Intelligent Automation And Soft Computing, 2019
Copyright © 2019, TSI® Press
Vol. 25, no. 3, 615–624
https://doi.org/10.31209/2019.100000116

Building an Open Cloud Virtual Dataspace Model for Materials Scientific Data

Yang Li [1], Jianjiang Li[1], Peng Shi[2], Xiaoning Qin[3]

[1]School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing, 100083, China
[2]National Center for Materials Service Safety, University of Science and Technology Beijing (USTB), Beijing, 100083, China
[3]Dawning Information Industry Co. ,Ltd, Sugon Building, No.36 Zhongguancun Software Park,No.8 Dongbeiwang West Road, Haidian District, Beijing,100193,China

ABSTRACT
The applications to process large amounts of materials scientific data at different scales, have been placed the field of materials science on the verge of a revolution. This domain faces serious challenges, including diversity of format of scientific data, and missing unified platform for sharing. A Virtual DataSpace model and the evolution model is introduced to organize heterogeneous data according to the user requirements and track the variations of data. The open cloud model is embedded in a materials scientific data sharing platform in our experiments to verify its effectiveness. The results show the model has made efforts making more information available and useful for researchers.

KEY WORDS: Cloud Computing, Intelligent services, Materials Scientific data, Ontology, Virtual Dataspaces.

1 INTRODUCTION

AS early as 1980, the "Big Data" concept is proposed which is introduced as the cadenza of "the third wave" by Toffler (1980). However, in recent years, new challenges have been made for better manage, reuse and share the "Big Data" proposed by Howe et.al. (2008) and Lynch (2008). The importance of data and information has become more prominent and data has dominated the information field. The main object of big data management is to provide intelligent services for the demands of the domain application. In engineering software domain, needs of big data management in domain expert systems are more urgent. To satisfy the requirements of different domain users and provide personalised intelligent services, research on how to efficiently process big data is essential. The main issue that facing the big data research field, is to extract valuable knowledge from distributed and heterogeneous data resources. To resolve this problem, researchers studied the related technical solutions as well as the methods of data analysis and model evolution.

Materials scientific data are mainly produced from scientific experiments, theoretical study, emulation, manufacture, simulation, manufacturing and etc. Materials scientific data has the same characteristics of big data. However, there are still some barriers to amplify data sharing and large-scale data analysis within the field of materials science.

(1) Different organization, softwares or materials researchers may prefer different data work flows. Data workflows patterns in materials engineering field vary differently and are depended on several factors, such as application's unique focuses, materials researchers' personal preferences, and different methods of data acquisition;

(2) Lack of platform to share data, use data from diverse areas including companies and research organizations;

(3) Diverse structure of data and limited data standards are existed in materials science field.

Although a wealth of knowledge of materials engineering has been provided, there are still tremendous difficulties in reusing and reproducing other's material data. Different organizations choose variety sets of tools, data warehousing methods, data analysis tools, which make the centralization use and sharing of materials data far more challenging.

In field materials engineering, the information systems problem is in how to manage data efficiently and apply intelligent services. The user requirements of data include deriving the relationships and hidden

CONTACT Jianjiang Li liyang@ustb.edu.cn

knowledge of the data, and acquiring data that change with the user requirements. Obviously, the technology of the traditional database methods is not fit for the new demands. Therefore, a new data management mode that provides interesting data and intelligent services for different users is needed.

The main aim of this paper is to propose the representation and evolution of VDS model for providing intelligent services for materials researchers. A new representation method based on ontology is proposed for big data, and a dynamic evolution algorithm for the continually changing data is presented. Finally, we construct a materials scientific data sharing platform in our experiments to verify the model's efficiency.

The paper is organized as follows. The related research about dataspace is investigated In Section 2. In Section 3, the representation and dynamic evolution method for VDS model is introduced. The application case of a material intelligent service is performed in experiments of Section 4. The conclusions and further work is described in Section 5.

2 RELATED WORK ON DATASPACE

THE concept of "dataspace" was introduced as a new data service mode in 2005, which is used for big data storage and management proposed by Baladron et.al. (2012) and G. Lin, W. Mingzhen. (2018). It is the extension of databases based on subjects. The subjects refer to the dataspace users, organizations, or specific applications. Thus the dataspace include subjects, the data sets that related to the subjects, and the relationship between different data sets. The main issue that addressed here is to collect, manage and organize the distributed and heterogeneous data resources to form a uniform data platform for better data services. Table 1 shows the different characteristics between traditional databases and dataspace.

It clearly shows that dataspace is more suitable for complex data environment which is illustrated by Y. Llsun, etc. (2018). It is essential for heterogeneous data resources and dynamic data flow. The dataspace technology can be used in many domains, such as Personal Information Mangement (PIM) proposed by Ballings and Poel (2012) and Blaunschi et.al. (2007), scientific data spaces proposed by Cafarelia (2009) and Cai (2009), companies and etc. Different dataspace models are applied for different domain data services. Recent research on dataspace model also include the Resource Space Model (RSM) proposed by Pradhan (2007), the PAD model proposed by Saema et. al (2009), and etc..

(1) **iMeMex** The iMeMex data model is a universal model for personal data management system which is developed by Dittrich J. P. et.al (2007). The data resources are organized and represented by resource view and resource view classes. A tetrad, is used for representing resource view, where η represents the name of the resources,τ is the property of data, χ represents the non-structured resources, andγ shows the relationships of the resources by using a directed graph. iDM is easy to understand. However, it is difficult to use for normal users, because iQL a query language based on XPaath.

