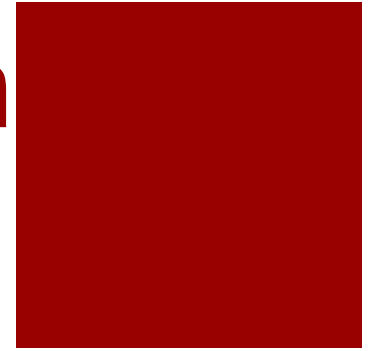


Is epistemological purism possible in Library and Information Science research?



Fidelia Ibekwe-SanJuan

ELICO – University of Lyon3 (France)

{fidelia.ibekwe-sanjuan@univ-lyon3.fr }

Thomas M. Dousa

Graduate School of Library and Information Science

University of Illinois, Urbana-Champaign, USA

{tdousa2@illinois.edu}

Motivation



- LIS is mostly driven by practical applications (Bates 2005, 2007)
- Need to ground the field in epistemologically-motivated methods (Bawden & Robinson 2012; Brookes 1980; Buckland 2012; Cornelius 2002)
- Incorporate epistemology and philosophy into LIS curricula (Furner 2010; Hjørland 1998, 2010)

What's at stake?



- “Epistemology is at the basis of the question “*what is the scientific method?*” (Hjørland 2010)
- Different scientific methods are coupled, more or less closely, with different epistemological positions.
- But there is need to draw a distinction between epistemological views and methods based on them (Hjørland 1998, 163).

Some major epistemological positions according to Hjørland



- Empiricism

knowledge acquisition through observations and experimentation, by induction

- Rationalism

knowledge gained through reasoning, an innate faculty

- Positivism

knowledge gained through analytico-deductive reasoning

- Historicism/Hermeneutics

knowledge acquisition through consideration of history, contexts, culture, geography and subjectivity

- Pragmatism

knowledge acquisition on the basis of satisfying goals and purposes of a person or a community

(Hjørland 2010)

Is epistemological purism possible?



- How can we answer these calls and make explicit the epistemological assumptions underlying a given piece of research work which we have carried out?
- Is one piece of research likely to embody one epistemological assumption from a given epistemological view, at various stages of its completion or various?
- Boundaries between epistemological theories can be fuzzy (e.g., in Philosophy, is Kant's epistemology empiricist and rationalist?).

Is epistemological purism possible?



- With respect to KO (Hjørland 2003, 2011):
 - Facet analysis is seen as a “rationalist” approach/stemming from rationalism
 - Some automatic indexing techniques are classified as “empiricist”/ stemming from empiricism
 - But in reality/practice, a single classification or indexing method may embody assumptions from more than one epistemological theory.

Is epistemological purism possible?



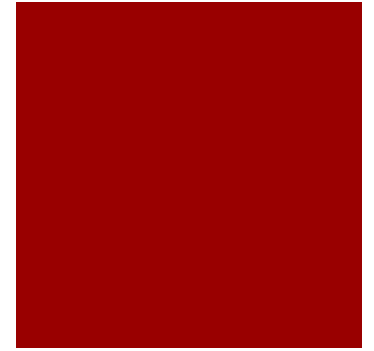
- For instance:
 - **S. I. Ranganathan's** Colon Classification appears to be grounded in a rationalist epistemology, in practice, he combined rationalism and pragmatism or “pragmatic-rationalism” in the actual realization of his classification (Tennis 2008).
 - **Julius Otto Kaiser's** method of Systematic Indexing: “his epistemological and methodological positions were hybrid in nature. Kaiser was primarily empiricist and pragmatist in theory whereas his methodology was pragmatist in aim but rationalist in mechanics”. (Dousa 2008).

Is epistemological purism possible?



- Hjørland is commendably aware of this: classification of epistemological theories into separate categories are “idealizations” which usually do not exist in pure forms.
- He observed that the four major epistemological views are interconnected: for “any kind of pragmatism is limited by constraints set by the real world through empirical evidence” (Hjørland 2003: 107)
- pragmatism is “closely related to historicism by understanding that observations are contextual” (Hjørland 2009: 1526), even if pragmatism places more emphasis on purpose.

Is epistemological purism possible?



- methodological hybridity underlies what would seem to be epistemologically “monolithic” approaches or schemes

Is epistemological purism possible?



