Migration of a living archive to a software-defined storage

Olga Zhelenkova (1,2), Vladimir Vitkovskij (1,2), Vladimir Chernenkov (1), Vladimir Shergin(1), Tatyana Plyaskina (1)

1- SAO RAS, 2 – ITMO University

Abstract. *For long-term data preservation, scaling of hardware and archive contents in context of increasing volume of observations and digital collections, access to data and their interpretation we considered a prototype of a persistent warehouse based on iRODS free software system and designed a migration strategy of the archive to a new storage environment.*

Since the 80s of the last century, the acquisition of data obtained with the telescopes of the Special Astrophysical Observatory of the RAS -- the radio telescope RATAN-600 and the optical telescope BTA -- started to be performed in digital format. It served as an impetus to the development of the concept of an observational archive. Now in the general archive of the observatory are stored heterogeneous digital collections. Each collection includes data obtained using one device mounted on a telescope.

The archive system itself consists of three interconnected subsystems -- a cascade archiving, a warehouse and an information system (see Fig.1).

I. The **cascade archiving** includes a few levels where files accumulated for a certain time and then, if it is need, are transferred between levels or stayed for an persistent storing:
(level 1) a hard disk of the acquisition system computer, where files obtained during an observational program are accumulated;
(level 2) the file server dedicated directory, where the files are copied upon completion of the program;
(level 3) an offline storage. When the volume of data on the level (2) becomes sufficient for writing to an optical disk, the administrator prepares a disk image with observations, service files and observation logs and writes a CD/DVD disk;
(level 4) the test online storage locates on a storage system and contains both an unchangeable copy of all optical disks and the contents of the same disks, but already processed by program filters so that names and structure of directories and files were according to the rules of the archive. On the test server we carries out the development and testing of archive software, as well as testing disks for new keywords and other non-standard situations when upload database tables;
(level 5) the work online storage is on the dedicated server. It contains the same as on the previous level (4) and it is used for supporting of data requests.

II. The **archive warehouse** contains the offline storage and two online storages. The offline storage includes 300 CD / DVD discs and the same number of duplicates. Each online storage consists from 20 digital collections with raw data, 3 local archive with files which coordinate calibrated and corrected FITS headers, and 2 archives with list of celestial objects detected on the direct images in total with 2.5 mln files in 2TB volume.

III. The search **information system**. The test information system is based on SciLinux + PostgreSQL on the test server. The information system supports three categories of users with a different set of privileges and the standard set of web requests. The work information system has the same configuration and possibilities and locates on the other server. They duplicate each other.