(2) **Unified Data Model** The data collected from desktop are organized by Unified Data Model (UDM) in a tree structured form. There are two core concepts in this model, desktop dataspace and dataspace fragment proposed by Pradhan (2007). The node of the tree in UDM can refers to a folder, a MatML file or the content of the MatML. Figure. 1 shows one use case of UDM tree structure. The root represents the top-level folder. Every folder, MatML and the content of the MatML is represented by the node of the tree. Database/information retrieval is used as the data integration method in UDM. A new query language, TALZBRA, has been introduced in UDM. UDM is mainly focused on desktop searching. The rational data, shortcuts searching are not supported.

Table 1. The comparison of traditional databases and VDS.

	Traditional Databases	**Dataspace**
Data Formats	Relational Data	Structured, Semi-structured and Non-structured data
Data mode	Relational Mode	Different Modes(Ontology and Relational)
Data sources	Single source, Isomorphic	Multi-source, Heterogeneous
Data association	Simple, Stable	Complex, Dynamic
Semantic	Without semantic	Progressive semantic
Data access	Accurate and complete results	Currently optimal results
Data Services	Construct before use (Pay-before -you-go)	Construct and optimize while using (Pay-as-you-go)

(3) **Probabilistic Semantic Data Model Semantic Data Model** (P-DM) use probabilistic mediate schema for data integration proposed by Saema et. al. (2009). They are applied for heterogeneous data resources integration to form the dataspace. The probabilistic mediate schema is constructed from a series of data resources schemas by using instructed algorithm. Data resource schema is then mapped to mediate schema using probabilistic semantic mapping method. The query is used for searching in the final mediate schema. The user query rewritten in a structured form, Qm=(Q1,Q2,…, Qm), which is called keyword reformulation. The subqueries are processed in the local data resources. Figure 2 shows the architecture of P-DM. The uncertainty is addressed in this model proposed by

Singh and Jain (2011). The quality of search is improved by using the top k query. But it is difficult to find the accurate schemas that match the users' expectations. And the P-DM model is difficult to extend.

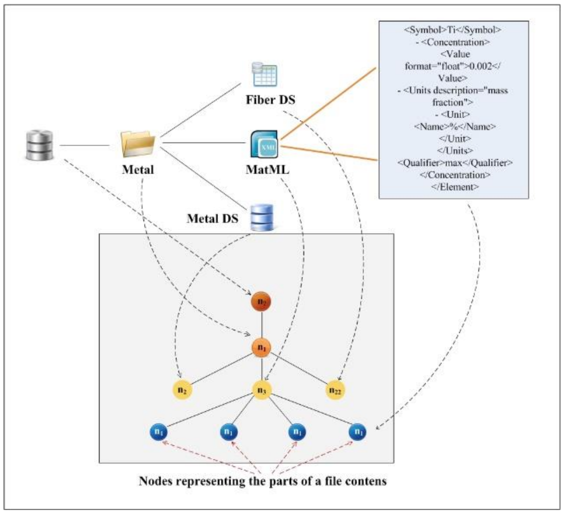


Figure 1. **An Example of UDM tree structure.**

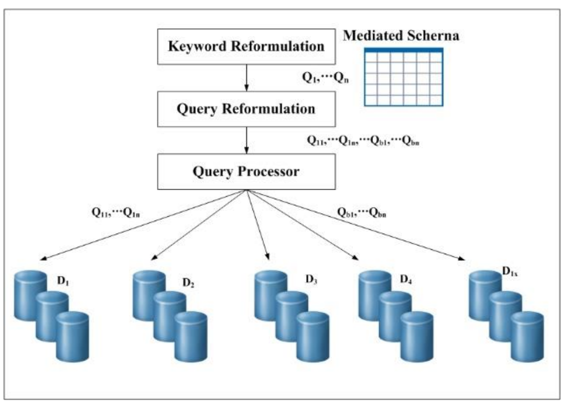


Figure 2. **The architecture of P-DM.**

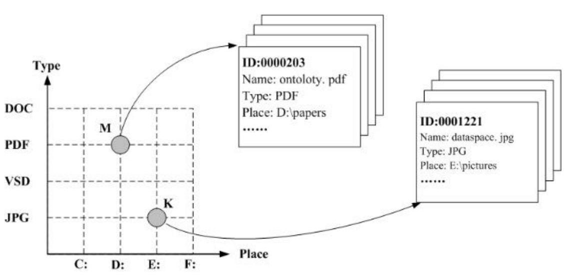


Figure 3. **A data example of the core dataspace.**

(4) **Domain Dataspace Model Domain Dataspace Model** is used by SEMEX, which is used for personal information management proposed by Dong (2005). The personal information data is wrapped, and the relationship between user and data is constructed by domain model. It is used for browsing and searching personal data. Classes and relationships among classes are included in domain model, which is a more ontology-like dataspace model. Thus the semantic searching is allowed in this model. The only problem is that users are involved in the model definition process. It is difficult to match the requirements because of the limitation of the users' understanding.

(5) **Core Space and Task Space Core** space and task space model is used in the OrientSpace, personal dataspace management system, which is also called core dataspace model proposed by Dong et. al (2009). This model is based on the personal requirements and characteristics of the personal information data. This work indicates that the customized dataspace, searching, evolving and integrations are closely related to subjective behavior and activities. Therefore the model is subject-oriented model, the multi-dimensions dataspace is constructed based on the frequency of the data access. Figure 3 shows the data example of the core dataspace. The node represents the data resource. Task space model is based on user tasks. Task is first recognized by taskspace model, the related data is then clustered according to the tasks, the relationships are constructed to satisfied the users' need. The inconsistency of the dataspace, safety of the personal data and the optimization of the data weight coefficients are still needed to be discussed.

(6) **Triple Model**

Triple Model uses triples to form the heterogeneous data, which is very similar to RDF proposed by Zhong (2008). The data are divided into small information blocks, which are called information unit. The triples are defined as (S, P, O), where S is the subject of the data set, P is the predicate element representing the property of the data set, O is the object element which is used for saving data. Figure 4 shows the hierarchical structure of the folder and file resources.

