КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ФИЗИКЕ

CONCEPT OF JINR CORPORATE INFORMATION SYSTEM

I. A. Filozova[1], *M. V. Bashashin, V. V. Korenkov, S. V. Kuniaev,*
G. Musulmanbekov, R. N. Semenov, G. V. Shestakova, T. A. Strizh,
P. V. Ustenko, T. N. Zaikina

Joint Institute for Nuclear Research, Dubna

The article presents the concept of JINR Corporate Information System (JINR CIS). Special attention is given to the information support of scientific researches — Current Research Information System as a part of the Corporate Information System. The objectives of such a system are focused on ensuring an effective implementation and research by using the modern information technology, computer technology and automation, creation, development and integration of digital resources in common conceptual framework. The project assumes continuous system development, introduction of new information technologies to ensure the technological system relevance.

PACS: 01.52.+r; 01.50.H-; 01.30.-y

INTRODUCTION

Unification of the institution information environment composing various heterogeneous resources is a serious integration problem. An insufficient integration of the institutional information resources is a consequence of various reasons, such as an operation on varied platforms, a scrappy automation because of phased automation of separate divisions, software development in different time periods and in various platforms, difficulties in providing the required information and technical documentation, a shortage of qualified staff, increased costs for support, operation, and development, etc. [1]. Integration problems are not limited to the software only — they concern with all IT-infrastructure of the institution. Corporate Information System may be a possible solution in such circumstances.

1. CONCEPT OF CORPORATE INFORMATION SYSTEM

There are several interpretations of the Corporate Information System. Organizationally, it reflects management ideology, combining business strategy and information technology, a characteristic of organization level. Technologically, it is an information processing tool covering major business processes and arranging the unified information space and realizing

---

[1]E-mail: fia@jinr.ru

a human–machine system to support the intellectual human activity. It should be providing an accumulation of experience and formalized knowledge, continuously developed, adaptive to the changing environmental conditions and new needs of the organization.

The concept of Corporate Information System includes an integrated approach to automation of all components of the organization's information technology as well. The main goal of the Corporate Information System for a scientific organization is a comprehensive automation of all activities for successful implementation of its mission — achievement of scientific results, satisfaction requirements of the scientific community. The objective of the Corporate Information System is a more efficient management of the resources: materials and equipment, financial, technological, human, intellectual, and informational.

A typical composition of the Corporate Information System may include a finance management system, management information system, electronic document management system, enterprise content management system, human resources management system, project management systems, computer-aided design systems, specialized user workstations, collaboration suites, software for implementing special problems, decision support system, Internet/Intranet environment.

Structurally, the Corporate Information System may be presented in the form of the management pyramid with two layers: the operative (bottom) and strategic (top) [2]. Information coming to the system input is about the basic resources (financial, material, human, informational) to be managed; the system output reflects the result of the main activities of the organization (Fig. 1). Primary data are structured, reduced, and filtered in workflow up in
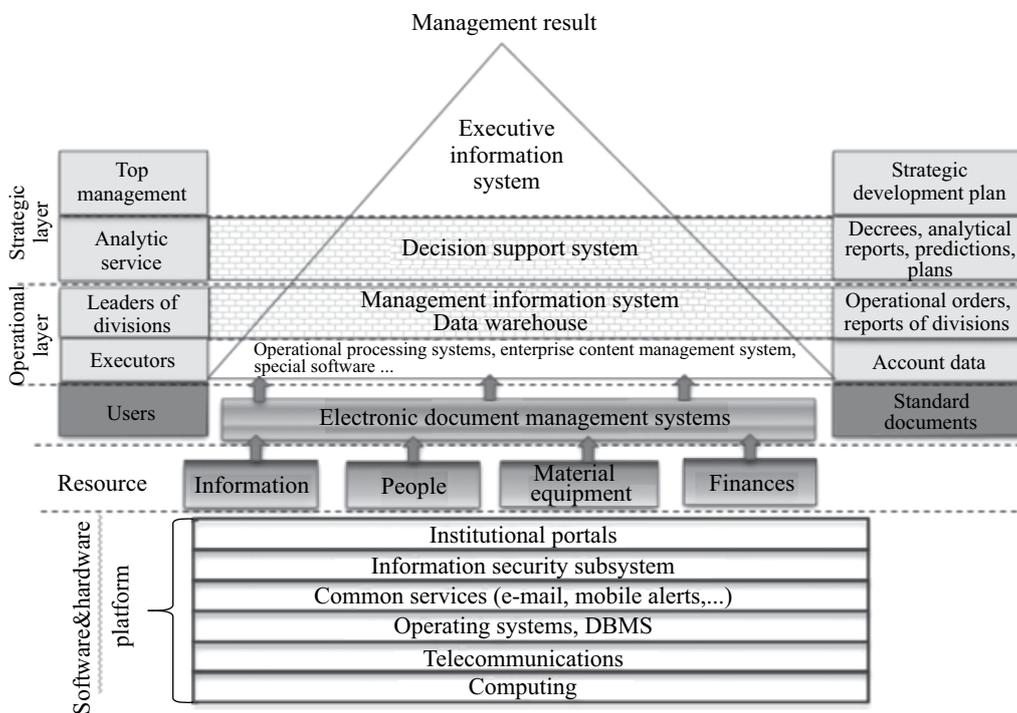

