and subsumption. There are individual attempts at creating collections of entity types and relationship types, for example:

- schema.org, sponsored by major players, including Microsoft, Google, and Yahoo. It collects entity types and relationships from quasi-standards (see Section 2.2 for examples). But schema.org is not well organized; it needs help. It is high time that more KO expertise be brought to bear on improving schemas of entity types and relationship types, especially those used with linked data;
- http://www.heppnetz.de/projects/goodrelations/ ;
- any of the many upper level ontologies, such as:
- BFO Basic Formal Ontology (http://www.ontobee.org/browser/index.php?o=BFO for entity types; http://www.ontobee.org/browser/index.php?o=RO for relationship types),
- DOLCE,
- SUMO.

But there is not the overarching neutral collection place many in the KOS community have long wished for. Such a more systematic (and appropriately funded) effort would greatly ease the finding and processing of information (aiding effectiveness and reducing costs) but is unlikely to materialize. Furthermore, most of these systems focus on binary relationships with the attendant limitations (see Section 3.1.1)

3.4.2 *Flexible unification at the entity value level*

There are many more entity values than there are entity types (over 10 million values for the entity type "place" alone), so problems of unification are massive. For example, every place should be identified by a single URI that is used everywhere, even in tags embedded in newspaper text. Flexibility is needed even in this seemingly simple domain: place boundaries change over time; United States government agencies divide the U.S. territory into administrative regions most of which are different from one agency to another.

For sensible unification of cultural concepts, flexibility is a must. For example, consider adapting (not just translating) the English-language Art and Architecture Thesaurus (AAT) to Chinese art (Chen, in press). We must add many artistic techniques specific to Chinese art. Then we must reorganize the hierarchy of artistic techniques to accommodate the new techniques; some will fit neatly into the existing hierarchy, others can be grouped with existing techniques under a new broader concept, and still others require a whole new branch in the hierarchy. We should also introduce appropriate related term cross-references when a new technique and an existing technique are similar but not the same. Each technique-Western and Chinese-needs a good definition in English and in Chinese. The resulting bilingual thesaurus will be useful for indexing Western art and Chinese art appropriately, but beyond that it will help Westerners to understand Chinese art and Chinese people to understand Western art. Quite similar considerations apply to developing a classification of anatomical structures in different organisms. For another example of unification, see Section 2.7, Figure 9. This type of unification requires a team of people who have considerable conceptual skill in the application of classification principles and good knowledge of the areas to be unified.

In some cases unification is exceedingly difficult and requires great intellectual effort. That should not keep us from pursuing unification where it is conceptually (if not necessarily politically) easy, as in assigning unique identifiers to places or organizations (while allowing for multiple names). And if we succeed in the difficult cases, we will do a great service to users and perhaps even foster mutual understanding across cultures, religions, or ideological divides.

The use of language may be broad and depend on the cultural context, even in translation. For example, African students in the United States may refer to their landlady as their "mother" (I thank Barbara Kwa ś nik for this example), implying that in their culture there is a broad concept of mother that includes, but is not limited to, what is normally referred to in English as "mother" (which itself is a radial category in Lakoff's terms (1987, 83), in that it could refer to birth mother, genetic mother, surrogate mother, caretaking mother, foster mother, stepmother). Such cases fall in the "exceedingly difficult" category. In the case at hand, the solution is to introduce the very broad concept of "mother in a extended sense" above "mother" as understood in Western culture, which in turn includes many narrower terms for specific types of mother.

Unification should focus on concepts not terminology. It should establish agreement between concepts or explain differences between concepts, however slight. Agreement on terminology is nice, even important in some cases, but not even desirable in others (consider expert and lay medical terminology-to each his own).

3.5 *Integrating bodies of data-KOS, knowledge bases, data sets, and causal maps and other node-link diagrams for stronger evidence and enhanced discovery*

Integrating (combining, merging, linking) bodies of data allows for more complete answers, inference considering more conditions, and data analyses not otherwise possible. It depends crucially on the unification of KOSs. This section explores, mostly through examples, issues in integrating data, interoperability, and compatibility. The discussion could even be extended to texts (a text can be considered a body of data), to partnerships and integration of companies and government agencies, and to interoperability in computer systems (both hardware and software). There may be some general underlying principles, but they manifest themselves slightly differently in different environments.