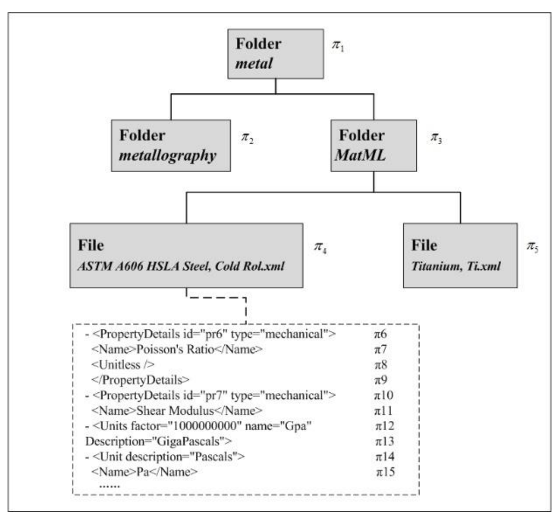


Figure 4. **An Example of Triple Model.**

The Triple Model is easy and flexible for heterogeneous data in dataspace. However, the path expressional queries are not supported in this model and subject predicate object queries are difficult to use.

Many dataspace prototype systems have been developed based on the dataspace model researches. iMeMex is a typical dataspace system based on iDM model. iMeMeX query language is used for searching language. The semantic searching is not supported introduced by Blunschi (2007). Semex is a personal information systems developed by University of Washington proposed by Dong and Halevy (2005). Semantics are used for collecting information. Neighborhood keyword query is used for browsing data and relationships of the data. The only problem of Semex is the performance of the queries and it lacks of full-text index.

Dataspace model is used for web data integration and management in PAYGO proposed by Madhavan et al. (2007). It realizes the ranking of the data resources and search results. The system evolves based on the user feedback. UDI is a dataspace system that integrates data resources in many domains. The integrate model is constructed by merging local models together. Thus the relationship of the data resources is simplified. A weight coefficient is applied for matching property, so the mapping between data resource and medial model is uncertainty. Another dataspace system that addressed in evolution is Roomba proposed by Cafarelia (2009). The user feedback is introduced in the mapping process. The instance string similarity matching method is used for comparing users' expectation. An interactive method for data integration is applied in CopyCat proposed by Ives et al. (2009). A Smart Copy and Paste (SCP) model has been developed in the integration and evolution process, the data is easy to manipulate for users instead of concentrating on different steps. Octopus is a multi-functional dataspace system which includes data integration, searching, data clearing, etc proposed by Cafarelia et al. (2009). The web data integration and keyword search are applied in this system. The searching results are sorted based on users' relevant degree, a best-effort operation is applied for users. The neural network is used for clustering high dimensional data in Self-Organising Maps (SOM) in medical care services proposed by Guerrero et al. (2012). The topological information is preserved in this model.

Most of the work are limited by its schemas and models, and seldom consider the demands of domain fields and data evolution as an important issue for e.g. the work of Yin et al. (2017), the work of Z. Jun and G. Zhenzhen. (2018). And these systems do not fit for the scientific materials data like proposed in the work of Yin et al. (2017) and Yin et al. (2016). Based on our previous research proposed by Hu et al. (2016), a modified model is introduced for better adapting dynamic lifecycle data and on-demand materials applications.

3 VIRTUAL DATASPACE MODEL BASED ON ONTOLOGY REPRESENTATION

AS mentioned above, the representation of virtual dataspace model is proposed in our work in the environment of open services. The model we introduced is the architecture that more emphasize on the demand of users and data characteristics. The concept of Virtual DataSpace (VDS) is established within the cloud computing environment proposed by Hu et al. (2016). This model meets the demands of materials engineering applications by providing a new mode of scientific data service.

3.1 VDS definition and lifecycle

In order to better match the demands of materials engineering applications, VDS applies a scientific data service model, which defines as, VDS= (SUR, DS, DRS, SS). SUR represents the user requirements of data; DS represents the data resources; DRS represents the relationships between data resources; SS represents the materials services.

The main demands of the subjects in VDS are the "data requirements" and the "service requirements". The "computing virtualization" is the core ideology of VDS, i.e., "Data as a Service, DaaS". Therefore, the analysis of the data evolution process in lifecycle of VDS is the key issue of the model construction.

From the requirements of the subjects, to the representation of data, data relationships, and the services that applied to researchers, data vanishing could be considered as the lifecycle of VDS. The process of the lifecycle is shown in Figure 5. Firstly, the data resources are identified in the initial step. Secondly, the matching and mapping progress is continued until the integration pattern is stabilized. The result data resources is the input of the next loop. Then the VDS model is improved based on the user feedback in the using and evaluating step. The evolution of data is supported in the improvement step. Finally, the VDS model is evaluated until it fits the requirements of data services.

To achieve a data service in VDS, first we should determine the conceptual model of VDS. The basic framework of VDS model includes, (see Figure 6) (1) large-scale of heterogeneous data resources, (2) a logic model based on various data types, (3) a semantic model based on ontology supporting data integration and information extraction, (4) a requirement model based on the interests of subjects and demands of application services, (5) the feedback and evolution of the subject, which obtains the user expectations, optimizes the VDS and improves the accuracy of data applications.