Fig. 1. Structure of Corporate Information System

the pyramid that provides reports for the top management, reflecting the most significant indicators to make strategic decisions on management and development. Thus, the Corporate Information System may be considered as a tool for information processing.

2. JINR CORPORATE INFORMATION SYSTEM. CURRENT STATUS

There are various information systems under operation at JINR, covering different functional areas: financial management, accounting and taxation, human resource and salary accounting, CRIS & OAR, project management, service center, business intelligence, material and equipment management. We plan to integrate them by developing the complex JINR CIS in next few years (Fig. 2). It should include General Informational Platform 1C, JINR Video Portal, JINR Document Server (JDS), JINR Events (IDC), Personal Information System (PIN), Monitoring and Accounting Systems, Activity Planning Tool Earned Management System (for experiment NICA). Some of them are functioning in test mode (JINR Project Management Service, JINR Help Desk), others are under development — GIS Applications Set, JINR Service Desk, Key Performance Indicators Dashboard (JINR KPI Dashboard), License Pool. Under the project, a single window system (Single Sign On, SSO) will be created with the aim to allow users to work with the subsystems of the JINR CIS. The SSO module in the institution grants authenticated users access to all subsystems, including client/server and web-based resources.

2.1. The Basis of Integration: CRIS & OAR. The component called Current Research Information System (CRIS) & Open Access Repository (OAR) is essential for a scientific organization. It reflects the results of its intellectual activity, and it is located in between operational and strategic layers of the Corporate Information System pyramid (Fig. 1).
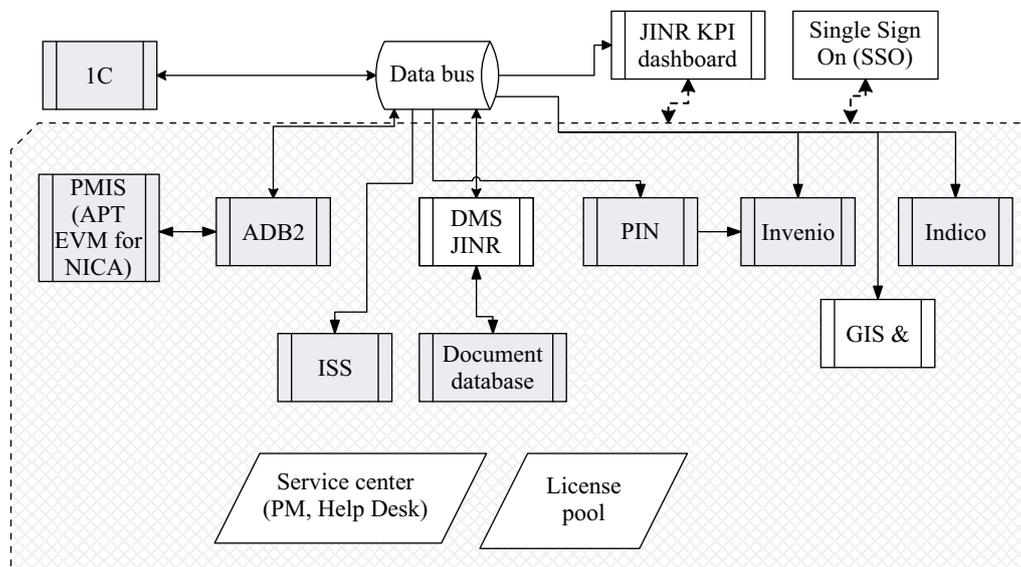

