Inorganic Materials Database Integration Platform

Victor A. Dudarev

A.A. Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, 49, Leninsky avenue, Moscow

Abstract: Current state of Russian databases on inorganic substances properties is considered. Several modern trends in materials database development are discovered. Special attention is paid to heterogeneous materials information systems integration. Several integration approaches are discussed regarding to consolidation of materials science information. Metabase approach used to integrate heterogeneous materials information systems in Russia on base of IMET RAS is discovered and means to international material information integration are announced.

Keywords: inorganic materials databases, heterogeneous information systems integration, materials science information consolidation.

Materials database development trends

Currently databases on properties of inorganic substances and materials are essential in providing information service for specialists in inorganic chemistry and materials science. As a rule every research center in a knowledge domain is aimed at its own database (DB) creation. So we have got a number of databases on properties of inorganic substances and materials that cover different aspects of materials science. At present rich variety of databases (DBs) on properties of inorganic substances and materials have been developed and maintained all around the world. Traditional areas, which are covered by DBs, are thermodynamic, thermo-chemical, crystallographic and crystal chemical properties. The majority of large industrial corporations support development of DBs that contain information on physical, technical and technological parameters of inorganic materials and substances. Nowadays several essential trends in modern DBs development can be highlighted [1]:

- Providing access to DB information via Internet.
- Using of powerful DBMS: Microsoft SQL Server, Oracle, etc.
- Information quality expert assessment. Great attention is concentrated on the quality (reliability) of stored information. Highly skilled specialists are engaged in development process of the most advanced commercial information systems for data gathering and expert estimation of data reliability. So users receive recommended values passed filtration for elimination of misprints.
- Providing DBs with information analysis tools from traditional thermodynamic calculations (statistical procedures) up to modern means for data regularities search allowing prediction of chemical objects behavior and making decisions.
- DBs on inorganic substances and materials integration. It means creation of interoperation mechanisms between DBs that allow user to obtain the complete cumulative information on properties of a certain substance.

In this paper the last in the list problem is discussed. Information integration is one of the most important tasks today. The data on various properties of a certain substance or material are distributed among different heterogeneous DBs. The chemist or material scientist has to look through a great number of DBs in order to find the necessary information. Therefore some superstructure above DBs is required that will allow obtaining some cumulative (integrated) information on all set of substance properties stored in different information systems. That is why DBs integration is necessary. Ten years ago heterogeneous DBs integration task was extremely complicated considering IT status, if possible. At present, period of such an information fragmentation seems to be coming to the end due to rapid IT-industry development and integration efforts. Current progress in science and technique stimulates concentration of diverse information on physicochemical substances properties. Modern polyfunctional materials development requires from us high standard of knowledge in different substance properties. Efficient online information service (for materials science engineers and chemists providing full data from reliable sources) decreases baseless papers' duplication and ultimately it reduces cost and time required for modern materials development.

Integration problem solution is concerned with several fundamental difficulties. Databases on inorganic substance and material properties were developed in various organizations and countries and thus they use different database management and operating systems. Taking into consideration differences in data quality, data expertise procedures, data formats, languages and many other aspects it should be stated that full and smooth integration of information resources is extremely difficult. However taking into consideration the peculiarities of DBs on inorganic substance and material properties some steps towards DBs integration can be made nowadays. In this paper integration approaches will be briefly discussed and the special attention will be devoted to approach developed in IMET RAS to integrate DBs at Web-application level. This approach can be used for close integration of Russian and Japanese DBs in this knowledge domain as well.

Database Integration Approaches Overview

Principally there are three approaches to database integration:

- Data Warehouse technology based on ETL (Extract, Transform, Load) paradigm [3].
- EII (Enterprise Information Integration) technology [4].
- EAI (Enterprise Application Integration) technology [5].

These approaches can be used to solve wide range of problems: from real-time integration to batch integration and from data integration to applications integration. Fig. 1 illustrates these approaches application area in relation to different task types [2]. The EII technology is the best approach for real-time data integration. The ETL technology allows the best batch data integration. The EAI technology gives the best results at applications integration in real-time or batch modes.