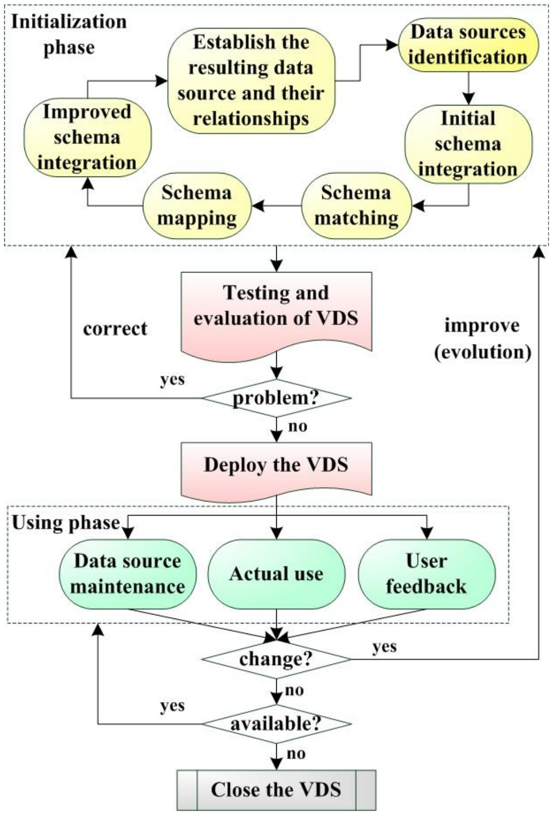


Figure 5. The evolution of data lifecycle in VDS.

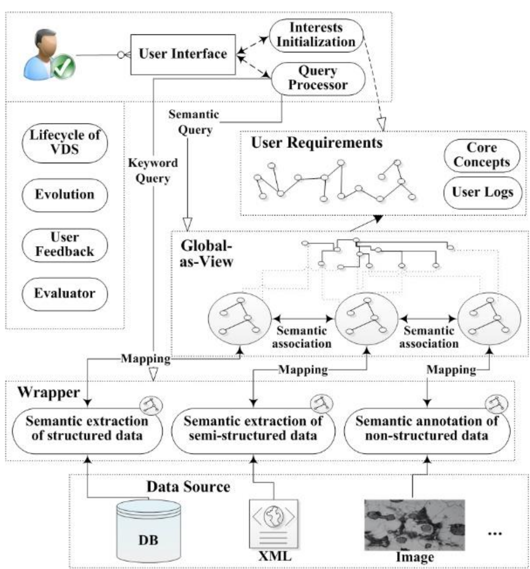


Figure 6. The Structure of the VDS Model.

3.2 Data representation based on ontology

More differences of solutions could be changed because of the small changes in data sets. Thus the knowledge information construction needs to represent the concepts and relationships of concepts based on ontology in materials engineering services. By using ontology, the heterogeneity of data resources could be resolved, and the complex data relationships could be built efficiently.

The five-tuples is the definition of VDS model, (Si, Cc, DSi, RV, OM), i=1,2,…,I, where I is the amount of subjects in VDS. Si represents the subject. Cc is the core concept sets which include subclasses and hierarchical classification structure. DSi is defined as, DSi = (DSstuc, DSsemi, DSnon), which represents the distributed data sets. RV represents the logical semantic model which is based on the keyword query. OM is the ontology model based on semantic mapping and searching.

VDS is composed of sub virtual dataspace (s-VDS). The data set of S-VDSi is defined as: DSi = $\sum$ DEid, d=1,2,⋯,Di, in which Di is the amount of data entities in the sub VDS of subject i. DEid is defined as: DEid = {Dt, Dlevel, Dsour, DMat, Donto, Ddesc, Wi-d}, where Dt represents the data type, which includes the numeric value, image, text, etc. Dlevel represents the data level, which includes the class, property, individual, etc. Dsour represents the resource of data, i.e. the location of meta data.

DMat represents the material data, like the experimental, performance, structural or simulated data. Donto represents the data ontology as the unique identifier of data entity. Ddesc represents the data description, which related to the Dt. Considering the numeric values, Ddesc is described as <value, range, accuracy, max, min, unit>, and considering the text format, Ddesc is defined as <text content, text length, text url>, considering the image format, Ddesc is defined as <image name, image type, image description, image size, image url>. Wi-d denotes the weight of data d for subject i, which is used for constructing mapping and reasoning ontology. Thus the semantic ontology of data could be built by Donto and Dtype according to different Dlevel.

Dtype is composed of three categories of different data types: the structured data, the semi-structured data, and the unstructured data. Ontology of knowledge representation is built by different methods according to table, XML, and image files. The ontology construction method could is shown in Figure 7.

(1) The OWL ontologies rules for structured table,
   (a) Ordinary tables are converted to class or subclass (OWL:Class or OWL:SubClass)
   (b) Join-tables and constraints of tables are converted to object properties (OWL:ObjectProperty).
   (c) Columns are converted to data properties (OWL:DataProperty).
   (d) Rows are converted to individuals. (OWL:Individual).

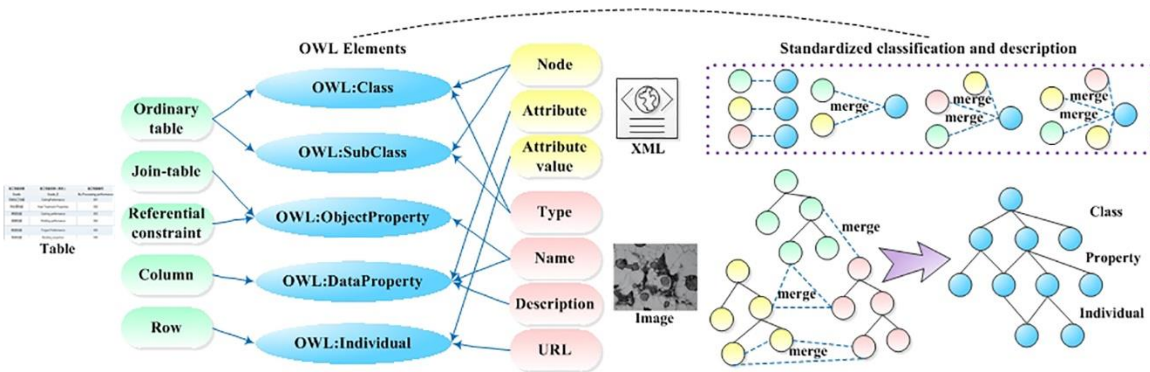


Figure 7. **The ontology construction method**

(2) The OWL ontologies for semi-structured XML,
   (a) XML nodes are converted to OWL classes.
   (b) XML attributes are converted to OWL data properties.
   (c) XML attribute values are converted to OWL individuals.
   (d) Parent-childs are converted to class-subclass relationships.
   (e) Element-attributes are converted to class-data property relationships.

