Development of the requirements template for the information support system in the context of developing new materials involving Big Data

Grechnikov F.V., Khaimovich A.I.

Samara State Aerospace University

Abstract. We consider a concept of databases for the numerical and physical experiments in the information management system used for new materials development. For this purpose, brief analysis of information systems and databases for large-scale experiments has been performed. Special features and general approach to the templates of the functional requirements are suggested.

Keywords: big data technology, the genome of materials innovation infrastructure materials, prediction of new materials, the pattern of requirements

Citation: Grechnikov F.V., Khaimovich A.I. Development of the requirements template for the information support system in the context of developing new materials involving Big Data. Proceedings of Information Technology and Nanotechnology (ITNT-2015), CEUR Workshop Proceedings, 2015; 1490: 364-375. DOI: 10.18287/1613-0073-2015-1490-364-375

One of the major issues tackled by the leading engineering departments and individual production engineers developing products for the aerospace branch with the use of new materials involves making all the strength calculations (structural analysis) with regard of such materials under consideration of mechanical, thermophysical and other features at the macro-level. At the micro-level though (the microstructure field) materials pertain to anisotropy of structure due to availability of heterogeneous phase composition, point-type and other localized defects, non-uniformity of crystal composition, foreign inclusions, etc. These localized defects as well as external micro-mechanical damages of the parts' surface give rise to crack formation and development, which in their turn lead to off-design cases and breakdowns of parts in the process of their operational use and maintenance. It is especially true with regard to composite materials with mechanical alloying, compounds of the 'binding-reinforcing fiber' type and others. Well-known are cases (e.g., at the AIRBUS and BOEING companies), when compound parts were withdrawn from manufacturing final products due to the mentioned reasons.

The material's life cycle consists of several stages:

1. development of the material's new composition;
2. optimization of its properties;
3. designing and developing products made of the material;
4. testing and certification;
5. production;
6. operation including repair and recovery;
7. disposal.

The innovations infrastructure in developing new materials is shown in Fig. 1 [1]. Designing of new materials requires creation of the existing materials' database: mathematical models of the materials, experimental data as to materials in form of photographic images and data based on vibroacoustic emission as well as digital data on the processes of generating the products' samples made of the respective materials.

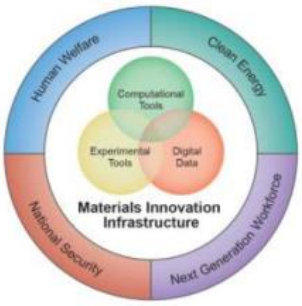


Fig. 1. – Innovations infrastructure in developing new materials [1]

Integration of the following methods is adapted in development of the materials database:

1. Tools of computer-based simulation (physical models of nano- and micro-level, numerical models and systems of simulation-type modeling of the macro-level continuum mechanics)
2. Toolware and experimental analysis methods
3. Digital data exchange and digital information storage formats

To enable works as to designing new material, it is necessary to create integrated medium in the BIG DATA system; the reason for that is that this database shall integrate sound data, experimental data photographs and existing materials mathematical models.

The input data (material database) for analysis are as below:

1. Results of simulation-type modeling of the structure – software modeling the material's structure (phase constitution, interrelationship, parts by volume, three-dimensional distribution and phases positioning)
2. Full-scale experiment results (standard tests) – microstructure pieces and material's submicrostructure obtained by means of light and electronic microscopy, results of the X-ray-structural analysis (describing material's structure and structural imperfections), and other analysis types. Final models as to the material's structure obtained as a result of the full-scale experiment data capture (development of special-purpose technique and software is required) linked to the models of P.1
3. Mechanical and other properties of the material (based on the experiment outcome, P. 2) – in strict compliance with the models of the structure.

