From Oceanographic Data Integration Systems to Knowledge Management – Issues and Approaches

Nurhan Hasanov[1] , Vasil Donev[2], Atanas Palasov[3]

(1) Institute of Oceanology, Bulgarian Academy of Sciences (nurhi@abv.bg)

(2) Institute of Oceanology, Bulgarian Academy of Sciences (donev@sirma.bg)
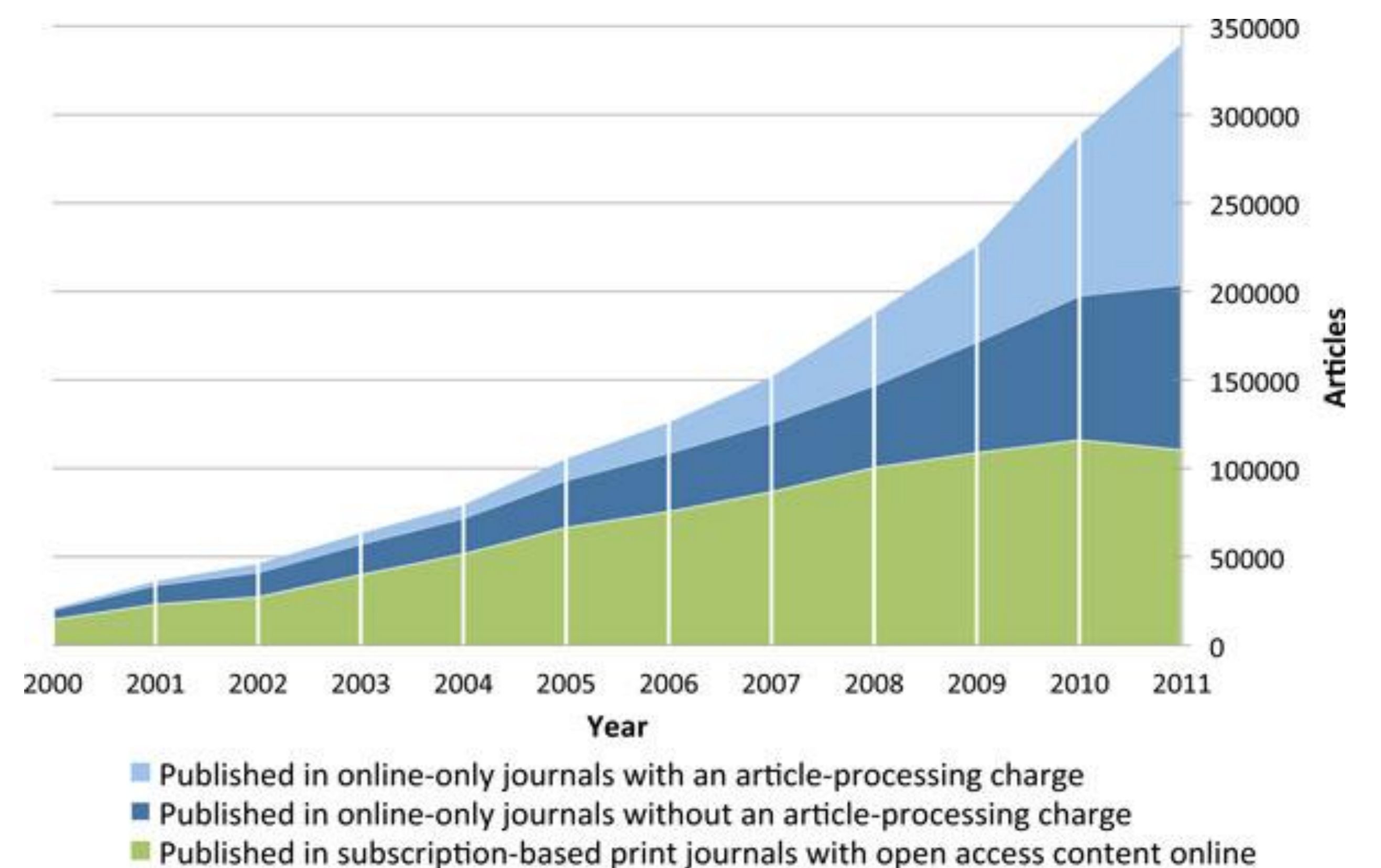
(3) Institute of Oceanology, Bulgarian Academy of Sciences (palazov@io-bas.bg)