Unification of KOS plays a key role

3.5.1 *Integrating knowledge organization systems (KOSs)*

We can distinguish three cases:

1. Horizontal integration of KOSs that cover the same domains and give the same kind of data for concepts and terms to produce a KOS that is more complete both in terms of the concepts and terms in the domain and in terms of the information (for example, synonyms, related concepts) given. Such a KOS could also be used as a unified KOS by multiple information systems, increasing reusability of data.
2. Horizontal integration of KOSs that cover different domains but give the same kind of data for concepts and terms to produce a KOS with wider domain coverage.

3. Vertical integration of KOSs that cover the same domain but give different kinds of data for concepts and terms to produce a richer KOS that can be used for multiple purposes (see also Section 3.3).

3.5.2 Integrating data sets for research

We can distinguish three ways of integrating data sets for research.

3.5.2.1 Horizontal integration of data sets with similar data for the same units of analysis to increase confidence that the data are correct.

This is a case of unification.

3.5.2.2 Horizontal integration of data sets with similar data for different units of analysis to increase coverage

Number of units of analysis; for example, assume we want to study the following topic:

School success in the United States differentiated into a number of specific concept understandings and skills (the "content" of learning) dependent on:
- the methods of learning and instruction used and the degree of freedom afforded to teachers;
- student characteristics, such as socio-economic status, race/ethnicity, and IQ; and,
- school characteristics, such as location and its characteristics, socio-economic and racial/ethnic mix in the classroom, and funding per student.

To do this, we integrate data sets with (anonymized) data about individual students from all states in the U.S. to create a very large sample, perhaps 20 million students, that allows for detecting effects in many specific situations characterized by a specific combination of independent variables and for studying interaction of variables in fine-grained detail.

The data include the independent variables mentioned, student answers to all questions on their state's standardized tests (which reflect the state's educational standards or learning objectives), and student grades in specific school subjects. To integrate the data sets from the different states into one large sample for analysis, we need to align the independent and the dependent variables, a classic case of KOS unification. Whether alignment is possible depends on the degree of agreement in defining and measuring the independent variables and the dependent variables, in this example the degree of uniformity of educational standards across states so that at least some of our specific school success variables are supported by some test question in every state test.

To sum up, this example illustrates integrating data sets that contain similar data on different instances or units of analysis (different students, patients, schools, programs,) to increase sample size. This approach has been used for a long time in meta- analyses. It requires that the variables of interest are defined the same or at least similar ways and that data values are represented the same way or at least mappable onto the system to be used for the study. For example, measurements provided by different wearable gadgets must be comparable.

3.5.2.3 Vertical integration of data sets with different data for the same units of analysis to increase variety

Number of variables; for example, assume we want to study the following topic:

relationships between
- per capita income,
- how people feel about the economy, and
- birth rate,

using place (city or town, county, or perhaps zip code) as the unit of analysis.

We have three large data sets:

data set 1: per capita income by place,
data set 2: Twitter messages (used to extract sentiment and place of the sender),
data set 3 birth rate by place.

We need a data set that gives for each place all three variables, so we need to integrate the three data sets. Our ability to do this depends on how places are defined and identified in each data set. Particularly in the Twitter data set, this is a big problem. Aligning the places in the three data sets requires 1) a large data base of places that gives names, nicknames, coordinates, and other information and 2) considerable KO expertise.

To sum up, this example illustrates the approach of vertical integration of data sets that contain different kinds of data on the same instances of units of analysis, so that a larger range of variables can be studied. Here (Dolbear and Hart 2008, Lyons et al. 2009) it is critical that in all sets to be integrated the units of analysis (places, persons, schools) are identified in such a way that sameness can be determined, so that data from different data sets about the same place or the same person or the same school can be properly combined.