The output information in the system of modeling new materials shall be viewed as the damage evaluation service based on the neural network. Principle of the service performance: on the basis of numerous selections of fragments of structures with defects as well as other parameters characterizing the material's macro- and micro-condition, the part's geometry parameters, operation conditions and calculated or experimentally obtained variants of damage or pathology development, training of the network is conducted. Flowsheet of the services' performance with regard to diagnostics and damage evaluation is presented in Fig. 2.

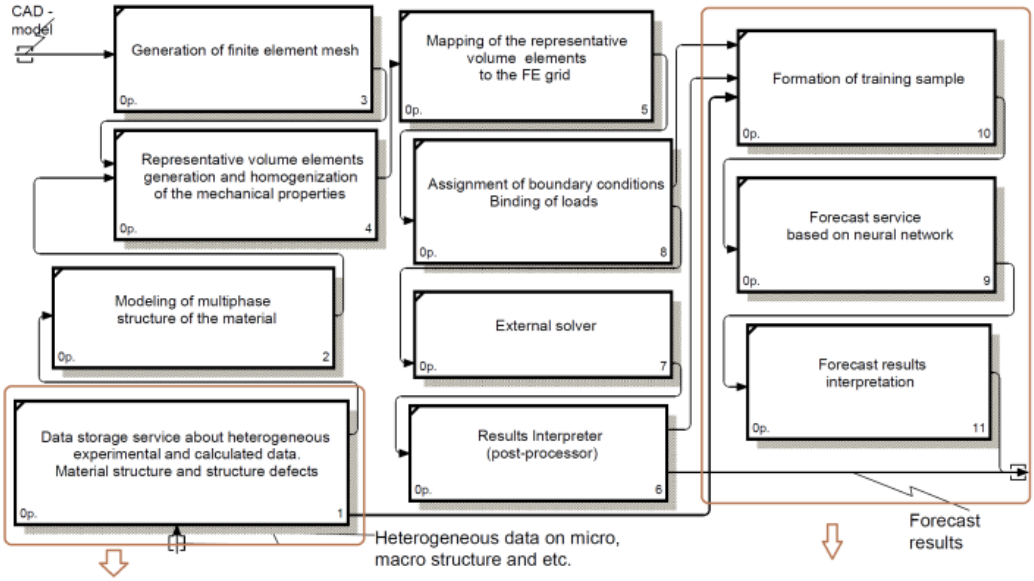


Fig. 2. – Flowsheet of the services performance with regard to diagnostics and damage evaluation. Arrows indicate the forecast service blocks.

The forecast is made in the following way: selected parameters are delivered at the input, which is similar to the elements of the training set of selected stuff. Selected parameters are obtained for the part material under testing, which in its turn has been selected from the industrial lot of the parts. For the required (possibly critical) working conditions, the grid generates a forecast of material's response.

In the event of unfavorable forecast, series of refined calculations and experiments are made, expert opinion is given as to ultimate solution and corrective actions are undertaken, which prevent failure of the product.

Training and clarification of the grid-related forecast is undertaken periodically by virtue of increase of the training set in size.

Some stages while creating new materials are implemented with the aid of hardware and special-purpose software. This being so, each stage can be implemented within special application package, which implies access to the complete package on the part of the material scientists. Moreover, scientific applications often require special skills for their installation, adjustment and startup, which the majority of researchers lack.

It is proposed to develop methods of producing cloud-based system of information support aimed at developing new materials, which would enable integration of the complete set of software tools into a unified medium and provide the researchers with the opportunity of operating it with the help of user-friendly web-interface having Internet access via web-browser. The web-interface as well as the computer modeling with the aid of application packages will be implemented in the cloud medium.

Start of computations in the cloud is proposed in the following way. In the cloud web-interface of the system, the user is expected to chose the system-supported application package from the list and assign the set of package- and task-specific input parameters (including the number of cores and RAM necessary for making relevant calculations) and in case of necessity – to indicate the file or directory with initial mega-data way, which is equally necessary to computations. After initiating the task start from the web-interface, the system will check availability of all the obligatory data and correctness of their input. In the event of successful check and availability of uncommitted resources, the system initiates startup of the corresponding number of virtual machines with preinstalled application software required for making calculations set by the user. After finalizing the calculations, the system will preserve results via the user-set way in the cloud storage.