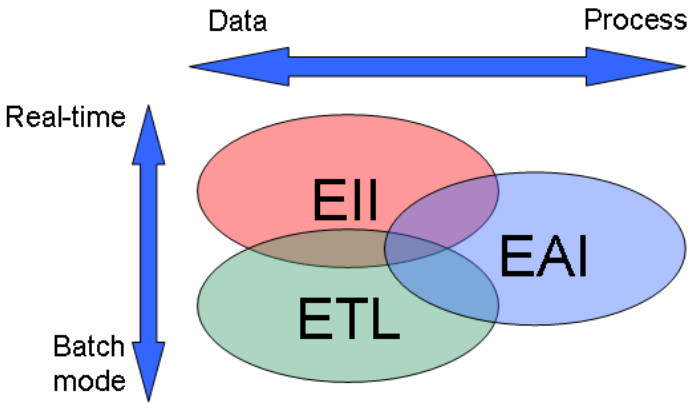

Fig. 1 Modern approaches for information systems integration.

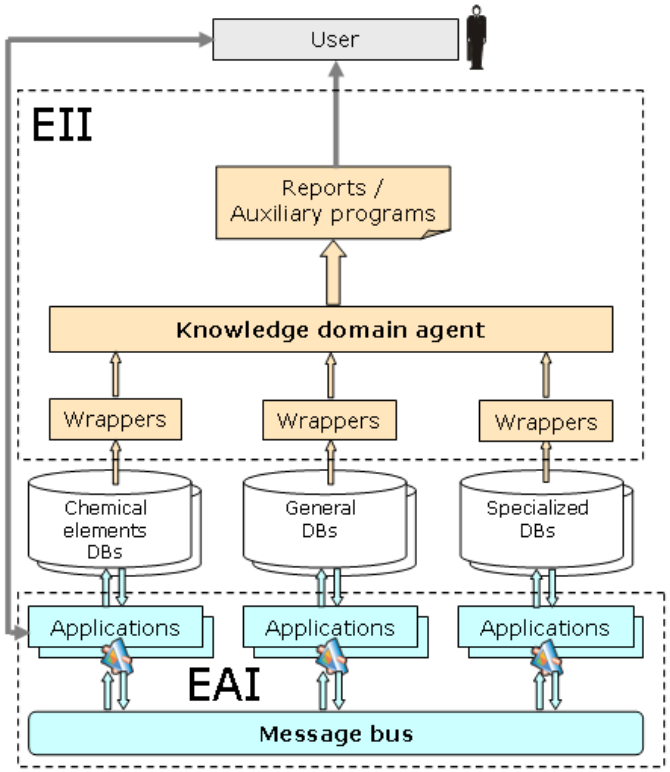

Fig. 2 IMET RAS database integration concept.

It is necessary to take into consideration that every data center on materials properties is a point of information concentration and data analytical processing based on different software and hardware. The technology of information accumulation and data processing has been settled down in each organization. So great investments that were made in hardware and software do not allow to transport mechanically all the data into some centralized database. Moreover many DBs on material and substance properties are equipped with auxiliary programs for substance parameters calculation. Therefore taking into consideration current

development conditions of databases on inorganic substance and material properties the integrated system based on both EAI- and EII-technologies was proposed at IMET RAS as a basis for database integration [6, 7] (fig. 2). This complex approach allows integrating dynamically a plenty of heterogeneous databases that are supplied with any computational subsystems.

Implementing EAI-integration

Historically EAI was the first integration approach applied at IMET RAS for integration of DBs on inorganic material and substance properties. In EAI database Web-interfaces are integrated. This integration approach main essence is that we are not going to integrate databases themselves, but we integrate their proprietary user interfaces only. Let's consider implementation of this integration technology.