3.5.3 Integrating knowledge bases for question-answering for cognitive computing

Horizontal integration here means merging two such knowledge bases with the same type of propositions to increase coverage or check agreement. Determining whether two propositions are the same, contradict each other, or are just different is not simple; this is especially important for propositions generated by information extraction from text and multimedia. The knowledge base should still include contradictory propositions, but link them with a relationship <“contradicts”> and provide sources/evidence for each. Vertical integration means linking two such knowledge bases with different types of propositions to enable answering different types of questions, including questions that require inference over multiple types of propositions. The Google knowledge graph is fed from multiple sources, using both horizontal and vertical integration.

3.5.4 Integrating node-link diagrams (causal maps, concept maps)

Section 2.7 has one example each for horizontal and vertical integration of causal maps.

3.5.5 Concluding remarks

Integration enables more powerful analysis using existing data. Often both horizontal and vertical integration of multiple bodies of data is required; this certainly applies to using linked data. Integration depends on the ability to match entity types, relationship types, and entity values used in the bodies of data to be integrated; here is a crucial contribution from KO. As a special case, this condition determines whether linked data contributed from different sources can truly be used together.

4.0 Conclusions

Knowledge organization has many and varied applications. Many of these applications use KOSs that are not informed by proper expertise and therefore poor. To benefit society and brighten its own future, the KO community should make itself useful and improve KOSs used in these areas. In improving existing KOSs and developing new ones, we should consider both:

- requirements for using KOSs for machine processing, specifically inference; and,
- requirements for using KOSs for human processing, specifically meaningful arrangements that assist in making sense as discussed in Section 2.8, especially Figures 16 and 17.

Many existing KOSs, especially those constructed by professionals trained in computer science or informatics, consider only use by machines, but the structures captured in them would be useful for people, too (see Section 2.8 on understanding and learning).

There are many opportunities for people with good training in KOSs, but they must be prepared to work in today's and tomorrow's environments-science labs, data repositories, consortia that work on KOS for their communities, instructional design labs, and more. Some examples:

- Help plan for the “data side” of scientific (including social science) studies, making sure variable definitions and measurement and data collection methods follow applicable standards (or explain the difference) and curate the resulting data so they can be shared.
- Participate in the development of causal maps and other node-link diagrams from the beginning to make the results interoperable and reusable.
- Participate in interdisciplinary studies to promote knowledge transfer among disciplines.
- Work with researchers to support reuse and integration of existing data, causal maps, computer models, etc.
- Work with instructional and user interface designers to create meaningful presentations.

Preparing KO professionals for such tasks requires an educational program that goes beyond what is offered in many information programs (including programs in library and information studies and in information schools) to foster new competencies and skills, specifically:

- A basic understanding of logic, formal ontology principles, inference, and complex queries so they can contribute to formal ontologies and incorporate pieces from systems that come from the ontology community into KOSs for more general use.
- Skills in using KOS software and in dealing with linked data.
- Ability to discern meaningful structures and then convey structure and meaning through good document.
- Ability to align classifications from different disciplines to support knowledge transfer.
- Ability to work with researchers on defining variables, determining measurement an data collection methods, and curating and sharing data, all to improve interoperability and reusability, as required increasingly by funders.

There are many communities that are involved in knowledge organization, but there is not enough communication across the boundaries. Among these communities are:

– Knowledge organization (as represented, for example, in ISKO);
– Semantics in linguistics and terminology;
– Knowledge representation in artificial intelligence;
– Ontology;
– Data modeling;
– Semantic web.

We need to improve communication across these communities through inviting speakers, having joint meetings, perhaps even merging ISKO with one or more other societies.

This paper points to opportunities for the field of knowledge organization to advance by improving and building upon our existing tools and practices. Development in these areas could enable KO, and the professionals who practise it, to occupy a more prominent and fulfilling role in diverse applications. For a final conclusion see Figure 18.

The Future of Knowledge Organization

Knowledge organization is needed everywhere

Create the future of KO

Think BIG. Think answers not pointers.
Focus on substantive data

Many areas, tasks, and functions that could profit from KO principles

Engage with Ontologies, AI, data modeling

Figure 18. The future of knowledge organization

References

Alexander, Keith, Richard Cyganiak, Michael Hausenblas, and Jun Zhao. 2009. "Describing Linked Data sets: On the Design and Usage of void, the 'Vocabulary of Interlinked Data sets.'" In *Linked Data on the Web Workshop (LDOW 09)*. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B6464C112CF51F71FEFF1768E17CBE59?doi=10.1.1.178.2369&rep=rep1&type=pdf.