- So in essence: LIS and KO theorists and practitioners may espouse one or more epistemological views as the basis of their research in theory (their discourse), but depart from it/them in practice (i.e., when building the knowledge artifact) ...
- How can we demonstrate/illustrate this better?

Our Approach



- Take 1 – 2 examples from the fields in LIS (i.e., a case study approach)
- Unpack the stages involved and uncover the epistemological assumptions operative at each stage as well as the limitations/biases of these assumptions

Fields for Our Case Study



- Information Retrieval (IR)

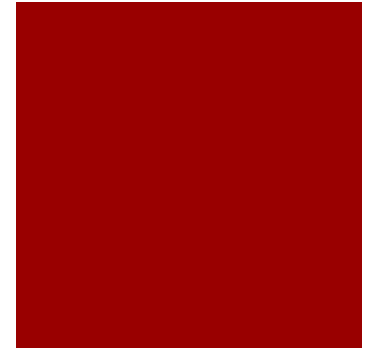
- Topic Mapping

- * Chen C., Ibekwe-SanJuan F., SanJuan E., & Vogeley M.S (2008)

- * SanJuan & Ibekwe-SanJuan (2006)

- * Ibekwe-SanJuan & SanJuan (2010)

Illustration



- IR (Information Retrieval)
- Topic Mapping

Combination of empiricism + positivism ...
with a dash of pragmatism + hermeneutics

Steps 1 & 2: Data collection & feature selection



■ From empiricist assumptions...

- Frequency is a topic marker
- Frequently co-occurring units reflect similarity of content and hence of documents;
- statistical behaviours of textual units capture what a document or a query is about.

■ ...To empirical models

- statistical or probabilistic models are built to approximate these observations
- Induction process: rules or models built from the observations of a particular set of documents
- not universal: may not hold in all cases

Steps 1 & 2: Data collection & feature selection

- Example: Astronomy domain (Chen *et al.* 2008),
- Corpus of 1,293 bibliographic records of SDSS-related publications retrieved from the ISI-WoS database. Sloan Digital Sky Survey (SDSS) <http://www.sdss.org/>

$$G(t) = \frac{F(t) \times V(t)}{\max\{F(t) : t \in T\} \times \max\{V(t) : t \in T\}}$$

$$F(t) = \sqrt{\sum_{d \in D} f_{t,d}^2 \times \ln(|\{d : t \in d\}|)}$$

$$V(t) = \sqrt{\sum_{d \in D} (\text{match } d \text{ against } t)^2}$$

Steps 1 & 2: Data collection & feature selection

1	G_Mean	tf.idf	Match_SQL_t	Freq	Docs	term
	0,55	0,95	0,57	78	43	white dwarf
	0,38	0,59	0,65	49	39	dark matter halo
	0,36	1	0,36	89	64	luminosity function
	0,32	0,72	0,45	57	38	velocity dispersion
	0,32	0,81	0,39	74	58	star formation
	0,3	0,63	0,47	59	53	active galactic nucleus
	0,28	0,77	0,36	62	40	early-type galaxy
	0,28	0,59	0,47	50	42	large-scale structure
	0,27	0,43	0,63	37	35	cosmic microwave background
	0,26	0,52	0,5	39	28	star formation history
	0,26	0,61	0,42	45	29	power spectrum
	0,26	0,62	0,41	56	49	emission line
	0,25	0,74	0,34	59	38	stellar mass
	0,25	0,39	0,63	27	18	black hole mass
	0,24	0,7	0,34	60	46	spectral type
	0,24	0,5	0,48	42	36	dark matter
	0,24	0,43	0,55	30	19	black hole
	0,23	0,47	0,49	37	30	gravitational lensing
	0,21	0,32	0,65	23	18	supermassive black hole
	0,21	0,62	0,33	48	33	photometric redshifts

Table 1. Top 20 terms ranked by $G(t)$ function on the SDSS corpus (Chen *et al.* 2008).