Fig. 2. Complex of JINR Corporate Information System

Namely, this subsystem is appropriate to provide integration of other informational systems, sharing and reusing the resources. Each subsystem of CRIS & OAR has both common features and particularities.

Commonalities of these systems include bibliographic information, affiliation, and project information. At the same time, CRIS is characterized as an administrative, comprehensive, integrative, person-centric, analytic system. CRIS manages financial and staff information, research and development (R&D) organization. It is aimed to record the R&D activity, to cover projects, people (expertise), organizational structure, R&D outputs, events, facilities, and equipment. On the other hand, OAR is an open-access, file-centric system, disseminating data via PMH-protocol. OAR is aimed to collect and preserve the R&D outputs, to provide services for collaboration members to manage and disseminate digital resources.

Since CRIS and OAR are intended to be the core of our corporate system, they should be combined for provision every possible services. Aggregation of the institutional information systems, namely, human resource, project, fund and other management systems is necessary for sharing, exchange and reusage of data. The task of efficient reusage of resources cannot be solved at the system level only. It requires permanent human curation of all resources: people (staff manager), projects (research project manager), finance (financial officer), materials and equipment (facility manager), bibliographic information (bibliography specialist, librarian, content manager, identity manager).

2.2. Authority Control. The key functional element of aggregation is the Authority Control subsystem of the Invenio software which can provide linkage and mutual control of the objects of Authority records with resources of the corporate system. Authority records are collections of entities representing names of authors and personalities, names of organizations, thematic terms, geographical names, uniform titles, projects, grants, etc. Organized by certain rules, they identify the objects and concepts uniquely. Authority control procedure provides overage variety of authorities (People, Institutes, Grants, Experiments, Projects, Journals, and so on), identifiers (DOI, ORCHID a.o.), linkages (multidimensional relations, vertical and horizontal linkages), history tracking (predecessors/successors linkages).

The data sources for Authority Control are institutional repositories, bibliographic databases, dictionaries, ontologies, authors' identification systems (international and institutional). Authority Control procedure provides accounting all options of names, e.g., different variants of an author's name. Furthermore, the usage of unique identifiers excludes ambiguities of author's names.

Authority Control procedure result is reliable and actual information reflecting the research and development organization, such as lists of publications, scientific reporting, bibliometrics and scientometrics reporting. Benefit of Authority Control is more qualitative, consistent data and minimizing the data input by end-users.

2.3. Authority Control Realization. There are three information management systems at JINR — JINR Document Server (JDS), Personal INformation System (PIN), Integrated Digital Conferencing System (IDC). The structures of each of them are built around different objects: a file (JDS), a person (PIN), an event (IDC) [3].

JDS functionality, provided by the software Invenio, covers all the aspects of modern digital library management. JDS was created and developed as an institutional repository of informational resources concerning the research activity at JINR: 1) the research and scientific-related documents of the following kinds: publications issued in coauthorship with JINR researchers; archive documents that describe all essential stages of the JINR research

activity; 2) documents providing informational support for scientific and technological research performed at JINR. JDS was implemented on Invenio platform (©CERN) possessing the module customizable structure. Invenio is a free software suite enabling one to create a digital library or document repository accessible on the web. Invenio complies with standards such as the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC21 as its underlying bibliographic format.