The system based on the developed techniques will also enable storage of various information on properties of already known materials, experimental data linked to previously conducted tests as well as performance search by these data with the purpose of forecasting properties of new materials without undertaking any complementary full-scale experiments.

In the USA and Europe, this direction is marked by similar works within the framework of developing material genome under the use of the Big Data approach. In Russia, elaboration of methods with the aim of further creation on their basis a portal for researchers in the field of material science is just at the initial stage.

June 24, 2011, the US President announced the Materials Genome Initiative – in order to double the speed, at which development and production of new materials is

done. Acceleration of developing advanced materials is of fundamental importance for achieving global competitiveness (http://www.whitehouse.gov/sites/default /files/microsites/ostp/materials_genome_initiative-final.pdf).

New materials are of composite nature – they have multi-component and multi-phase structure: without adequate simulation modeling, information resource and data exchange development of new-generation materials (given the existing empirical approach) is either impossible of too time-consuming due to their complexity. The Materials Genome Initiative is aimed at promoting development of new materials; it accelerates the process and makes it more cost efficient. December 4, 2014, strategic plan of the Initiative was published (http://www.nist.gov/mpi/uploacl/MGI-StrateQicPlan-2014.pdf) NSTC (National Science and Technology Council). The subcommittee for genome includes NIST, Department of Energy, Department of Defense, NSF, NASA, NIH, USGS, DARPA in coordination with the infrastructure of Nanotechnologies Knowledge.

In much the same way that the Human Genome Project accelerated development of biological sciences owing to identification and decoding of basic construction blocks of human genetic code, the Materials Genome Initiative (MGI) will accelerate understanding of the material science fundamentals, providing information ensuring creation of new products and processes. MGI will require unprecedented level of cooperation of participants including government, industry, academic entities, professional communities, national laboratories, thus resulting in rebirth of the respective American industrial sectors. In order to integrate results of experiments, numerical methods (finite elements method, etc.), theoretical approach, the strategic plan envisages creation of a network of the MGI (Material Genome Initiative) resources – with the purpose of developing reliable techniques of precise simulation modeling, improving tools of performing and processing experiments' outcomes. Directional effect of these integrated data – discovery of new materials, development of analytical information for increasing value of the obtained experimental and computation data. Other objectives of the plan, linked to intensive data processing, include creation of means for materials' information-specific infrastructure implementation as well as development of the first rate techniques of maintaining proper storages of databases as per materials.

In order to start a dialogue within the framework of the MSE community (Engineering of Development and Study of Materials), the National Institute for Standards and Technologies (NIST, USA) conducted a workshop on materials' digital data in May 2012 under the aegis of MGI (Materials Genome Initiative). The workshop determined a number of issues that have to be settled in the process of creating data-specific strategy for materials; these include: schemes/ontologies of materials, standards of presenting data and metadata, data repositories/data achieves, data quality, incentives for joint use of data, intellectual systems and tools for data search [2].

The European Union is involved in elaborating standards of data exchange as for engineering materials within the framework of the European Committee for Standardization [3]. These standards mainly emphasize engineering materials for the aerospace branch. The European Commission finances activity of the expert group

called Integrated Computational Materials Engineering (ICMEg); the Group has been established as an entity for integrated materials computation engineering – with the purpose of developing standards and protocols required for supporting digital data exchange on materials [4].