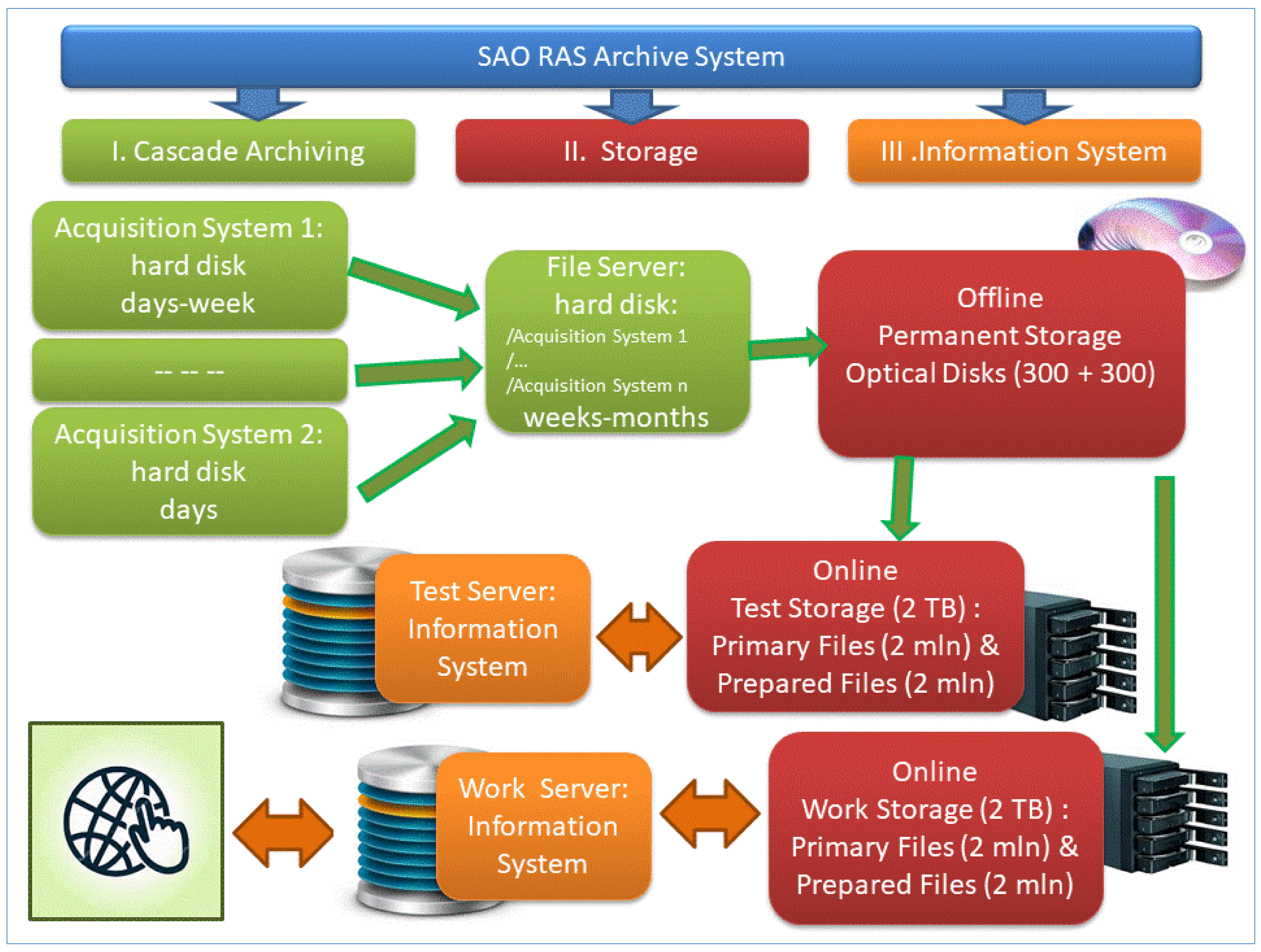

Fig.1. The archive system schema.

We could designate the following functionality of a system providing a permanent archive:

- updating existing resources, what includes replenishing collections and correcting data mistakes;
- bringing specifications of collections to a unified data model;
- the addition of new digital collections;
- primary data processing;
- web access to data and compliance with the standards of the International Virtual Observatory Allience (http://www.ivoa.org);
- compliance with the Reference Model for an Open Archival Information System and possibly other archival standards;
- making plan of partial or full data migrations, including the transformation to other storage formats.

A persistent archive is the save of digital information for decades, and even better -- without a time limit. This requires to store with files of the digital collection a description of its organization and metadata that allow information to be interpreted. Despite the fact that the reliability of devices is growing, and errors that occur during reading / writing can be controlled programmatically, but the life of the read / write devices is estimated at 5--10 years, and the life cycle of relevant software is estimated at 5--7 years, that less then life of the machine-readable data carrier. So with long-term storage it is necessary to monitor the status of the archive storage and periodically transfer data to new media. Software-defined storage (SDS) is a rapidly evolving technology where software manages storage resources, which allows data to be located on non-specialized equipment under control of general-purpose operating systems like Linux, Windows.