Auer, Soren, Jens Lehmann, Axel-Cyrille Ngonga Ngomo and Amrapali Zaveri. 2013. "Introduction to Linked Data and Its Lifecycle on the Web." In *Reasoning Web. Semantic Technologies for Intelligent Data Access: Proceedings of 9th International Summer School 2013, July 30-August 2, 2013, Mannheim, Germany,* edited by S. Pudolph, G. Gottlob, H. Ian and F. van Harmelen. Berlin; New York: Springer. doi: 10.1007/978-3-642-39784-4.

Baierer, Konstantin, Evelyn Dröge, Vivien Petras and Violeta Trkulja. 2014 "Linked Data Mapping Cultures: An Evaluation of Metadata Usage and Distribution in a Linked Data Environment." In *Proceeding of International Conference on Dublin Core and Metadata Applications 2014, 8-11 October 2014, Austin, Texas.*

Bilal, Dania and Peiling Wang. 2005. "Children's Conceptual Structures of Science Categories and the Design of Web Directories." *Journal of the American Society for Information & Technology* 56: 1303-13.

Borshchev, Andrei and Alexei Filippov. 2004. "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools." *The 22nd International Conference of the System Dynamics Society, July 25 – 29, Oxford, England.* http://www.anylogic.com/upload/iblock/296/296cb8fa43251a1fc803e435b93c8ff2.pdf.

Breton, Ernest J. 1991. "Indexing for Invention." *Journal of the American Society for Information Science* 42, no. 3: 173-7.

Brickley, Dan and Libby Miller. nd. "FOAF (2000-2015)." http://www.foaf-project.org/.

Cañas, Alberto J. and Joseph D. Novak. 2006/2008. *The Theory Underlying Concept Maps and How to Construct Them.* FL: Institute for Human and Machine Cognition. http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf.

Ceusters, Werner, Maria Capolupo, Barry Smith and Georges De Moor. 2009. "An Evolutionary Approach to the Representation of Adverse Events." *Studies in Health Technology and Informatics* 150: 537-41.

Chaouiyal, Claudine, Duncan Bérenguier, Sarah M. Keating, Aurélien Naldi, Martijn P. van Iersel, Nicolas Rodriguez, Andreas Dräger, Finja Büchel, Thomas Cokelaer, Bryan Kowal, Benjamin Wicks, Emanuel Gonçalves, Julien Dorier, Michel Page, Pedro T. Monteiro, Axel von Kamp, Ioannis Xenarios, Hidde de Jong, Michael Hucka, Steffen Klamt, Denis Thieffry, Nicolas Le Novère, Julio Saez-Rodriguez and Tomáš Helikar. 2013. "SBML Qualitative Models: A Model Representation Format and Infrastructure to Foster Interactions between Qualitative Modelling Formalisms and Tools." *BMC Systems Biology* 7: 135. http://link.springer.com/article/10.1186%2F1752-0509-7-135.

Chen, Peter Pin-Shan. 1976. "The Entity-Relationship Model—Toward a Unified View of Data." *ACM Transactions on Database Systems* 1: 9-36. http://www.inf.unibz.it/~nutt/IDBs1011/IDBPapers/chen-ER-TODS-76.pdf.

Chen, Shu-Jiun, Marcia Lei Zeng and Hsueh-Hua Chen. (in press). "Alignment of Conceptual Structures in Controlled Vocabularies in the Domain of Chinese Art: A Discussion of Issues and Patterns." *International Journal on Digital Libraries.*

Davies, Tim and Duncan Edwards. 2012. "Emerging Implications of Open and Linked Data for Knowledge Sharing in Development." *IDS Bulletin* 43, no. 5: 117-27. doi: 10.1111/j.1759-5436.2012.00372.x.

Dobish, Elizabeth 2013. *Management Styles & What Makes an Effective Teacher.* http://www2.mcdaniel.edu/slm/student/dobishe/present580.ppt.

