SIPAD-NG: architecture and methodology of customization for a Data or a Mission Center

Claire Pottier [(1)], Silvia Enache [(2)], Dominique Heulet [(3)],

[(1)] CNES
18 avenue Edouard Belin 31401 Toulouse Cedex 9, France
EMail: claire.pottier@cnes.fr

[(2)] CNES
18 avenue Edouard Belin 31401 Toulouse Cedex 9, France
EMail: silvia.enache@cnes.fr

[(3)] CNES
18 avenue Edouard Belin 31401 Toulouse Cedex 9, France
EMail: dominique.heulet@cnes.fr

ABSTRACT

SIPAD-NG (“Système d’Information, de Préservation et d’Accès aux Données – Nouvelle Génération” – “Information System for Data Preservation and Access – New Generation”) was designed to access data through a catalog, whatever their scientific context is.
At present, SIPAD-NG is used by Data and Mission Centers covering various fields such as plasma physics, altimetry, clouds and aerosols, material physics.
SIPAD-NG system was presented during previous meetings (PV2004, PV2007 and PV2009).
Recently, it was equipped with a new functionality of data acquisition. Therefore, SIPAD-NG is now implementing the functions “ingest", "data management" and "access" of the OAIS model.
A methodology was developed around SIPAD-NG to allow its customization for a Data Center or a Mission Center.

SIPAD-NG FUNCTIONAL ARCHITECTURE

SIPAD-NG functional architecture is based on the OAIS model, as shown in figure 1.

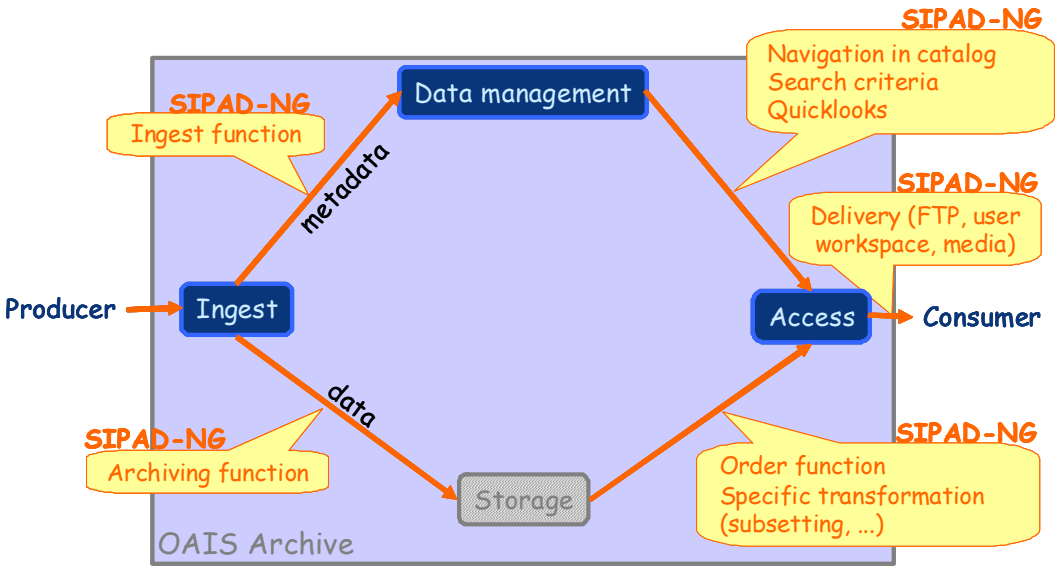

Figure 1: SIPAD-NG functional architecture

SIPAD-NG fulfils the functions from producer to consumer through:

- getting data from producer and transfer them to a storage system: it may be a disk or, at CNES, a dedicated service called STAF which assumes the long term conservation of data files,
- ingestion of metadata in a catalog: these metadata are XML files that contain all the information needed to describe and find the stored products,
- consultation of this catalog by navigation in a data tree, search criteria, or quicklooks,
- ordering of selected data and maybe application of specific transformations,
- delivery of these data to the end user, by FTP or on a user workspace or on a media (DVD for example).