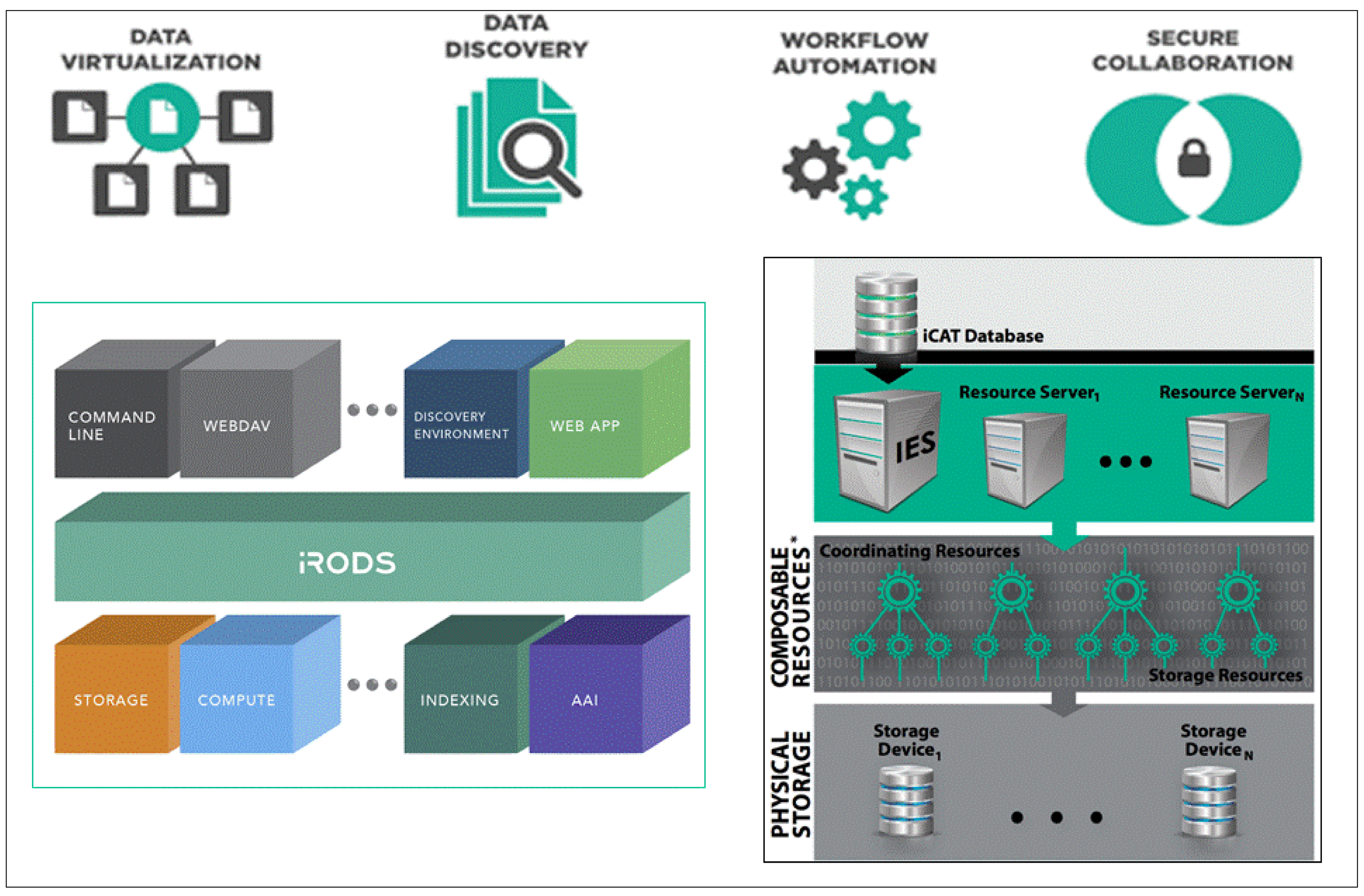

Fig.2. The iRODS system possibilities.

SDS-based storage also allows you to organize a hierarchy of stored data with the distribution of files on servers between hard drives with different characteristics. This approach is appropriate when it is necessary to make changes in the IT-infrastructure, and also if there is a lot of equipment in it that does not functionally fully meet the challenges, but is suitable for software-defined storage. In prospects considering data volume increasing and existing hardware and, we decided to organize a software-defined storage for our archive based on the iRODS (integrated Rule-Oriented Data System, (http://irods.org), see Fig.2. The system has functionalities that as we think it meets our needs.

Use Case. The observatory has three geographically distributed data centers – optical telescopes, the radio telescope and the headquarter. These autonomous centers, can be deployed into the three connected iRODS zones. The archiving scheme used in the observatory is can be better reflected in the architecture, which consists of several levels of resources. At the lowest level, observation data is accumulated on computers of acquisition systems (storage resources), the second level is the telescope file server (coordinating storage resource / resource), and the third level (coordinating storage / storage resource) is the online storage of the archive. Since the observations obtained with telescopes are accumulated on computers of different collecting systems, a hierarchy of several storage resources and composable resources is required. Since the centers exist autonomously, there may be situations when one of them is not accessible due to a power outage or an accident on the telecommunication line. In such cases, a replication strategy is also needed to support data availability. The introduction of iRODS capabilities into the archive system is not rational to start in full in the scheme that arises from the existing scheme of data flows from collection systems to permanent storage, but in stages – step by step. And we begin with the introduction of the system in the maintenance of online storage, where already in the existing information system there is information about the location of files and their meta descriptions.

Acknowledgements. *This project was supported by a grant from the Russian Foundation for Basic Research No.17-07-01367.*