Döhling, Lars and Ulf Leser. 2014. "Extracting and Aggregating Temporal Events from Text." In *WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web*, 7-11 April, 2014, Seoul, Korea. Geneva, Switzerland: ACM, 839-44. https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/research/publications/2014/tempweb2014.pdf.

Dolbear, Catherine and Glen Hart. 2007. "Ontological Bridge Building – Using Ontologies to Merge Spatial Data sets." In *AAAI Spring Symposium: Semantic Scientific Knowledge Integration*, 15-20. https://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-05/SS08-05-004.pdf.

DuCharme, Bob 2011. "Introduction to SKOS." http://www.snee.com/skos/20111006/.

Dunsire, Gordon, Corey Harper, Diane Hillmann and Jon Phipps. 2012. "Linked Data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability." *Information Standards Quarterly* 24, no. 2/3: 4-13. http://www.niso.org/apps/group_public/download.php/9411/FE_Dunsire-etal_VocabMgmt_isqv24no2-3.pdf.

Eseryel, Deniz and Victor Law. 2012. "Effect of Cognitive Regulation in Understanding Complex Science Systems During during Simulation-Based Inquiry Learning." *Technology, Instruction, Cognition, & Learning* 9: 111-32.

Euzenat, Jérôme and Pavel Shvaiko. 2013. *Ontology Matching*. Heidelberg: Springer.

Ferucci, D. A. 2012. "Introduction to 'This is Watson'." *IBM Journal of Research and Development* 56, no. 3/4: 1-15. http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6177717.

Foley, Rider. 2012. *Obesity – Causal Diagram: Integrating Logic And and Research To to Display Complex Systems.* https://prezi.com/rsn0kbwpjg0g/copy-of-copy-of-obesity-causal-diagram/.

Food4Me Project. 2015. *Personalised Nutrition: Paving a Way to Better Population Health.* http://www.food4me.org/images/White_Paper/food4me_white_paper-7-04-15.pdf .

French, Simone A., Leonard H. Epstein, Robert W. Jeffery, John E. Blundell and Jane Wardle. 2012. "Eating Behavior Dimensions: Associations with Energy Intake and Body Weight: A Review." *Appetite* 59: 541-9. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3454469/.

Fridsma, Doug. 2013. *EHR Interoperability: The Structured Data Capture Initiative.* http://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/ehr-interoperability-structured-data-capture-initiative/.

Graudina, Vita, Janis Grundspenkis, Sigita Milasevica and Riga Technical University. 2012. "Ontology Merging in the Context of Concept Maps." *Applied Computer Systems* 13: 29-36.

Hao, Jin-Xing, Ron Chi-Wai Kwok and Angela Yan Yu. 2007. "Automatic Semantic Causal Map Integration." In *2007 Proceedings of the Pacific Asia Conference on* Information *Systems (PACIS).* http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1092&context=pacis2007.

Hu, Jingyuan, Hongming Cai, Boyi Xu, and Cheng Xie. 2014. "A Linked Data Based Decision Support System for Cancer Treatment." In *Proceeding of the Enterprise Systems Conference (ES), 2-3 Aug. 2014, Shanghai*, China. *Enterprise Systems Conference (ES)*, IEEE, 39-44.

Huang, Xiaoli and Dagobert Soergel. 2006. "An Evidence Perspective on Topical Relevance Types and its Implications for Exploratory and Task-based Retrieval." *Information Research* 12: 281-91.

Kaye, K. and Win Thandar Aung. 2013. "Word Sense Disambiguation: A Briefly Survey." *International Journal of Computer & Communication Engineering Research (IJCCER)* 1: 1118-23

Lakoff, George. 1987. *Women, Fire, and Dangerous Things.* Chicago, IL: University of Chicago Press.

Le Novère, Nicolas, et al. 2009. "The Systems Biology Graphical Notation." *Nature Biotechnology* 27: 735-41.

Le Novère, Nicolas. 2013. *COMBINE: The Computational Modeling in Biology Network.* http://co.mbine.org/system/files/COMBINE-LeNovere.pdf. See also the website http:/co.mbine.org.