SYSTEM ARCHITECTURE

SIPAD-NG system architecture is showed in figure 2.

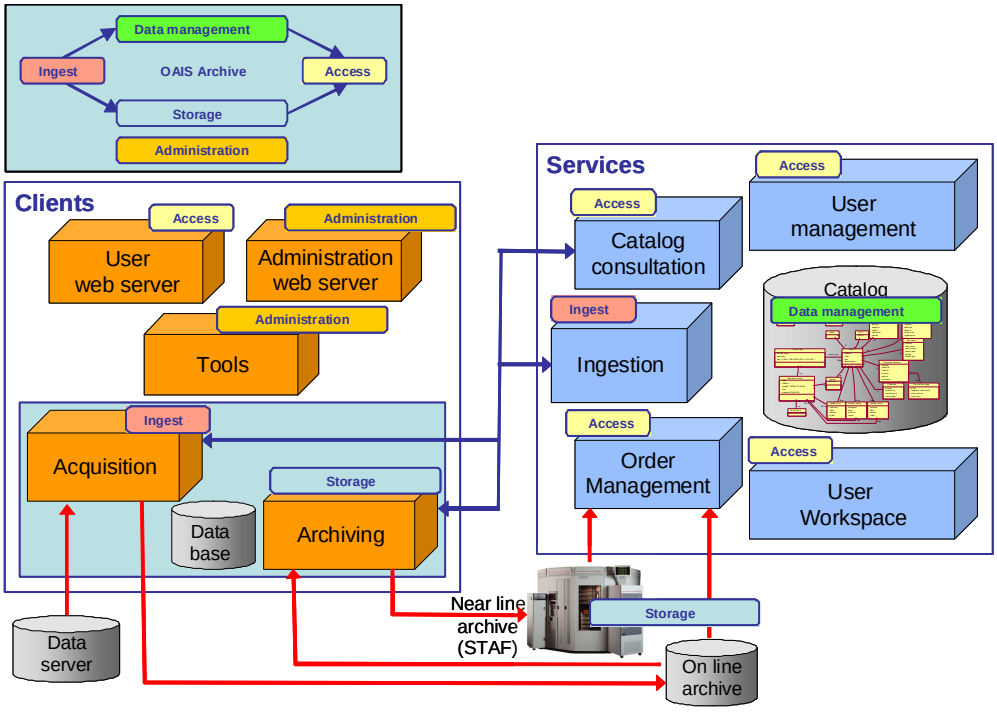

Figure 2: SIPAD-NG architecture

One **instance** of SIPAD-NG consists of six basic services and an administration web server. The basic services implement functions which can be found in any data access system:

- **consultation** is used to search for data by navigating in data graphs, by browsing quicklooks and by applying selection criteria,
- **user management** is used to centralize all information about users' profiles, access rights and quotas,
- **order management** is used to extract data from the archives and to possibly process them with added value services before delivering them to users via network or media,
- **user workspace** offers users a workspace in which they can retrieve data,
- **ingestion** can be used to acquire or update information on the SIPAD-NG catalog, as well as information related to the data (browse products, documents),
- **administration and supervision** (not showed in figure 2) is used to manage the system and to ensure that it is operating correctly. An instance provides also an administration web server.

One SIPAD-NG instance can manage one or several **projects**. A project consists in:

- A **catalog**: this is an ORACLE database (used by the basic services) which references all of the information on data, users and added value services. Note that project data are not stored by SIPAD-NG; they can be accessed in an archive (which may be a near-line system or simply an on-line disk space); the project catalog memorizes metadata and the location of the data in the archive,
- One or several **client applications** which use the programming interface provided by the services to consult the project catalog and to order data. Important client applications are the **user web server**, several **administration tools** and the **acquisition and archiving** client. This client is composed of a specific ORACLE database and two applications. The **acquisition** application retrieves data files from a data server, transfers them to the on line archive and references them in the SIPAD catalog. The **archiving** application creates TAR files from the online data, transfers them to the near line archiving facility (STAF) and updates the SIPAD catalog. Data files are then deleted from the online archive after a retention period.