(3) The OWL ontologies of images,
   (a) Image types are converted to classes or subclasses.
   (b) Image names are converted to object properties or data properties.
   (c) Image descriptions are converted to data properties.
   (d) Image URLs are converted to individuals.

Dlevel classified the converted data to different categories. The semantic information of data is described by using Donto. The similar data entities are merged according to the semantic information. Thus the final global semantic view is achieved.

3.3 Dynamic evolution of VDS model

The VDS model construction is processed by using the feedback results of the VDS in the evolution. The user expectations are shown in the feedback results.

There are three different types feedbacks from requirements in the model,

(1) Active feedback: It shows both the acceptance and negation. And also the feedback from application is included.
(2) Behavior feedback: The behavior information is extracted by analyzing the user behavior on services, pages browsing and data manipulation.
(3) Variations of data resources: The variations of data can be detected by mining the behavior of active response and large-scale of data.

The algorithm of evolution of our model by using feedback instances is shown in Table 2. Firstly, the candidate mappings are selected as input. The input mappings contain the knowledge of classes, relationships and attributes of the local s-VDS. Then the user feedback is used for mapping and match local data resources and sematic view of the data sets. The cases that not fit for users' requirements are identified. Finally, the new mapping that meets the user's requirements is constructed by merging and correcting the existing mappings and the refined mappings is formed.

Table 2. **The evolution algorithm based on the feedback instances**

Algorithm. RefineMappings (Map *Mi*, UFI *instan*)

```
Inputs Map: candidate mappings
       UFI: user feedback instances
Outputs Map: refined mappings
Begin
 If ( Mi ≠ null){
   S_Map = Mi;
   O_Map;
   Foreach O_L ∈ S_Map {
     If (O_L ≠ null)
       {Add <O_L, O_G> To O_Map}
   }
   AnnotateMappings(O_Map, UFI);
   C_Map = CombineMappings(O_Map);
   Return C_Map;
 }
End
```

Overall, the accuracy of data access, the expectations of users and the service quality of VDS are improved by using the user feedback mechanism in the evolution algorithm of VDS. The dynamic evolution is suitable for the lifecycle of VDS. The evolutionary algorithm performs well for the dynamic process, which considers an important characteristic which distinguishes the model from other traditional models of data management.

4 EXPERIMENTS AND RESULTS IN MATERIALS SCIENCE ENGINEERING

IN order to testify our VDS Model, a materials science data services platform is built. The structure of the platform is shown in Figure 8.

Distributed and heterogeneous data resources are integrated in this platform in materials science field. Three different types of data are organized in this cloud platform, including values from databases, the XML, HTML files, images, PDF files and etc. The data can be obtained by browsing meta data, navigation of the data, keyword searching and etc. The data can be updated by using the submit module.

Different converted rules are developed to transform data resources into a unified form. For example, the relationships of the databases are used for converting structured data. The annotation method is used for unstructured data like images. And natural language methods are used for PDF files. The domain ontology of the semantic could be formed by using the converted data sets. The semantic information could be extracted from complex data resources based on the open cloud service environment in order to reach the users' requirements (see Figure 9). The knowledge of materials engineering is established by semantic mapping and evolution. "Protégé" is used for generating materials ontology OWL file. "TouchGraph" is applied to generate the visual semantic view of the model (see Figure 10).

Materials scientific engineering applications are applied based on the ontology representation model. A specific materials scientific engineering service. materials selection, is applied in our platform, which is shown in Figure 11. The material category and parameters of performance are selected in the first place. The retrieved relevant materials are shown according to the combination of different conditions. Further detailed information could also be viewed in order to get more information. Finally, the recommended information could be sorted according to the suggested degree. The effectiveness of the ontology representation and intelligent materials services are verified in this application.

The materials scientific data sharing platform has been onlined for more than 2 years. The visit quantity has achieved 200,000. 600,000 different materials data has been collected by this platform, and its number is rapidly growing. 1829 database tables, 729 xml files, and 5293 unstructured data files including images, videos are accumulated in this platform. 50,000 data items are selected from different data types as the testing sets, in order to compare the accuracy and time expense. The experiments are performed to compare the query of tables, XPath query for XML files, the