Abstract

Oceanographic raw/primary data increases exponentially last decades and this phenomenon strongly desired by the scientists in the past now threatens to become a serious issue while interpreting or simply searching for a particular type of information. Such newly emerged challenge has been seriously considered both institutionally and professionally and is now addressed with some preventive measures in the field of the information technologies sciences which already provide some kind of order in respect of data storage and integration. However, with such vast amount of information now easily available the question of its correct analyzing, rationalization and finally getting certain new facts out of it pose serious task for all participants in the process – scientists, government officials, business bodies, analysts. The bottom line of all these developments is that the focus moves from gathering of primary information to producing new knowledge out of all available information sources. This in turn changes the approaches for managing the information flows – from simple arrangement and search for certain data in the databases to obtaining some kind of existing or new knowledge. The issue is not unknown in other fields of human activities, for instance the business management. For the oceanology science however, the matter is further complicated by the extensive kinds of data being otherwise reliably stored and we may assume nowadays relatively easily available. The transition from data management to knowledge management obviously requires implementation of new instruments for information processing or at least some kind of an upgrade of contemporary data management tools.

Based on the above considerations the article deals mainly with the issues that go in line with the data integration as basis for further knowledge creation or access. Following this path, an attempt is made to tie up the conclusions and the recommendations concerning the knowledge management process to the needs of broader span of end users rather than limiting the results for the oceanographers' community only. Such open approach is determined first of all by the constant deepening of the interrelations between the science and the economy deemed to be an economy of knowledge as often called nowadays. An analysis of the major drawbacks of some of the most used systems for oceanographic data integration is presented and the results of this analysis outline the main directions which should be followed by the institutions in search of extra value from the yet accumulated oceanographic information. Given the complex nature of the oceanographic information, the still existing variety of formats, storages and forms of data, the broad span of interests for this information as well as the new technological instruments for processing, a conclusion can be drawn up that further improvement of the usage of the existing oceanographic information could be reached only through the implementation of semantic technologies in the search process for the information. The idea is exemplified through several approaches adapted to the needs of the institution. On the basis of the proposed solution the presentation also examines the opportunities for the creation of an expert system for niche studies which draws resources from the already known model of semantic integration system.

The fundamental idea that lies behind the approach described in the presentation is the understanding that both science and business are mutually dependent in respect of the progress they seek for and only through creations of technical instruments which open the information space of the science to wider community of users – business, administration, social and nongovernmental entities - we can reach maximum efficiency in transformation of data into knowledge. To prove this the presentation demonstrates both technical and methodological solutions for data and knowledge management. A brief description of the structure of the combined data management system is given with elements of analysis for the most critical parts of it. Finally, the presentation opens a discussion on some further fields of development in the context of the latest achievements of the information technologies and how these in our opinion may affect the functionality and the access features of the system.

- Published in online-only journals with an article-processing charge
- Published in online-only journals without an article-processing charge
- Published in subscription-based print journals with open access content online

Continuously increasing volumes of available oceanographic studies

Practically it becomes impossible even for experienced scientists to keep up with the latest findings which cover more and more fields of interest. The issues related to the search of the necessary information now tend to move from purely quantitative (as commonly being in the past) to more qualitative in view of the expanding multidisciplinary nature of the contemporary primary oceanographic data.

### Information search: Quantitative → Qualitative

The issues related to the search of the necessary information now tend to move from purely quantitative (as commonly being in the past) to more qualitative in view of the expanding multidisciplinary nature of the contemporary primary oceanographic data.