Lyons, Ronan A., Kerina H. Jones, Gareth John, Caroline J. Brooks, Jean-Philippe Verplancke, David V. Ford, Ginevra Brown and Ken Leake. 2009. "The SAIL Databank: Linking Multiple Health and Social Care Data sets." *BMC Medical Informatics and Decision Making* 9, no. 3. http://www.biomedcentral.com/1472-6947/9/3.

Marshall, Byron B., Hsinchun Chen, Rao Shen and Edward A. Fox. 2006. "Moving Digital Libraries into the Student Learning Space: The GetSmart Experience." *Journal on Educational Resources in Computing* 6, no.1: np.

Maybury, Mark T. 2012. *Multimedia Information Extraction: Advances in Video, Audio, and Imagery Analysis for Search,* *Data Mining, Surveillance and Authoring*. Hoboken, NJ.: John Wiley & Sons, Inc.

McDonald, Ryan, Fernando Pereira, Seth Kulick, Scott Winters, Yang Jin and Pete White. 2005. "Simple Algorithms for Complex Relation Extraction with Applications to Biomedical IE." In *ACL '05 Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 25-30 June 2005, Ann Arbor, MI*. Stroudsburg, PA: ACM, 491-8. http://www.ryanmcd.com/papers/relationACL2005.pdf.

Milstein, Bobby and Jack Homer. 2009. *System Dynamics Simulation in Support of Obesity Prevention Decision-Making*. https://www.iom.edu/~/media/Files/Activity%20Files/PublicHealth/ObesFramework/IOMIrvine16Mar09v52MilsteinHomer.pdf.

Moreno-Conde, Alberto, David Moner, Wellington Dimas da Cruz, Marcelo R. Santos, José Alberto Maldonado, Montserrat Robles and Dipak Kalra. 2015. "Clinical Information Modeling Processes for Semantic Interoperability of Electronic Health Records: Systematic Review and Inductive Analysis." *Journal of the American Medical Informatics Association* 22: 925-34.

Nuhoğlu, Hasret. 2010. "The Effect of the System Dynamics Approach on Understanding Causal Relationship Skills in Science Education." *Procedia—Social and Behavioral Sciences* 2: 3614-18.

OM. 2015. *Tenth International Workshop on Ontology Matching*. Bethlehem, PA:2015-10-12. http://om2015.ontologymatching.org/.

Papenhausen, Chris and Satyanarayana Parayitam. 2015. "Business Management Simulations as a Pedagogical Tool to Improve Student's System Dynamics Thinking." *Developments in Business Simulation and Experiential Learning* 42: 76. https://journals.tdl.org/absel/index.php/absel/article/view/2913.

Peroni, Silvio. 2014. "The Semantic Publishing and Referencing Ontologies." In *Semantic Web Technologies and Legal Scholarly Publishing,* 121-94. London: Springer.

Roberts, Mark S. 2013. *Best Practices in Applying Models to Healthcare Policy*. http://hssr.duke-nus.edu.sg/sites/duke-nus.edu.sg.hssr/files/files/Mark%20Roberts%20pdf.pdf.

Sander, Herbert Sauro, Jacky L. Snoep, Kurt Kohn and Hiroaki Kitano. 2009. "The Systems Biology Graphical Notation." *Nature Biotechnology* 27: 735-41.

Smith, Barry et al. nd. *Why are Ontologies Needed to Achieve EHR Interoperability?* http://www.slideserve.com/kylee/why-are-ontologies-needed-to-achieve-ehr-interoperability.

Soergel, Dagobert, and Denisa Popescu. 2015. "Organization Authority Database Design with Classification Principles." *UDC Seminar 2015 October*. http://seminar.udcc.org/2015/abstracts.php.

Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles, CA.: Melville.

Soergel, Dagobert. 2009. *Knowledge Organization Systems. Overview.* www.dsoergel.com/SoergelKOSOverview.pdf.