In Russia, it is necessary to launch developments as to creating genome of materials under utilization of the Big Data technology. For this, it is necessary to form requirements with regard to large-volume digital data covering the stage of collecting and processing of experiments' outcomes as for mechanical properties studies (in dynamics, for example) or microstructure fragments, stage of the numerical simulation modeling of the process of materials' responses – with the view of further use of the collected experimental data for shaping an integrated model of materials-specific knowledge. The integrated knowledge model includes information required for concluding forecast related to the materials' response (including those with noncritical defects) – in the course of maintenance or in the technological process. Testing of the end-to-end information support and shaping the list of requirements as to materials' genome (for the damage evaluation case) could be performed as shown in Table 1.

Table 1. Application examples (use cases) of the materials information support services as per damage evaluation

Experimental data on pathology (defects, heterogeneities) in the material's structure	Simulation model of the material's response (computational model). Technology of identifying imperfections by indirect information	Response prediction model
Unstructured data storage on microstructure of material with micro-cracks	Modeling of crack development in CAE – system with preparing finite element pattern on the basis of selected micro-structure fragments with the aid of cloud service	Forecast (training) model of material's strength variation on the basis of neural networks
Fragment (sample) of vibroacoustic emission signal in the event of critical material load (e.g., in static test, deviation of the operational conditions and geometry from nominal ones (mechanical treatment, sheet pressworking, etc. for the know and new materials))	Spectral analysis technology as per emergencies identification on the basis of identifying emission parameters	Forecast (training) system of monitoring materials' responses in emergencies

Currently, there are precedents in Russia of global distributed information systems supporting experiments operating large data volumes. For illustrative purposes, one can mention the Central Information-Computer Complex of the Joint Institute for Nuclear Research (JINR) (in the town of Dubna), which develops as a multifunctional center for storing, processing and analyzing data and which is designed for providing a broad spectrum of opportunities to users on the basis of components integrated into it: grid-infrastructure of Tier-1 and Tier-2 – with the purpose of supporting experiments on LHC (ATLAS, ALICE, CMS, LHCb), FAIR (CBM, PANDA) and large-scale experiments; general-purpose computation cluster; cloud computation infrastructure; computation heterogenic cluster HybriLIT; training-and-research entity for distributed and parallel calculations [5].

Based on the experience of designing projects information support systems ATLAS, CMS of the JINR employees [6-10], it is possible to define a number of requirements concerning the databases conceptual structure with regard to information support of experiments within the genome of material project.

For full CMS databases' structure and hierarchy see Table 2, Appendix A [6].

Conclusion

High competition under the global market conditions requires continuous improvement of consumer properties in the output products', the major part of which depends on materials used in manufacturing. In this regard, rapid development of new materials with required properties is critically important; moreover, it is important often not just for individual companies but for branches of nation's economy. Development of methods of organizing interrelations of heterogeneous virtual services and applications in Big Data will enable a system of digital support of new materials development in Russia. The distributed system will make it possible to integrate on a uniform (from the uses' point of view) platform the whole set of diversified data required by the material scientists – with the purpose of forecasting defects development in the process of creating new materials.

Worthwhile mentioning is availability of commercial software in form of interfaced modules, designed for solving adjoint problems of the structural analysis, in which the composite material microstructure's influence on the developed structure is taken into account, e.g., Digimat, MultiMech. However, there is no available forecast service (e.g., solution on the neural networks basis). This enables rapid forecast solution on the basis of pre-conducted experiments and calculations with the aid of Big Data.

Recommendations on the information services assisting experiments have been validated and are more evident. For the dynamically resizable data, nested and user-defined arbitrary structure (multimedia, for example) NoSQL warehouses are preferable. Thus, the usage of the object model is recommended for distributed databases with a large number of complex relationships: cross-reference links, many-to-many relationships between objects.

A flexible object database can be a better choice for Configuration, but DB Conditions, Component and Event Data tag databases with slowly changing structures and simple relationship can use RDBMS for their implementation.