PIN is the information system collecting personal data on the JINR staff scientific activity including publications, scientific and organizational work, teaching, and innovation. It provides an authorized access to enter and edit information; preparation of documents for certification; formation of summary reports of departments and the institution as a whole. Since PIN data format differs from OAI-PMH, the data exchange channel between PIN and JDS has been developed that allows one to search the content of PIN using the JDS web-interface.

IDC is built on the top of the Indico platform (©CERN) which allows managing the entire process of creating scientific events (conferences, seminars, lectures) and automating many functions of the organizers during the life cycle of the conference. Indico is currently intensively used at JINR. Most of events held at JINR are scheduled through Indico, e.g., Programme Advisory Committees, Scientific Council, conferences, workshops, and seminars. Sections and group meetings are easily manageable, allowing the participants to submit materials and share them with others [5].

The Authority Control procedure became possible in release of the Invenio version 1.2.0 with corresponding functional modules implemented. At a solution of the task of aggregation the bibliographic formats MARC21 and MARCXML together with the functionality of the Authority Control subsystem can be chosen as base elements. MARC21 format for Authority Data includes: repeatable linking fields (4xx, 5xx); horizontal linking (subfield $w: $wa — predecessor, $wb — successor); vertical linking (subfield $w: $wt — parent); repeatable System Control Number (field 035); repeatable Standard Technical Report Number (field 027).

The Invenio module *BibAuthority* provides enriching of bibliographic data with data from authority records, reindexing of bibliographic records containing links to recently updated authority records, cross-referencing between MARC records (using $0 subfields) [4].

The estimate of *BibAuthority* functionality is currently performed, that involves: Authority Collections addition, Authority Records verification, curation processes providing, output formats configuration, search engine configuration for displaying the search results from other information systems via JDS web-interface. Test authorities collections *People, Institutes, Journals, Subjects, Themes & Projects, Experiments, Grants* are created and configured. Each entity or object is associated with a set of data entered in one or more authority records. Each bibliography is associated with the number of authorities included in the list of search elements to be controlled. Thus, most part of operations with content can be partially automated. One of the tasks is to increase the efficiency of the corporate system as much as possible by automation of creation and management of the authority collections. To solve this problem, at first the authority records are prepared and uploaded into JDS by packages in batch mode, then machine learning libraries for author disambiguation and for parsing unstructured pdf-files in structured XML documents are configured and tested. Availability of authority collections allows one to receive full and exact information from the system quickly. A user browses the authority records and linked resources by navigation or by search. For example, it is possible to view person's profile with lists of his publications, projects, experiments, grants, etc. Some typical ways of navigation are: institute → publication, author → publication,

experiment → publication, grant → author → publication, etc. Thus, repository is easy to use while ensuring high data quality.

CONCLUSIONS

Building of a corporate information system is a rather long and difficult process. The JINR Corporate Information System is developed with the use of layering principle. Specialized software, specific workflows, document management, supporting software are allocated in the separate layers. The advantage of this approach is the ability to make changes in the separate program components located in a single layer, without the need for radical alterations in the other layers, to provide a formal specification of the interfaces between the layers. Expected result of this concept realization is the manageability improvement of the organization, the quality of decisions.

REFERENCES

1. *Kusov A.* Problems of Introduction of Information Systems // Management of Economic Systems. 2011; http://uecs.ru/uecs-28-282011/item/411-2011-04-25-10-08-37 (in Russian).

2. *Korol I.* The Structure of the Corporate Information System // Informatization of Education. 2009. No. 1(54). P. 69–82 (in Russian).

3. *Borisovski V. F. et al.* Open Access Archive of Scientific Publications. JINR Document Server; *Borisovski V. F. et al.* // Proc. of XII Rus. Conf. RCDL'2010, Kazan, Russia, Oct. 13–17, 2010. Kazan: Kazan State Univ., 2010. P. 162–167 (in Russian).

4. A Digital Library Framework Invenio. http://invenio-software.org/.

5. Indico Project Site. http://indico-software.org.