Soergel, Dagobert. 2011. *A General Model for Searching Linked Data OR Design of an Integrated Information Structure Interface. A Unified Framework for Indexing and Searching in Database, Expert, Information Retrieval, and* Hypermedia *Systems*. http://www.dsoergel.com/UBLIS571DS-04.1-1Reading2SoergelSearchingLinkedData.pdf. A much revised and expanded version of Soergel, Dagobert. 1994. "Challenges in Indexing Electronic Text and Images." In *Information Structure Management. A Unified Framework for Indexing and Searching in Database, Expert, Information-Retrieval, and Hypermedia Systems*. Medford, NJ: Learned Information, 111-56.

Soergel, Dagobert. 2013 [2015]. "Knowledge Organization for Learning. Conjectures and Methods of Study." *Ciência da Informação. Special issue Informação e Tecnologia* 42, no. 2: 232-54 http://revista.ibict.br/index.php/ciinf/article/download/2288/1911. A revised and much expanded version of Soergel, Dagobert. 2014. "Knowledge Organization for Learning." In *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proceedings of the Thirteenth International ISKO Conference, 19 -22 May 2014, Krak*ów, *Poland*, edited by Wiesław Babik. Advances in knowledge organization 14. Würzburg: Ergon Verlag, 22-32.

Soergel, David, Kirindi Choi, Ty Thomson, Jay Doane, Brian George, Ross Morgan-Liniall, Roger Brent and Drew Endy. 2004. *MONOD: A Collaborative Tool for Manipulating Biological Knowledge Management System*. http://download.communicationx.net/MONOD-Biological-Knowledge-Management-System-Soergel-et-download-w81930.html.

Sonnenberger, Gabriele 1995. *Der Wit-interpreter: Ein Textanalysesystem, das seine eigene Wissensbasis durch die Analyse von Texten erweitert*. [*The wit-interpreter. A text analysis system that adds to its own knowledge base during text analysis.*] Konstanz: Hartung-Gorre Verlag.

Starr, Barbara. 2014. *Demystifying the Google Knowledge Graph*. http://searchengineland.com/demystifying-knowledge-graph-201976.

Stötzel, C., J. Plöntzke, W. Heuwieser and S. Röblitz. 2012. "Advances in Modeling of the Bovine Estrous Cycle: Synchronization with PGF2." *Theriogenology* 78: 1415–28.

Todorović, Jelisaveta, Marina Matejević and Ivana Simić. 2012. "Educational Styles, Communication and Contentment of Students with their Families." *Procedia—Social and Behavioral Sciences* 69: 899-906.

Vandenbroeck, Philippe, Jo Goossens and Marshall Clemens. 2007a. *Tackling Obesities: Future Choices—Building the Obesity System Map*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295154/07-1179-obesity-building-system-map.pdf.

Vandenbroeck, Philippe, Jo Goossens and Marshall Clemens. 2007b. *Tackling Obesities: Future Choices – Obesity System Atlas*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295153/07-1177-obesity-system-atlas.pdf.

W3C (World Wide Web Consortium). 2012. *SKOS Simple Knowledge Organization System. Home page*. http://www.w3.org/2004/02/skos/.

W3C (World Wide Web Consortium). 2014. *The Organization Ontology*. http://www.w3.org/TR/vocab-org/.

Washington Post. 2015. *Wearable Gadgets Portend Vast Health, Research, and Privacy Consequences*. http://cacm.acm.org/news/186878-wearable-gadgets-portend-vast-health-research-and-privacy-consequences/fulltext.

Wiek, Arnim, Rider W. Foley and David H. Guston. 2012. "Nanotechnology for Sustainability: What does Nanotechnology Offer to Address Complex Sustainability Problems?" *Journal of Nanoparticle Research* 14, no. 9: 1-20.

Williams, Paul. 2012. *A Short History of the ER Diagram and Information Modeling*. http://www.dataversity.net/a-short-history-of-the-er-diagram-and-information-modeling/.

Winiwarter, Werner 2011. "Automatic Linguistic Knowledge Acquisition for the Web." *International Journal of Web Information Systems* 7, no. 1: 18-43.

Zhang, Pengyi and Dagobert Soergel. 2014. "Towards a Comprehensive Model of the Cognitive Process and Mechanisms of Individual Sensemaking." *Journal of the Association for Information Science and Technology* 65: 1733-56.