Step 3: Implementation phase



■ From empiricism to positivism

- Deduction of general rules which translate previous observations into formal models on the basis of which an algorithm could be written to perform the task at hand.
- The algorithm → rendered as a mathematical or logical statement and then translated into a computer program via a programming language
- Tacit presupposition that there is an ideal way in which the task at hand can be modeled in a mathematical language.
- Reflects positivist assumptions

Step 4: Representation phase



- **Sliding into hermeneutics and pragmatism**
- Interpreting knowledge artifacts
 - Lists
 - Graphs
 - the global form or layout,
 - the relative positions of nodes with regard to one another,
 - their sizes, color, shape, and links to one another.
 - Requires pragmatics (goals, purposes) and hermeneutics (background knowledge; subjectivity; cultural, historical, and literary context)

Step 4: Representation phase



- **Sliding into hermeneutics and pragmatism**
- Case of IR
- concept of relevance in IR is slippery (Buckland 2012)
- cannot be measured by intrinsically
- can only be judged by the user
- Relevance judgments are contextual and situational, can change over time even with regard to the same documents (Bawden & Robinson, 2012)

Why should we care?



- **Research design in general**

- Epistemological theories come with own ontological commitments, i.e., assumptions about what counts as an entity for the purposes of a given piece of research and what can be said about it (Quine 1948, Hjørland 2013).
- Epistemological theories determine choice of methods ... and the interpretation of results gained from those methods.

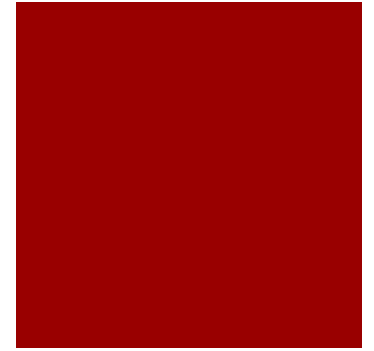
Why should we care?



■ KOS design in particular

- Hjørland (2013: 1) “the field of knowledge organization itself is based on different approaches and traditions such as user-based and cognitive views, facet-analytical views, numeric taxonomic approaches, bibliometrics and domain-analytic approaches: these approaches and traditions are again connected to epistemological views, which have to be considered.”
- each epistemological view entails ontological commitments that determine what concepts are chosen to represent the subject content of documents and how these concepts are related to one another
- It is important to know which epistemological criteria were used to construct a knowledge organization system (KOS)

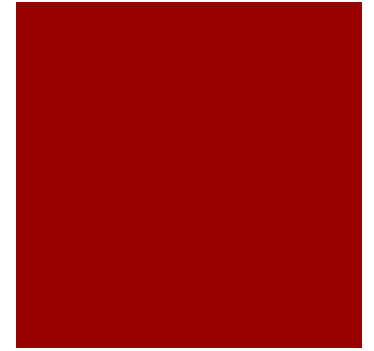
Why should we care?



■ KOS design in particular

- Design decisions about what concepts to include within a KOS, what linguistic labels to use to refer to them, and what kinds of relationships to posit between them, are made at a given time and in light of a given purpose and within the horizon of a given theory or system.
- All KOSs have a pragmatic and a historical dimension that determine both their form and content.

Why should we care?



■ KOS design in particular

- Yet, traditional admonitions to follow the “usage” of the users in the choice of terms (Cutter, 1904), select vocabulary from the literature, as a basis for facet classification (Vickery 1960) or to seek “literary warrant” in the definition of classes point to **empiricist** dimensions as well while the use of categories points to **rationalist** ones.
- Recognition of the “**different epistemological moments**” in KOS design affords better understanding of the process of creating a KOS.

Why should we care?



■ Implications for academics & info professionals

- Clarify aspects of phenomena that are favored or brought to the forefront and which are kept out of account or relegated to the background in the course of R & D.
- simulation/modelling/design are reduction processes, i.e., simplifications of otherwise complex and multifaceted realities.
- Hence, practitioners should be aware of the various factors that have contributed to the particular *Gestalten* of the KOSs that they design or apply.

Why should we care?



- Enable stakeholders, policy makers, users, and, on a broader level, members of the general public to evaluate the results of a given piece of research with due knowledge of the presuppositions that shaped its design
- Loss in users' confidence and trust if biases and limitations of research and practical results are not made explicit



One may well wonder whether, ultimately, pragmatics and hermeneutics are not the overall driving forces, conforming knowledge artifacts, to paraphrase Kant, to our thought.

- Thank you!

- Questions?

