6th International Young Scientists Conference in HPC and Simulation, YSC 2017, 1-3 November 2017, Kotka, Finland

Toolkit for intensive work with metadata in specialized information systems

Andrey Polyakov[a, *], Dmitry Kokovin[a], Alexey Poyda[a], Mikhail Zhizhin[b], Alexander Andreev[b], Alexander Govorov[b,c] and Viacheslav Ilyin[a,d]

[a]NRC Kurchatov Institute, Moscow, Russia
[b] Space Research Institute of the Russian Academy of Science, Moscow, Russia
[c] Moscow State University of Geodesy and Cartography, Moscow, Russia
[d] Moscow Institute of Physics and Technology (MIPT), Moscow, Russia

Abstract

We present a concept of the information system for intensive work with metadata in heterogeneous data store to modify dynamically both the data and the metadata models by user. This concept is illustrated by example of such information system for the project where the *digital registrar* is developing for the culture heritage of Old Russian paintings, namely Dionisius frescoes in the Cathedral of the Nativity of the Virgin at the Ferapontov Monastery painted in 1502.

Keywords: metadata, data model, information system, Dionisius frescoes, digital registrar

* Corresponding author.
 E-mail address: andrew@kiae.ru

1. Introduction

In various applications, key problem is the choice of information system to provide effective work with data. Now there are many IT solutions with wide spectrum of services to meet different types of application requests such as cost of the system and its technical support, flexibility, scalability, openness, DBMS choice etc. One of the most popular information system of this type is the CRM (Customer Relationship Management) ones [1] targeted on automation of working processes in companies and organizations. However, for small companies and organizations CRMs are too expensive. In these cases, work with data is reduced to data storage in file system on local discs, or in public cloud resources [2] such as GoogleDrive and OneDrive.

It is worth to note that decisive factors, such as data types, are fixed in existing information systems. It brings to the problems in case when data heterogeneity is changed during the life of the application project. Now dynamic modification of data models and types of connections between them becomes to be more and more important, as well developing functionality of the system and developing methods of data elaboration and analysis. Furthermore, the time dimension becomes to be important too, when data, coming in time, should be jointed in tracks according hand-picked data attribute. An illustrative example can be medical registers where data of medical investigations and analysis should be collected in the time tracks for each patient.

Note that for such data attributes as heterogeneity, links to other data (e.g. time tracks), data models binding to the elaboration and analysis methods, history of the data birth and changing, special terminological concept *metadata (data about data)* has been developed much last 10-15 years. One can mention *ontologies* [3] as one of the most productive developments in this direction. However, work with metadata has static character in existing IT solutions, without possibility of dynamical modification or formation of new data models. In following we will use the term *data* for input data and for derived data of the same type, and will use the term *metadata* for derived data describing *data* structure, details of the *data* provenance (history), links with other *data* or with *metadata* etc.

Thus, it is an open problem now to develop the toolkit to equip users with ability to organize the dynamical modification of data and metadata models in transparent and easy way.

Formulations above have a conception character but we will not develop here the corresponding theoretical constructions in deep. Instead we present key points of the realization of this conception for specific class of applications, where

- ✓ data are stored in file system while metadata can be placed in relational database to ensure high performance for search and filtering;
- ✓ users work with metadata during the preparation of data processing jobs while the actual data are used when the job is performed.

As a pilot application, we consider in this paper the information system under developing for Dionisius frescoes painted in 1502 on the walls in the Cathedral of the Nativity of the Virgin at the Ferapontov Monastery [4]. The Dionisius frescoes have been included in the UNESCO World Heritage list in 2000. In 80's years of last century there a unique project was undertaken on conservation and preservation of the Dionisius frescoes [5,6]. Then, in 00's years, special heating and climate control system was established in the Cathedral of the Nativity of the Virgin. As a result, the physical conditions of the frescoes are good and stable for now. It gives additional argument for creation specialized information system described above for this culture heritage. Important point is that such a system could (or even, should) be extended with data provided by future investigations of the physical condition of the frescoes. The information (data+metadata) stored and created in the system under elaboration will allow to snapshot current condition of the frescoes at high resolution and accuracy requested by specialists in the field of heritage conservation [7]. For these purposes the main source of the data is macro and micro photography in different spectral bands – UV (ultraviolet), visible and IR (infrared) lights. In following, we will use a term digital registrar for the above concept in application to the Ancient Culture Heritage, and a name SuBMiT (Service Based Metadata management sysTem) for the software implementation of this concept.

The paper is organized as follows. In Section 2 we give basic information about the object of culture heritage, wall paintings (frescoes) by Dionisius in the Cathedral of the Nativity of the Virgin at the Ferapontov Monastery, for which the *digital registrar* is under creating. In Section 3 an architecture is discussed for the SuBMiT toolkit designed taking into account specifics of the application chosen. In Section 4 the pilot programming implementation is discussing in