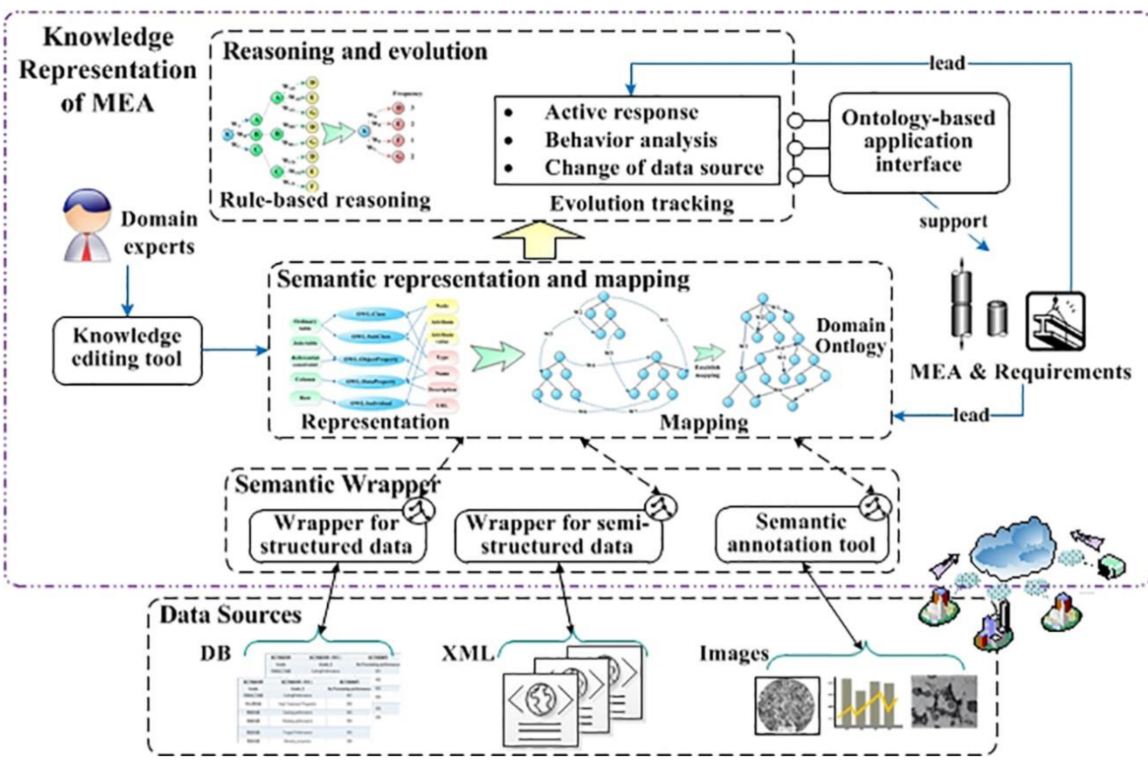


Figure 8. The structure of Materials Scientific Data Services Platform

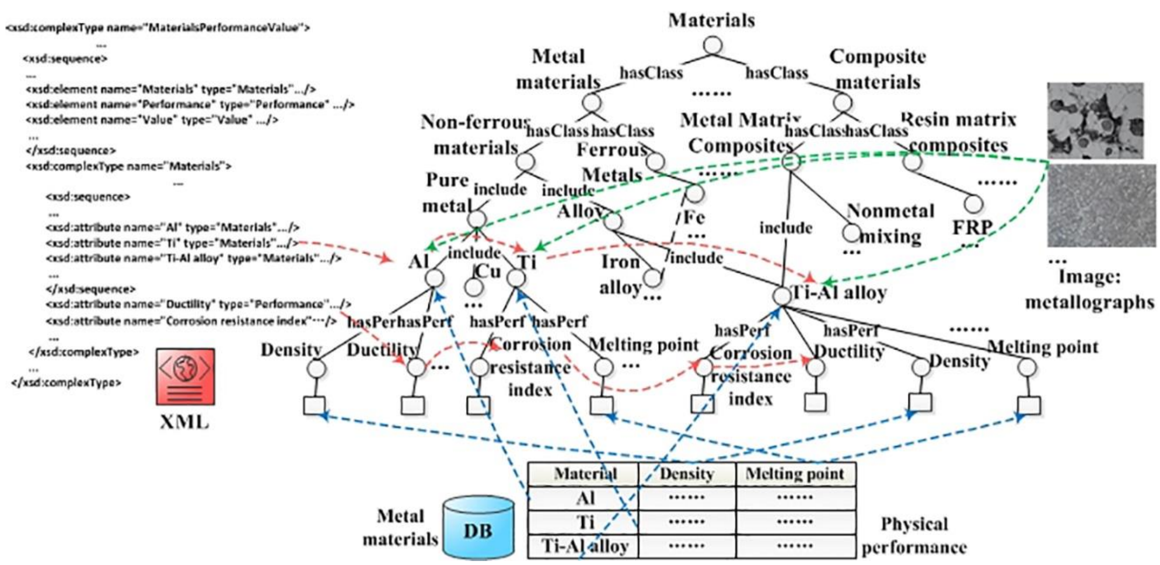


Figure 9. **The structure of Materials Scientific Data Services**

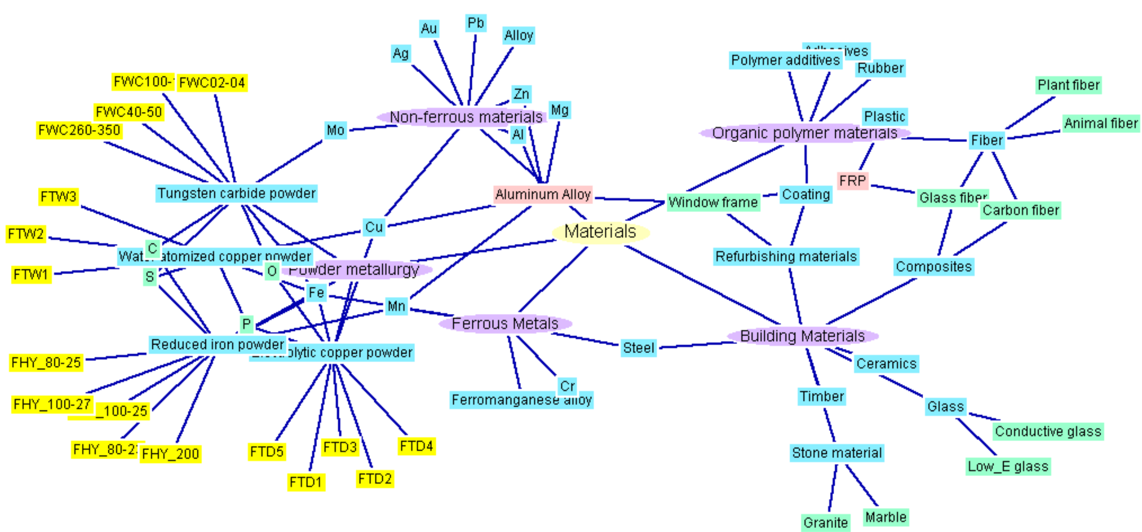


Figure. 10. **Partial visual knowledge representation model of VD**

query for unstructured data files, and the semantic query based on the ontology representation. It shows that with the increasing number of data items, significant increase of accuracy and less time expense are achieved by our model (see Figure 12). Thus the ontology representation for the data resources are more efficient for the data services in our experiments. Therefore, personalized and accurate data service for materials scientific engineering is supported by our platform.