References

Bates M.J. (2005). Information and knowledge: an evolutionary framework for information science, *Information Science Research*, n° 10(4), Online at <http://InformationR.net/ir/10-4/paper239.html>.

Bates M.J. (2007). Defining the information disciplines in Encyclopedia development, *Information Science Research*, 12(4). Online at <http://informationr.net/ir/12-4/colis/colis29.html>

Bawden D., & Robinson L. (2012). *Introduction to Information Science*, Facet Publishing, London.

Brookes B.C., (1980). The foundations of information science: part I. Philosophical aspects, *Journal of Information Science*, n° 2 (3): 125-133.



References

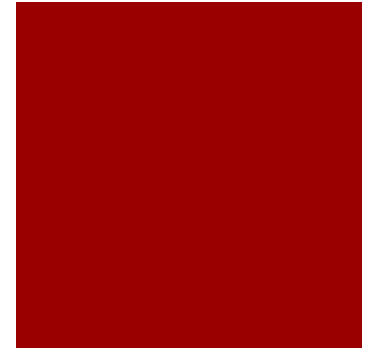
Buckland M. (2012). What kind of science can Information Science be? *Journal of the American Society of Information Science & Technology*, n° 63(1): 1-7.

Chen C., Ibekwe-SanJuan F., SanJuan E., & Vogeley M.S. (2008). Identifying Thematic Variations in SDSS Research, *Proceedings of the 9th International Conference on Textual Data Statistical Analysis (JADT-08)*, Lyon, France, 12-14 March 2008, 319-330.

Cornelius I. (2002). Theorizing Information for Information Science, *Annual Review of Information Science and Technology*, n° 36: 393-425.



References



Dousa T. M. (2008). Empirical observation, rational structures, and pragmatist aims: epistemology and method in Julius Otto Kaiser's theory of systematic indexing. In C. Arsenault & J. T. Tennis (eds.), *Culture and Identity in Knowledge Organization: Proceedings of the Tenth International ISKO Conference, 5–8 August 2008, Montréal, Canada*. Advances in Knowledge Organization, vol. 11. Würzburg: Ergon Verlag,, 242–248.

Furner J. (2010). Philosophy and Information Studies, *Annual Review of Information Science & Technology*, n° 44: 159-200.

Hjørland B. (1998). Theory and metatheory of information science: a new interpretation, *Journal of Documentation*, n° 54 (5) : 606-621.

References

Hjørland B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2): 87-111.

Hjørland B. Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8), 2009: 1519-1536.

Hjørland B., Epistemological paradigms in knowledge organization, *11th International Conference of the International Society for Knowledge Organization (ISKO)*, Rome, 23-26 February, 2010. [unpublished seminar paper].

Hjørland B. (2013). Theories of Knowledge Organization — Theories of Knowledge, 13th Meeting of the German chapter of ISKO (International Society for Knowledge Organization), Potsdam, 19th March 2013, 20 pages. [Keynote paper]. Online at <http://www.isko-de.org/data/uploads/KOAndTheoriesOfKnowlede.pdf>

References



Ibekwe-SanJuan F., & SanJuan E., Knowledge Organization Research in the last two decades: 1988-2008. (2010). In C. Gnoli & F. Mazzocchi (eds.), *Paradigms and conceptual systems in KO: Proceedings of the 11th International Conference of the International Society for Knowledge Organization (ISKO), 23-26 February 2010, Rome, Italy*. Advances in Knowledge Organization, Vol. 12. Würzburg: Ergon Verlag, 115-121.

Quine, W. V. O. (1948). On what there is. *Review of Metaphysics*, 2(5): 21–28.

SanJuan E., & Ibekwe-SanJuan F. (2006). Textmining without document context, *Information Processing & Management* 42(6), 2006: 1532-1552.

References

SanJuan E., Flavier N., Ibekwe-SanJuan F., Bellot P. (2008). Universities of Avignon & Lyon3 at TREC 2008 Enterprise Track, *17th TREC Conference Proceedings*, NIST Special Publication SP 500-277, Gaithersburg, Maryland, 18-21 Nov. 2008, 6p. Online at http://trec.nist.gov/pubs/trec17/t17_proceedings.html

Tennis J. T. (2008). Epistemology, theory, and methodology in knowledge organization, *Knowledge Organization*, n° 35 2/3: 102-112.