Appendix A

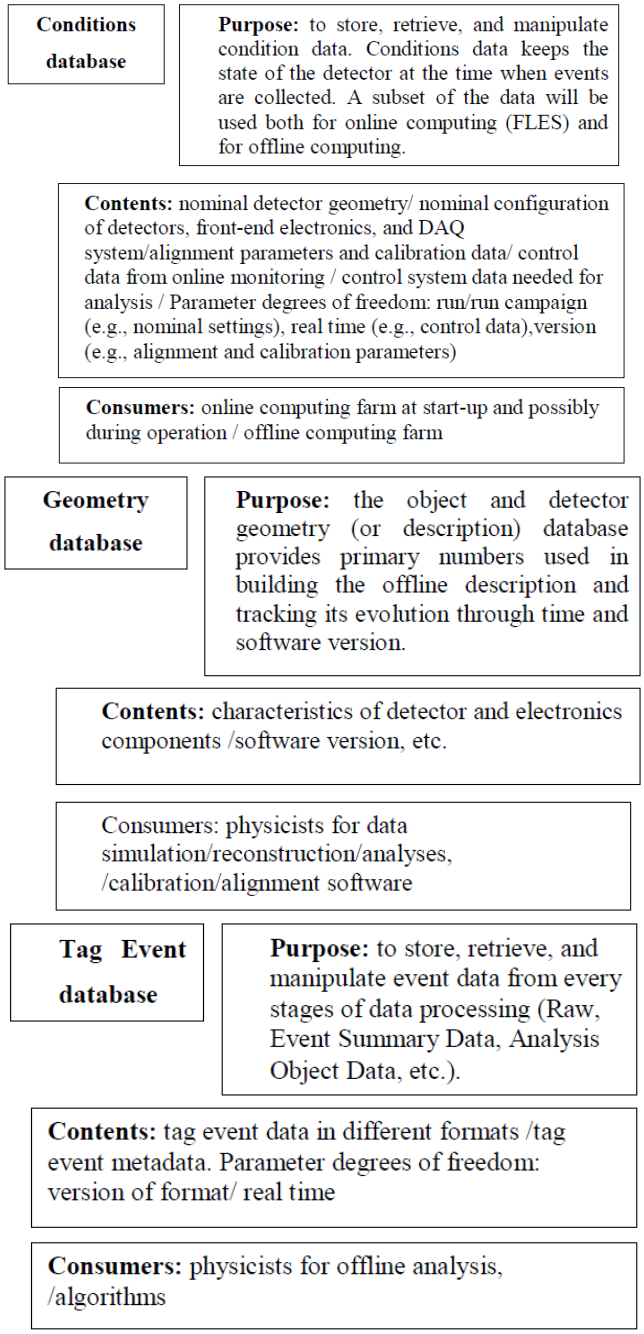
The Configuration DB will support only the data that is directly required to start and support the DAQ (Data AcQuisition) system in an efficient operating condition. This database should be elastic because not only data but the structures of data can vary. This means that the Configuration DB must be developed in close collaboration with the DAQ group.

Table 2. CMS databases structure

Database level	Database purpose
Equipment management database	Holds structured data about all detectors parts as equipment elements.
Construction database	Holds all information about relations between different equipment elements.
Conditions database	Holds all information about detector conditions (data on operating conditions).
Configuration database	Holds all information required to bring the detector in any running mode.

Typical use cases of DBs for physical experiments [6].

Configuration database	**Purpose:** to store the parameters describing the topology of the DAQ system, hardware and software components, and run parameters.
	Contents: nominal configuration of detectors, front-end electronics, and data acquisition/ hardware and software components, run parameters
	Consumers: setup of the configurations of detectors, front-end electronics/ data acquisition at run start through appropriate interfaces/ online computing farm at start-up and possibly during operation


