The vast quantities of information generated by academic and industrial research groups are reflected in a rapidly growing body of scientific literature and exponentially expanding resources of formalized data including experimental data from "omics" platforms, phenotype information, clinical data, etc. For information technologies, the challenge remains to support scientists in identifying relevant information, integrating this information in specific "knowledge bases" and formalizing this knowledge across multiple scientific domains to facilitate hypothesis generation and validation and, therefore, the generation of new knowledge. The BioXM™ Knowledge Management Environment efficiently models such complex research environments. This platform is designed for the aggregation of information and the semantic modeling of scientific processes. Any particular area of scientific interest can be modeled as a structured network of related objects. Thus, the BioXM system allows an efficient modularization and abstraction of knowledge.

The actual definition of “knowledge” is indistinct and multifaceted. One aspect of it, certainly, is the awareness of a validated interconnection of details, which are of lesser value when used in an isolated environment. In the BioXM system, knowledge is conceptualized as relationships between semantic objects representing “elements of a scientific domain” (such as genes or drugs). Those relationships are supplemented by the annotation of evidence providing validation for the related objects. Further validated relations with other “elements of a scientific domain” (such as cell types or diseases) may exist, expanding the knowledge network. Specific parts of the knowledge model may be organized in sub-networks (such as a particular signal transduction pathway in an organism of interest), allowing a hierarchical structuring of knowledge. The organization of information in specific projects provides a further efficient mechanism of distinction between separate parts of the knowledge network.

The conceptualization of entire areas of interest in ontologies allows one to use the inherent inference relationships for the exploration of knowledge networks. Entities from external public or proprietary databases accessible by the embedded BioRS™ Integration and Retrieval System can serve as “virtual semantic objects” in the knowledge network. They can also be used as “read-only” annotation of the “real” semantic objects. All semantic objects (such as elements, relations, contexts, ontology instances, or BioRS database entries) can be annotated with additional information. Annotations are form-based and support hierarchical organization of information.

Using the BioLT™ Literature Mining Tool, relationships based on co-occurrence can be extracted from the literature. Curated dictionaries of semantic categories are used for the search, allowing semantic objects in BioXM to be referenced by the search results. This enables the inclusion of the revealed relationships in the existing knowledge network. Search results may be shared with new objects or created for previously retrieved or created objects using the content of specified columns as new attribute values. Data modification sequences can be controlled by the number of results of a preceding operation (e.g., an object lookup) or by the content of a specified column. In the example shown here, the creation of a new annotation and its assignment to a previously defined context are carried out only if the preceding query returns no results.

The BioXM system provides graphical browsing through the network. An advanced query builder allows a flexible exploration of the knowledge with complex queries using a natural-language-like syntax. Flexible reporting allows the generation of new knowledge. The BioXM system allows efficient modularization and abstraction of knowledge.