5 CONCLUSIONS AND FUTURE WORK

AN open cloud virtual dataspace model for materials scientific data is proposed in this paper. This model provides a new mode of data sharing, management pattern and data services in specific domain field. The definition of the model, construction method, and data representation based on ontology are described in details. And a dynamic evolutionary algorithm is also introduced for better matching the data lifecycle and meeting the user's changing requirements. Finally, a material scientific data sharing platform is constructed to manage materials data and offer intelligent services for material researchers. The model is also verified in the platform, which shows the efficiency and effectiveness of the data accessing and data services.

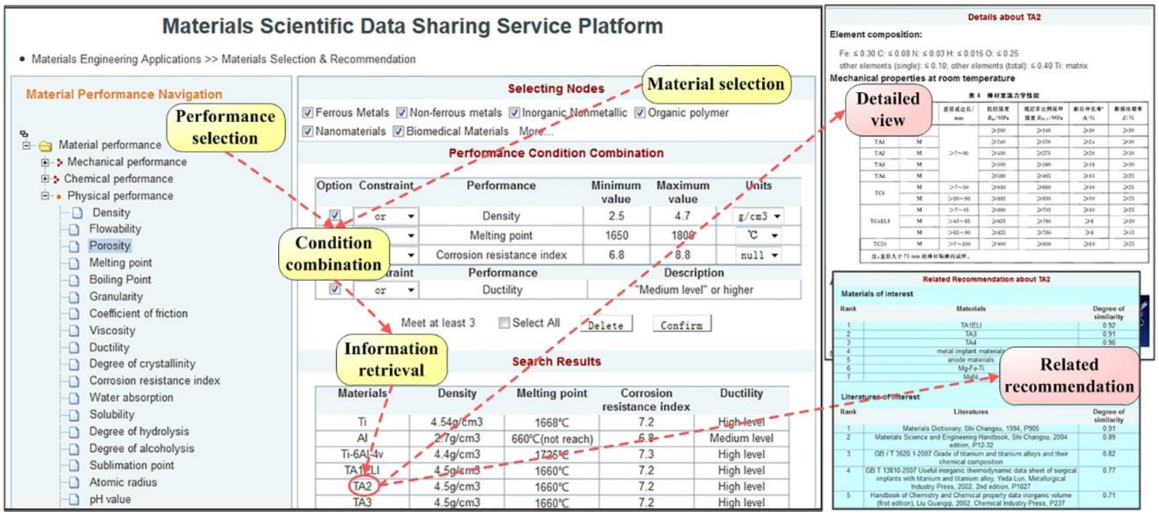


Figure. 11. **Materials selection service case in our platform**

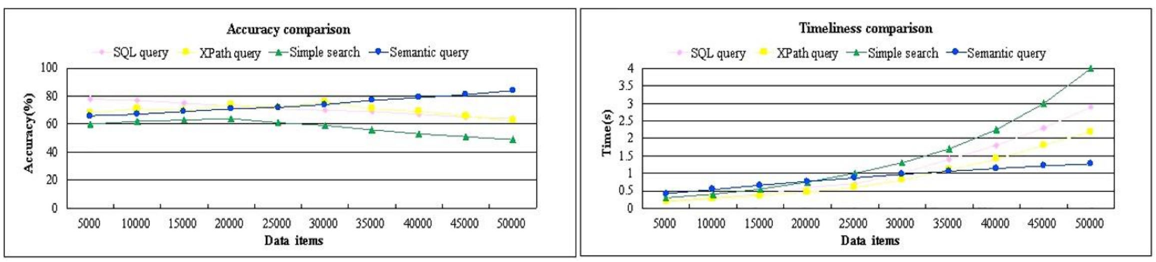


Figure. 12. **The experimental results of ontology representation model**

The future work of the open cloud virtual dataspace model includes, improving representations by designing an automatic construction method, developing an evolutionary algorithm for more complex workflows in materials engineering, and performance more rigorous and large-scale of experiments on the materials data sharing platform.

6 ACKNOWLEDGMENT

THIS work was supported by in part by the National Key R&D Program of China (No.2017YFB0203703).