Conditions database
Purpose: to store, retrieve, and manipulate condition data. Conditions data keeps the state of the detector at the time when events are collected. A subset of the data will be used both for online computing (FLES) and for offline computing.
Contents: nominal detector geometry/ nominal configuration of detectors, front-end electronics, and DAQ system/alignment parameters and calibration data/ control data from online monitoring / control system data needed for analysis / Parameter degrees of freedom: run/run campaign (e.g., nominal settings), real time (e.g., control data),version (e.g., alignment and calibration parameters)
Consumers: online computing farm at start-up and possibly during operation / offline computing farm
Geometry database
Purpose: the object and detector geometry (or description) database provides primary numbers used in building the offline description and tracking its evolution through time and software version.
Contents: characteristics of detector and electronics components /software version, etc.
Consumers: physicists for data simulation/reconstruction/analyses, /calibration/alignment software
Tag Event database
Purpose: to store, retrieve, and manipulate event data from every stages of data processing (Raw, Event Summary Data, Analysis Object Data, etc.).
Contents: tag event data in different formats /tag event metadata. Parameter degrees of freedom: version of format/ real time
Consumers: physicists for offline analysis, /algorithms

Component database	**Purpose:** manage properties of detector and electronics components obtained in a QA process after and during mass fabrication. This includes both mechanical/engineering production data and electronics data (calibrations, chip thresholds, etc.) which are generally kept completely separately.
Contents: tag event data in different formats /tag event metadata. Parameter degrees of freedom: version of format/ real time.	
Consumers: physicists for offline analysis /algorithms.	

The analysis of databases of physical experiments, such as ATLAS and CMS, shows that they have similar use cases and close structure of corresponding databases. Below the typical use cases for Configuration (Fig. 3), Conditions (Fig. 4), Component and Event (tag) (Fig. 5), Geometry (Fig. 6) databases is presents [6-10].

