Knowledge Organisation Systems as Enablers to the Conduct of Science

Patrick Lambe

Knowledge Organisation & Science

"To organise knowledge is to gather together what we know into a comprehensive organised structure, to show its parts and their relationships. This is the work of scholars and encyclopaedists. It is not the role of the information profession. Our tasks are to make knowledge (whether organised or unorganised) available to those who seek it, to store it in an accessible way, and to provide tools and procedures that make it easier for people to find what they seek in those stores."

'On Knowledge Organisation' 2008

Photograph of Brian Vickery removed for copyright reasons

Two Assumptions

1. Separation between scholars and information professionals
2. Knowledge organisation mainly supports search and discovery

Photograph of Brian Vickery removed for copyright reasons

Pursuing Science

Three essentials for the conduct of science:

- Consistent ways of describing science
- Conspectus across scientific domains
- Continuity between science memory, science activity, and new knowledge creation

"The ability to achieve innovation in a competitive global information society hinges on the capability to swiftly and reliably find, understand, share, and apply complex information from widely distributed sources for discovery, progress, and productivity." (Interagency Working Group on Digital Data, 2009).

KOS & Conduct of Science

KOS:

- Standardises language
- Structure shows connections and relationships
- Enables sensemaking

Knowledge Organisation in Science

Linnaeus

- Rules for lexical stabilization enabled coordination
- Meaningful structure prefigured evolutionary theory

Portrait of Linneeus removed for copyright reasons

Mendeleev

- Sensemaking structure - periodic table prompted hypotheses about gaps and relationships between elements, supported new discoveries

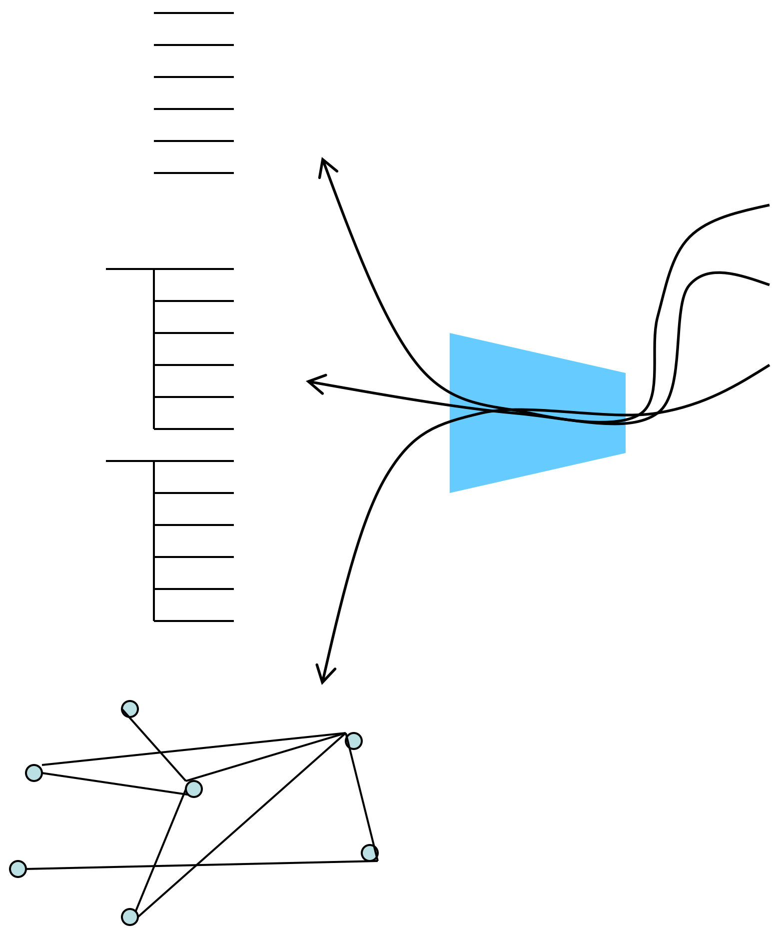
Portrait of Mendeleev removed for copyright reasons

Orders of Complexity in KOS

1. Controlled vocabularies

2. Taxonomies

3. Ontologies



4. Mechanisms for collecting emerging language

Eg. folksonomies, topic maps

Five Principles

1. The complexity of a KOS needs to match the complexity of the domain it attempts to describe, and the complexity of the coordination, connection and sensemaking work it needs to support.

2. When the complexity of the KOS exceeds human cognitive capabilities, designed interfaces using taxonomies are necessary to serve the working needs of users in their own normal working contexts.

3. It is not sufficient to use semantic technology to describe science activity. This does not get at all the functions of a KOS. Linnaeus and Mendeleev had the impact they had, because they engaged in a work of design, not simply description.

4. A KOS that effectively supports the conduct of science must be able to identify, observe and represent informal social activity and relationships beyond the boundaries of traditional formal outputs and records of science activity

5. A KOS that effectively supports the conduct of science must be able to observe and connect formal and informal activity streams, using designed taxonomy structures and visualization tools as 'human-oriented middleware' between emerging new language and concept usage, and existing ontologies.

Case Study: STAR METRICS

Straits Knowledge
Redefining corporate knowledge

Linking researcher profiles with products of their research and administrative records data on science funding requires a KOS-like data infrastructure

nature Vol 464|25 March 2010

OPINION

Let's make science metrics more scientific

To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

Measuring and assessing academic performance is now a fact of scientific life. Decisions ranging from tenure to the ranking and funding of universities depend on metrics. Yet current systems of measurement are inadequate. Widely used metrics, from the newly-fashionable Hirsch index to the 50-year-old citation index, are of limited use[1]. Their well-known flaws include favouring older researchers, capturing few aspects of scientists' jobs and lumping together verified and discredited science. Many funding agencies use these metrics to evaluate institutional performance, compounding the problems[2]. Existing metrics do not capture the full range of activities that support and transmit scientific ideas, which can be as varied as mentoring, blogging or creating industrial prototypes.