Metabase Concept

When integrating databases at Web-interfaces level it's required to provide facilities for browsing information contained in other databases. This information should be relevant to the data on some chemical system currently being browsed by user. Let's consider the following example. User who browses information on Ga-As system from "Diagram" database should have an opportunity to get information for example on piezoelectric effect or non-linearoptical properties of GaAs substance contained in "Crystal" database.

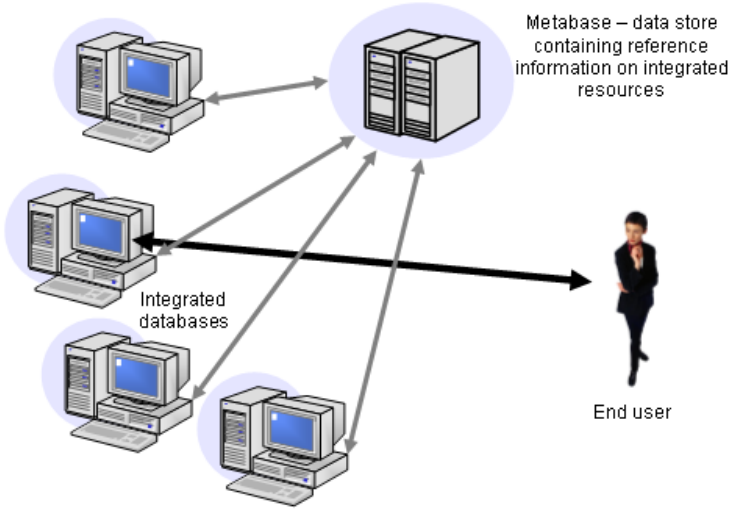

Fig. 3 Metabase concept.

So it's obvious that when designing distributed information system, it's required to provide search for relevant information contained in other databases of distributed system. Thus, we hardly need to have some active data center that should know what information is contained in every integrated database. Obviously some data store should exist that somehow describes information contained in integrated database resources. In this manner, we come to the metabase concept – a special database that contains some reference information on integrated databases contents (fig. 3).

In our case, it is information on chemical systems and their properties. The amount of this metainformation should be enough to perform search for relevant information on systems and corresponding properties.

Service Oriented Architecture

As far as it is required to access metabase from different software platforms Web-service was developed that based on SOAP (Simple Object Access Protocol) and XML (eXtensible Markup Language). So it is a special Web-service that is used by external databases to load information into metabase (fig. 4). Web-service is a kind of entry point for integrated databases and is responsible for all information update in metabase. This Web-service is accessible at http://meta.imet-db.ru/MUService/MUService.asmx. In information exchange XML-documents of special format is used. Such documents contain information on external database contents update. Xml-document format should satisfy a particular XML-schema [7].

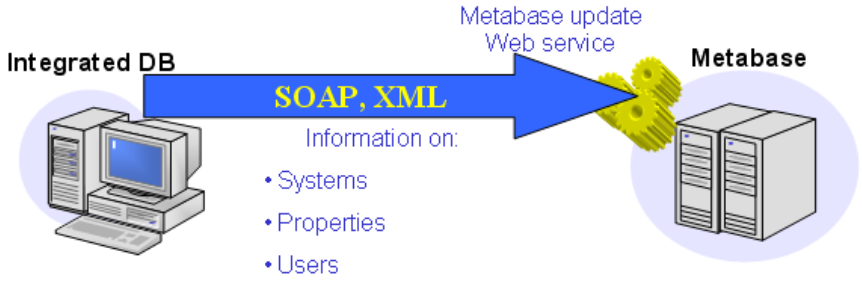

Fig. 4 Metabase update Web-service.

Web-service for relevant information search is available at http://meta.imet-db.ru/Service/Service.asmx. It is used for relevant information search by integrated system according to above mentioned rules. So it is entry point for integrated database systems when it is required to find relevant information according to metabase. Communication with this service is implemented via SOAP protocol according to WSDL-description at http://meta.imet-db.ru/Service/Service.asmx?wsdl. Information on relevant chemical systems and its properties that reside in other integrated databases returns in XML-document. XML-document format satisfies XML-schema available at http://meta.imet-db.ru/Relevance.xsd.