7 REFERENCES

M. Ballings M, Poel D V. (2012). Customer event history for churn prediction: How long is long enough?. *Expert Systems with Applications*, 39(18): 13517--13522

C. Baladron, M. Aguiar J, B. Carro, et al, (2012). *Framework for intelligent service adaptation to user's context in next generation networks. IEEE Communications Magazine*, 50(3): 18--25

L. Blunschi, J.P. Dittrich, O. R. Girard, et al, (2007). A dataspace odyssey: The iMeMex personal dataspace management system. *In CIDR*, 114--119

M. J. Cafarelia, A. Halevy, N. Khoussainova, (2009).*Data integration for the relational web. VLDB Endowment*,2(1): 1090—1101

H. Cai, X. Hu, Q. Ly, Q. Cao, (2009). A novel intelligent service selection algorithm and application for ubiquitous web services environment. *Expert Systems with Applications* 36(2): 2200--2212

X. Dong, A. Y. Halevy, (2005). A platform for personal information management and integration. *In: PhD Workshop of VLDB, Trondheim, Norway*, 26-30

X. Dong, A. Y. Halevy, J. Madhavan, (2005).Reference reconciliation in complex information spaces. *In: SIGMOD/PODS, Baltimore, USA*. 85-96

Y. L. Dong, D. R. Shen, Y. Kou, T. Z. Nie. (2009). Data Organization Model and Relationship Discovering Model in Dataspace. *Journal of Computer Research and Development*, 46(z2): 191-199

J. D. M. Guerrero, D. Marcelli, E. Soria-Olivas, et al. (2012). Self-Organising Maps: A new way to screen the level of satisfaction of dialysis patients. *Expert Systems with Applications*, 39(10): 8793—8798

D. Howe, M. Costanzo, P. Fey, et al, (2008). Big data: The future of biocuration. *Nature*, 455: 47—50

C. J. Hu., Y. Li., X. Cheng, Z. Y. Liu. (2016). A Virtual Dataspaces Model for large-scale materials scientific data access. *Future Generation Comp. Syst.*, 54: 456-468.

Z. Ives, C. Knoblock, S. Minton, et al. (2009). Interactive data integration through smart copy and paste. *In: CIDR, Asilomar, USA*, 299-310

S. R. Jeffery, M. J. Franklin, A. Y. Halevy, (2008). Pay-as-you-go user feedback for dataspace systems. *In: SIGMOD/PODS, Vancouver, Canada*, 847--860

Z. Jun, G. Zhenzhen. (2018). Intelligent Control for Integrated Guidance and Control based on Intelligent Characteristic Model. *Intelligent Automation and Soft Computing.* 1-11. 10.31209/2018.100000029.

G. Lin & W. Mingzhen. (2018). The Data Analyses of Vertical Storage Tank using Finite Element Soft Computing. *Intelligent Automation and Soft Computing.* 1-10. 10.31209/2018.100000031.

Y. Llsun, C. Chang, S. Vishal, W. Isaac, B. Bharat. (2018). Guest Editorial: Advances In Security and Privacy Technologies for Forthcoming Smart Systems, Services, Computing, and Networks. *Intelligent Automation and Soft Computing.* -1. 10.31209/2018.100000048.

C. Lynch, (2008). Big data: How do your data grow?. *Nature*, 455: 28-29.

J. Madhavan, S. R. Jeffery, S. Cohen, et al. (2007). Web-scale data integration: You can only afford to pay as you go. *In: Third Biennial Coneference on Innovative DataSystems Research (CIDR), Asilomar, USA*, 342—350

S. Pradhan, Towards a novel desktop search technique. (2007). *Database and Expert Systems Applications*, 4653: 192-201

A. D. Saema, X. Dong, A. Y. Halevy, (2009). Data modeling in dataspace support platforms. *Conceptual Modeling: Foundations and Applications*, 5600: 122--138

M. Singh, S. K. Jain. (2011). A Survey on Dataspace. *Advances in Network Security and Applications*, 196(3): 608--621

A. Toffler, (1980). The third wave: The classic study of tomorrow. USA: Bantam.

Y. Yin, F. Yu; Y. Xu,; L. Yu, J. Mu, (2017). Network Location-Aware Service Recommendation with Random Walk in Cyber-Physical Systems. *Sensors*, 17, 2059.

Y. Yin, S. Aihua, G. Min, Y. Xu, S. Wang, (2016). QoS Prediction for Web Service Recommendation with Network Location-Aware Neighbor Selection. *International Journal of Software Engineering and Knowledge Engineering*, 26(4): 611-632.

Y. Yin, W. Xu, Y. Xu, H. Li, L. Yu, (2017). Collaborative QoS Prediction for Mobile Service with Data Filtering and SlopeOne Model. *Mobile Information Systems:* 7356213:1-7356213:14

Y. Yin, Y. Xu, W. Xu, M. Gao, L. Yu, Y. Pei, (2017). Collaborative Service Selection via Ensemble Learning in Mixed Mobile Network Environments. *Entropy*, 19(7): 358

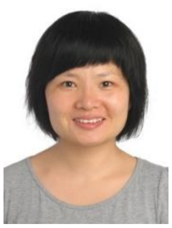
M. Zhong, M. C. Liu, C. Qian. (2008). Modeling heterogeneous data in dataspace. *In: Information Reuse and Integration (IRI), Las Vegas, USA*, 404--409

8 NOTES ON CONTRIBUTORS

Y. Li received the Ph.D. degree from the University of Science and Technology Beijing, China, in 2010, where from she is had also received her B.Sc., degree. Her research interests include data integration, semantic web, cloud computing and software engineering. Currently, she is the associate professor at University of Science and Technology Beijing, Beijing, China.

J. J. Li received the Ph.D. degree from Peking University, Beijing, China, in 2001. He is currently a Professor at the School of Information Engineering at the University of Science and Technology Beijing, China. His main research interests include parallel computing, parallel compilation technology, parallel software engineering, network storage system, data engineering and software engineering.

P. Shi received his B.E. in mathematics from Shandong University, China in 1999. Then he received his M.E. in computer application from Ocean University of China, China in 2002. He received his Ph.D. in computer science from Institute of Computing Technology, Chinese Academy of Science in 2007. His interested area includes online social network, high performance computing and data engineering.



X.N. Qin received her Master degree in computer science. She is currently a senior engineer, VP of the Dawning information Industry Co., Ltd, GM of the Server product department. She is engaged in research and design of server architecture, high performance servers and so on. She has presided over the R&D of eight socket server A950&I980, Blade Server TC6600 and owned many patents of invention.