Taking into account that the newest trends in the information technologies move more to knowledge management rather than data or even information management it becomes clear that commonly accepted approaches so far tend to produce less benefits to the interested parties and hence threaten to turn a great deal of the incoming and even existing data sets into useless piles of data which besides the expenses for their production and storage start to generate unacceptable levels of information pollution or noise.
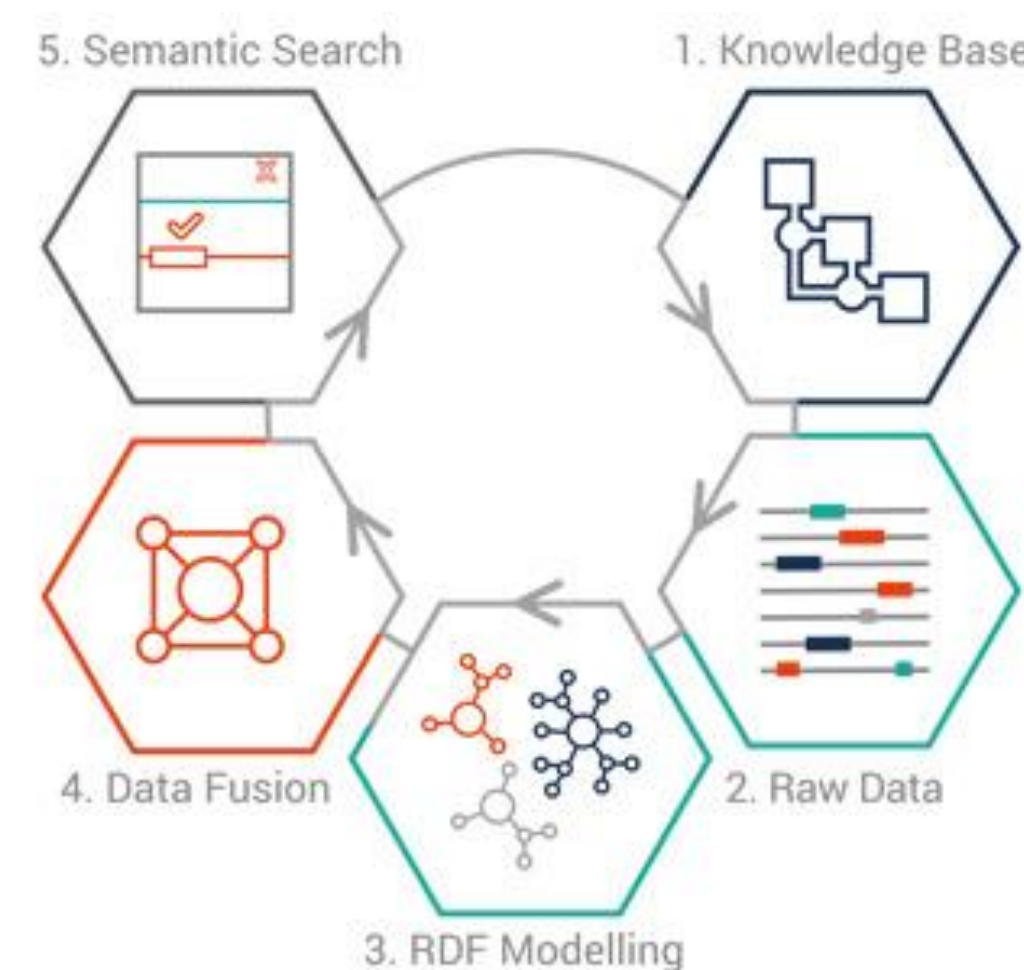
Heterogeneous data, information and knowledge from many sources