The dangers of poor metrics are well known — and science should learn lessons from the experiences of other fields, such as business. The management literature is rich in sad exam- prioritize research money risks missing out on an important discovery from left field. It is true that good metrics are difficult to develop, but this is not a reason to abandon them. Rather it should be a spur to basing their development in sound science. If we do not press harder for better metrics, we risk making poor funding decisions or sidelining good scientists.

SUMMARY

- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

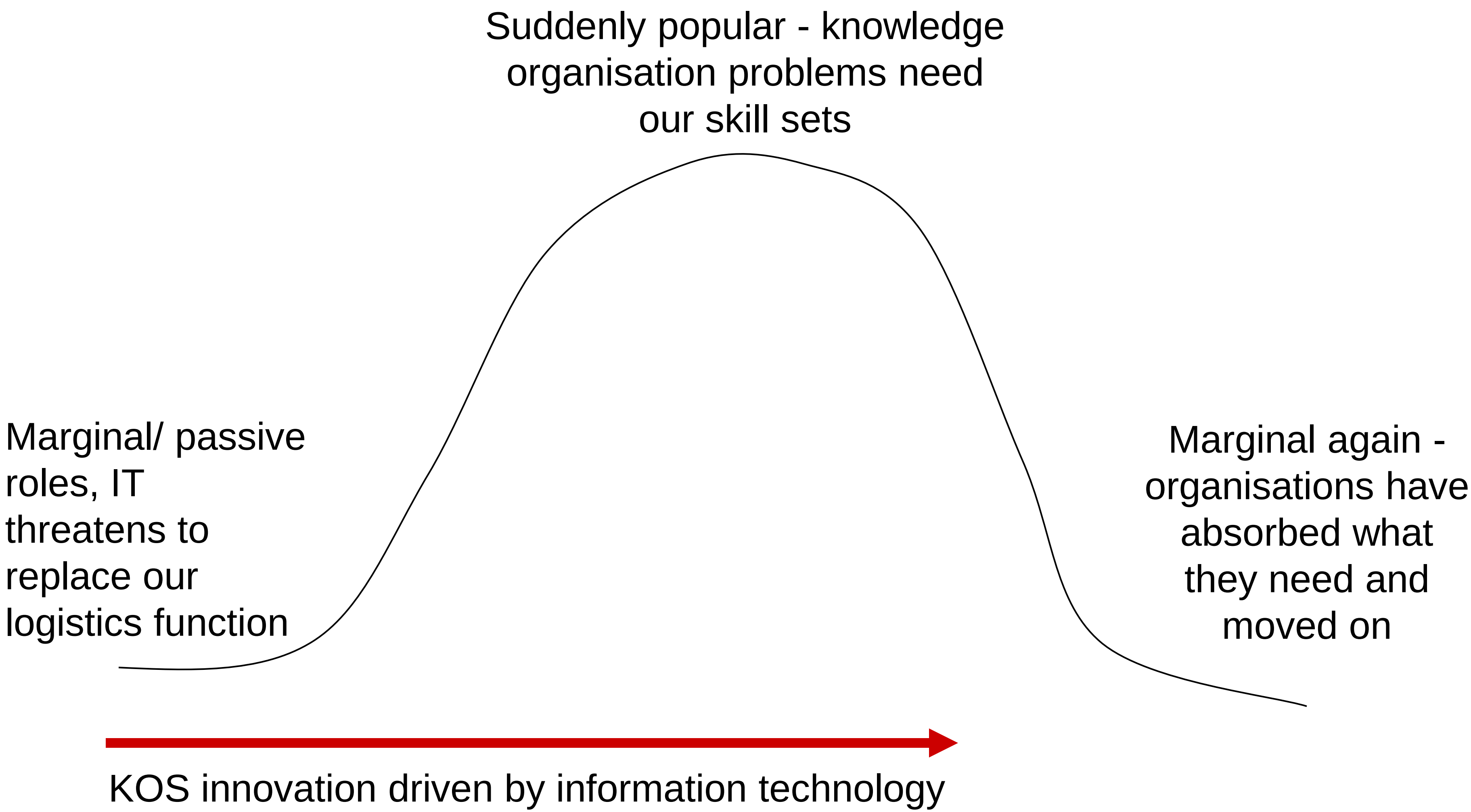
Clean data

Metrics are data driven, so developing a reliable, joined-up infrastructure is a necessary first step. Today, important, but fragmented, efforts such as the Thomson Reuters Web of Knowledge and the US National Bureau of Economic Research Patent Database have been created to track scientific outcomes such as publications, citations and patents. These efforts are all useful, but they are labour intensive and rely on transient funding, some are proprietary and non-transparent, and many cannot talk to each other through compatible software.

and publishing communities to create unique researcher identifiers using the same principles as the Digital Object Identifier (DOI) protocol, which has become the international standard for identifying unique documents. The ORCID (Open Researcher and Contributor ID) project, for example, was launched in December 2009 by parties including Thompson Reuters and Nature Publishing Group. The engagement of international funding agencies would help to push this movement towards an international standard.

Nature March 2010

Implications



Any Questions?

plambe@straitsknowledge.com