Relating Intellectual Property Products to the Corporate Context

Keith G Jeffery CCLRC, Anne Asserson UiB

Abstract

The knowledge society demands that organizations utilise to the full the IP (intellectual property) they generate and acquire. In a context of research, development and innovation, the IP consists of products, patents and publications (in the widest sense – any stored representation of human intellect). While conventional research publications (white literature) provide much of the visible IP, the 'submerged part of the iceberg' is the grey literature. With OA (Open Access) technologies the line between white and grey becomes blurred, because of accessible postprints and preprints. Significantly, white literature (and grey literature) may commonly be hyperlinked to further grey literature in the form of technical reports, procedure manuals, learning materials, software, data (in all media), communications (e.g. e- and voice-mail) etc.   The hypothesis is in three parts:

a) hypermedia: hyperlinks from white (and grey) literature to the 'submerged iceberg' of grey publications enriches and completes the IP of the original publication;

(b) CRIS: (Current Research Information Systems) that cross-link the material to other entities in the business domain (where the business here considered is R&D) provides a context to enable the end-user to utilize more effectively the publication or collection of publications for their purpose;

(c) Metadata: of high quality for controlling integrity, providing access, providing a description and managing rights is essential.

At both UiB, Norway and CCLRC, UK systems have been built utilising CERIF (Common European Research Information Format which provides a data structure to match (b) above) linked with publications information. Both are based on joint research by the authors, the initial design results having been reported in GL'99 (Jeffery 1999). Both implementations have demonstrated effectiveness with enthusiastic end-users. The UiB system is targeted at providing UiB with research productivity information and uses a relational database structure extending CERIF 2000 with bibliographic entities/attributes. This allows publications to be related to people, to organizational units and to projects. At CCLRC, using relational database technology, a CERIF-based extended corporate data repository (including people, organizational units and projects) is linked with an OA eprints institutional archive, containing postprints and preprints (and also other grey publications). In each case, the relationships are annotated with role (such as author or reviewer) and start date/time, end date/time. The relationships can be many-to-many, thus a publication may have many authors each of which may have affiliations with many organizational units and projects. Each instance of organizational unit, project or person may be recursively related to another; this allows correct representation of hierarchic or network (connected graph) relationships.   For both implementations the next steps involve implementing fully the formalised Dublin Core metadata proposed initially by the authors in GL'99 and since developed further by experience (Asserson and Jeffery 2004). This will provide easier syntactic (structural) interoperation and allow semantic interoperability. However, more importantly this development will allow the end-user to utilise the virtual collection, whatever the media and whatever the quality, in a knowledgeable and context-aware way.

1   BACKGROUND

In a context of research, development and innovation, the IP (Intellectual Property) consists of products, patents and publications (in the widest sense – any stored representation of human intellect). While conventional research publications (white literature) provide much of the visible IP, the 'submerged part of the iceberg' is the organisation's grey literature. This commonly represents its 'know how' or knowledge base (Jeffery 1999), (Jeffery, Asserson et al 2000). There are also legal considerations: many organizations protect their IPR with patents or pre-publication; copyright and database right are counterbalanced by Freedom of Information and Data Protection legislation. Innovation, Technology Transfer, Wealth Creation, Quality of Life are major objectives of R&D, and the reason why national governments, commercial organizations, charitable organizations and even individuals invest in it. Most of the technology upon which we depend today is the result of R&D years ago, and similarly the quality of life we enjoy is largely the result of R&D in topics such as medicine, education, environment. This is the IP.

The WWW (World Wide Web) (W3C), has made e-publishing inexpensive and easy. This has led to an explosive growth of institutional repositories. The Open Access Initiative (OAI) utilised the Dublin Core (DC) metadata standard and harvesting software (OAI-PMH) to link the repositories. The two great challenges for the Web as outlined in (Berners-Lee 1999) are the semantic web (to make the web understandable) and the web of trust (to make it secure). The semantic web is now being constructed, largely by (a) more formal data structures which are suitable for manipulation by first order logic, commonly involving the use of structured metadata and (b) use of domain ontologies to provide