Due to historical reasons the existing oceanographic data is stored and still processed for search in a non-semantic way. And while for the primary, raw data there might be some excuse and explanation for such approach nowadays, there is no such logic reason that this should be applied to the unstructured information, the more that it is the exactly this type of information that carries most of the knowledge content. Because of the more directly practical essence of the oceanology it may be expected that more and professionally differently specialized people shall be interested about the knowledge for the ocean environment. A great deal of these users are not supposed to be acquainted in depth for the specifics of the particular field of knowledge they are interested in. The straightforward solution for them is to have at disposal integration and search tools with well developed set of ontologies, which when applied in the appropriate search engines would produce meaningful results even from stored unstructured information. Once created such set of ontologies may be implemented in the structures of the appropriate data integration systems or interlinked with specially designed for semantic search engines.
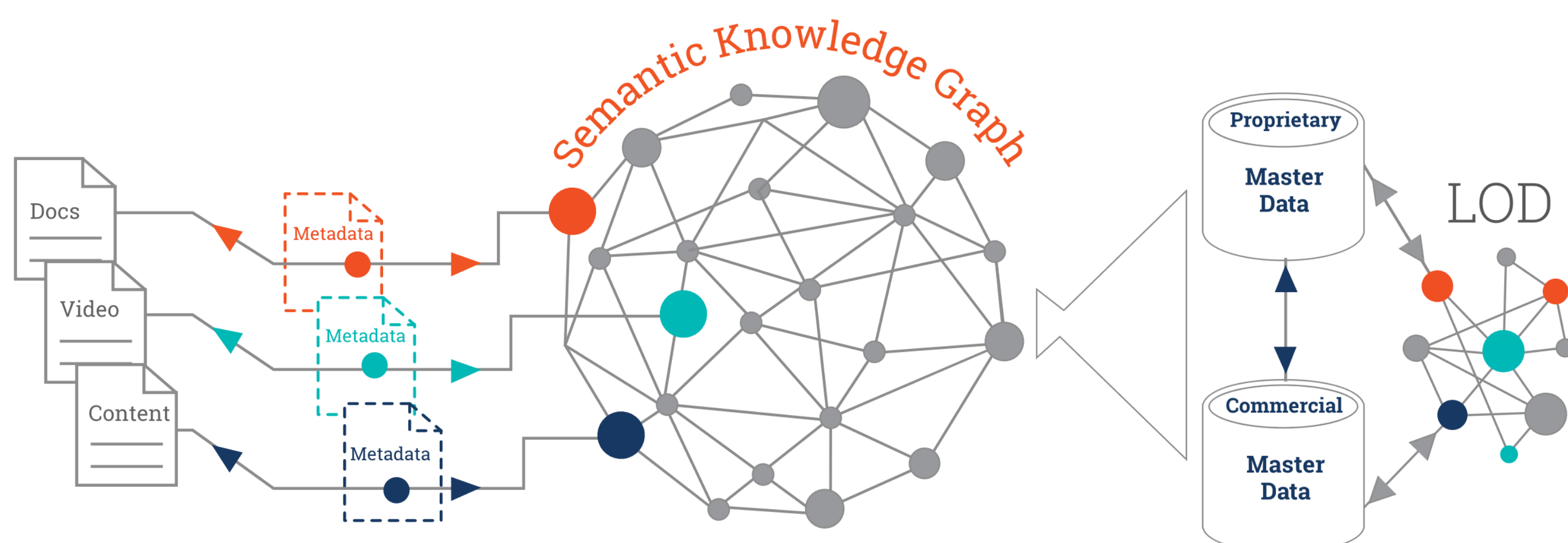
Conclusions

The use of semantic technologies in the integration and search of oceanographic data, information and knowledge is a challenge at the present moment. The reason for this is because there is a little to almost no work done on the creation of the domain ontologies as foundation stone beneath the semantic base system to service the needs of both scientists and other non-scientific entities (governmental, private, non-profit, business). The main advantages of so created semantic data base should be sought not that much in direction of search for primary data (this falls within the interest of a narrower group of ocean specialists) but rather as an opportunity to provide meaningful access to useful data, information and knowledge for non-scientific entities. Not to be neglected as well is the fact that the semantic knowledge base operates much faster compared to traditional relation data bases.