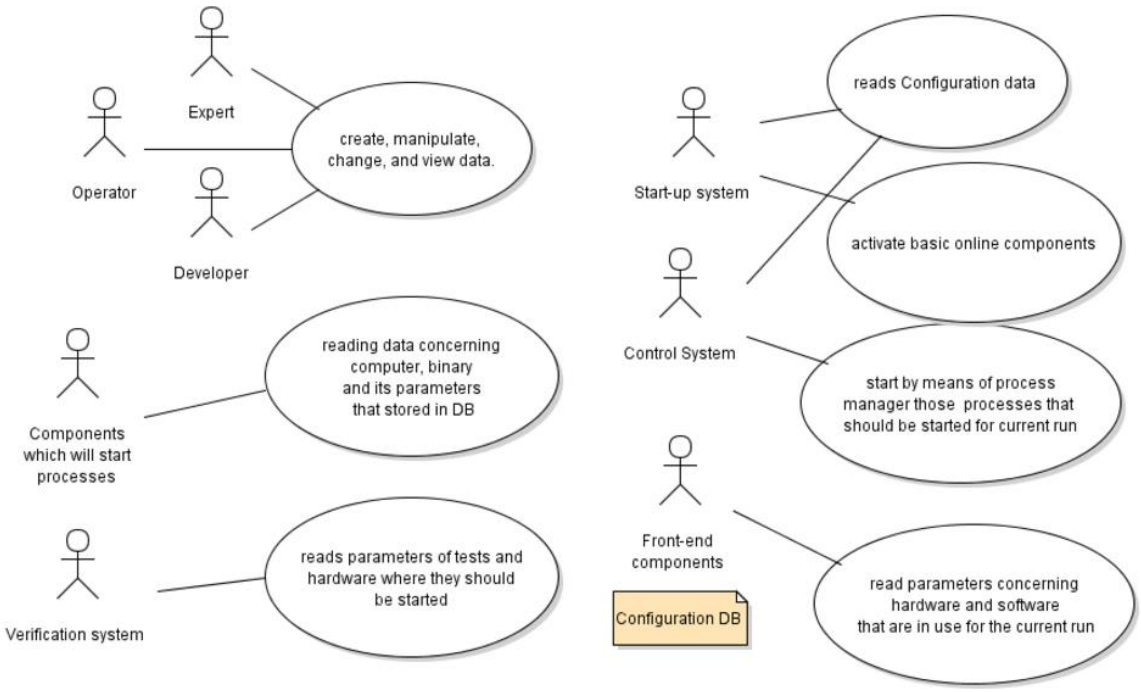


Fig. 3. – Use cases for Configuration DB

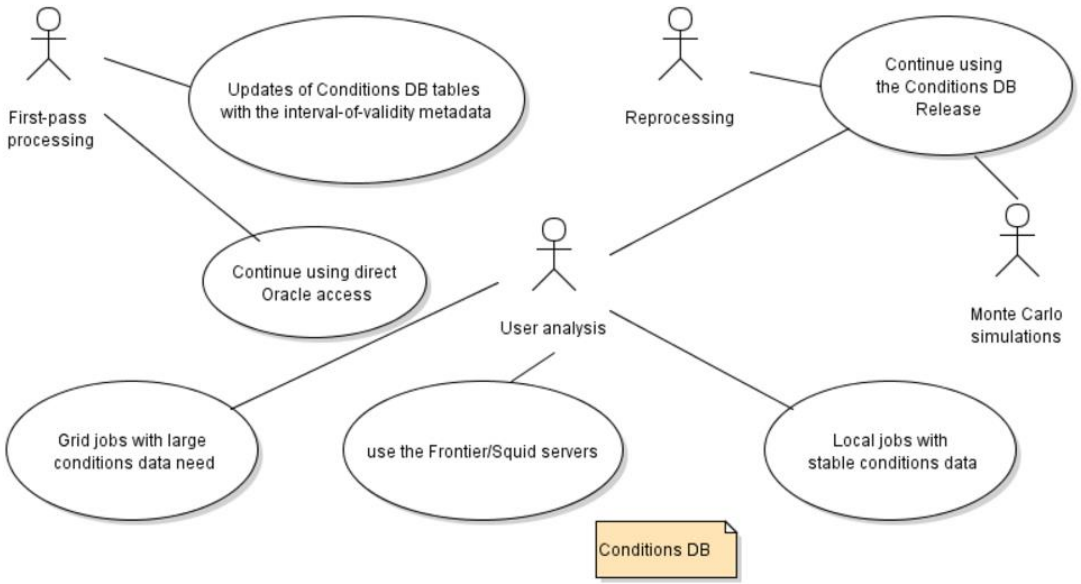


Fig. 4. – Use cases for Condition DB

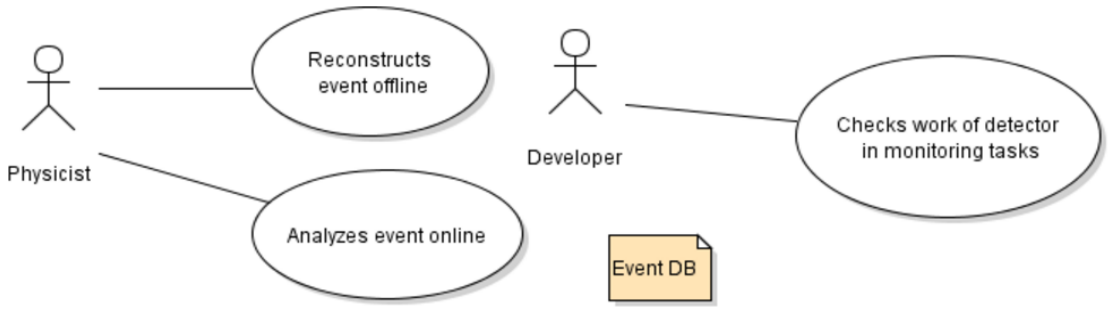


Fig. 5. – Use cases for Event DB

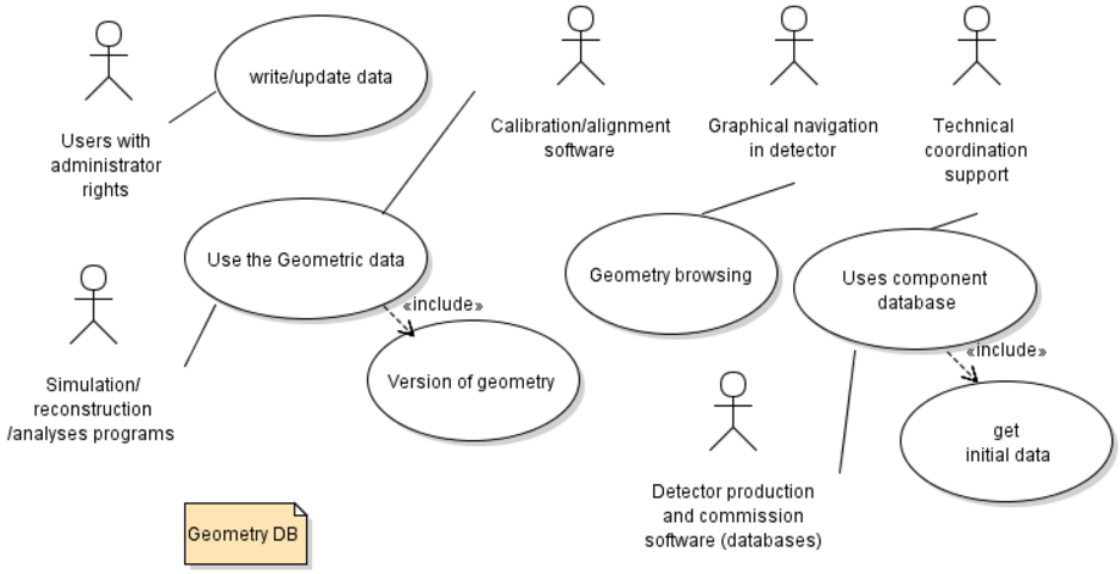


Fig. 6. – Use cases for Geometry DB

References

1. Materials Genome Initiative National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative http://ssau.ru/files/science/conferences/itnt2015/itnt_2015_46.pdf
2. **Ward CH, Warren JA, Hanisch RJ.** Making materials science and engineering data more valuable research products, 2014. http://www.immijournal.com/content/3/1/22
3. **Austin T, Bullough C, Gagliardi D, Leal D, Loveday M.** Prenormative research into standard messaging formats for engineering materials data. Int J Dig Curation, 2013; 8: 5-13. doi:10.2218/ijdc.v8i1.245.
4. **Schmitz GJ, Prahl U.** ICMEg, the integrated computational materials engineering expert group a new European coordination action. Integr Mater Manuf Innov, 2014; 3:2. doi:10.1186/2193.
5. Joint Institute for Nuclear Research, 2014. Annual Report, 2015; 14.
6. **Akishina EP et al.** Conceptual Considerations for CBM Databases. Communications of the Joint Institute for Nuclear Research, 2014. E10-2014-103: 25.
7. **Gallas EJ et al.** An integrated overview of metadata in ATLAS. CHEP 2009 Conference, Prague, Czech Republic, J March 23, 2009.
8. **Miotto GL, Aleksandrov I, Amorim A, Avolio G et. al.** Configuration and control of the ATLAS trigger and data acquisition. Nuclear Instruments and Methods in Physics Research Section, 2010, A, 623(1): 549:551.
9. **Almeida J, Dobson M, Kazarov A et. al.** The ATLAS DAQ system online configurations database service challenge. J. Phys. Conf. Ser., 2008; 119(2): 022004.
10. **Vaniachine AV, von der Schmitt JG.** Development, deployment and operations of ATLAS databases. J. Phys. Conf. Ser., 2008; 119(2): 072031.