Conclusion

The complex approach to information integration combining the integration at a level of user and data interfaces (EAI+EII) is proposed in IMET RAS. EAI-based integration of databases on inorganic materials is described in detail. Such tasks as relevant information search in integrated system and transparent user transition between Web-applications (taking into account the security issues) implementation were solved at Web-applications integration. Special database called metabase is introduced and its role is briefly described for relevant information search mechanisms.

Nowadays the EAI-integrated information system includes five DBs that were developed at IMET RAS: DB on the properties of inorganic compounds "Phases" [8], DB on phase diagrams of semiconducting systems "Diagram" [9], DB on substances with significant acousto-optical, electro-optical and nonlinear-optical properties "Crystal" [10], DB on width of the forbidden zone of inorganic substances "Bandgap" [6] and DB on properties of chemical elements “Elements” [11]. One of the most important features of the developed integrated system is that DBs which have been included into integrated system were created with various DBMS using essentially different computer platforms: Sun (DB "Diagram") and Intel (other DBs) and different operational systems: Sun Solaris (DB "Diagram") and Microsoft Windows 2008 Server (other DBs). This fact proves that integration infrastructure

architecture developed at IMET RAS is efficient to consolidate heterogeneous DBs Web-applications. So the system allows to users to move freely between different applications.

Currently IMET RAS databases are integrated with NIMS AtomWork database, so Russian users can transparently browse relevant information in the Japanese database. Taking into consideration English user interfaces of "Crystal" and “Bandgap” it’s possible to smoothly add other databases to the integrated system, e.g. other NIMS databases. This will allow users of every integrated database to see information contained in other databases of EAI-integrated system. This would be the first step to international database integration between Russia and Japan. In future database integration at data level can be discussed and corresponding standards be developed. That will allow providing strong international information service for specialists in inorganic chemistry and materials science.

References

[1] N. Kiselyova, S. Iwata, V. Dudarev, *et al.* Integration Principles of Russian and Japanese Databases on Inorganic Materials // Int.J. “Information Technologies and Knowledge”. 2008. V.2. №4. P.366-372.

[2] C.Imhoff. Intelligent Solutions: Understanding the Three E's of Integration EAI, EII and ETL. DM Review Magazine, 2005, apr. (http://www.information-management.com/issues/20050401/1023893-1.html).

[3] R.Kimball and J.Caserta. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons, 2004.

[4] J.P.Morgenthal. Enterprise Information Integration: A Pragmatic Approach. Lulu.com, 2005

[5] J.P.Morgenthal. Enterprise Applications Integration with XML and Java. Prentice Hall PTR; Bk&CD Rom edition, 2000.

[6] V.A.Dudarev, N.N.Kiselyova, V.S.Zemskov. Integrated system of databases on properties of materials for electronics. Perspektivnye Materialy, 2006, N.5 (Russ.).

[7] V.Kornuyshko and V.Dudarev. Software Development for Distributed System of Russian Databases on Electronics Materials. Int. J. “Information Theories & Applications”, 2006, v.13.

[8] N.Kiselyova, D.Murat, A.Stolyarenko, et al. Database on ternary inorganic compound properties “Phases” in Internet. Informazionnye resursy Rossii, 2006, N.4 (Russ.).

[9] Yu.I.Khristoforov, V.V.Khorbenko, N.N.Kiselyova, et al. Internet-accessible database on phase diagrams of semiconductor systems. Izvestiya VUZov. Materialy elektron.tekhniki, 2001, №4 (Russ.).

[10] N.N.Kiselyova, I.V.Prokoshev, V.A.Dudarev, et al. Internet-accessible electronic materials database system. Inorganic materials, 2004, v.42, №3.

[11] N.N.Kiselyova. Computer Design of Inorganic Compounds. Application of Databases and Artificial Intelligence. Nauka, Moscow, 2005 (Russ.).