METHODOLOGY

Figure 3 shows the methodology used to configure SIPAD-NG to be used by a Data or a Mission Center.

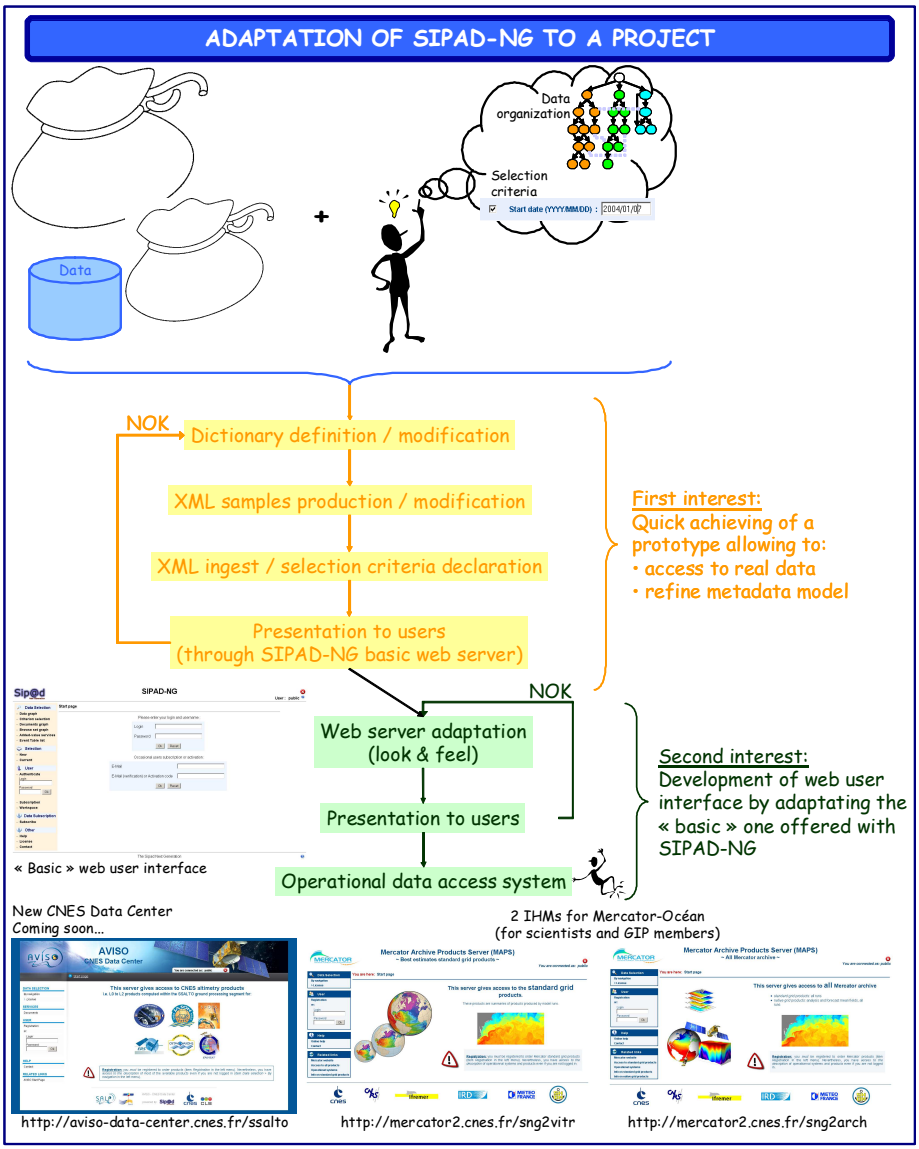

Figure 3: SIPAD-NG configuration

The first step consists in the definition of the **metadata dictionary**. A dictionary is an XML schema to describe datasets and data objects that are to be referenced in the catalog. Important inputs for this step are:

- The way you want to organize the navigation though the catalog,
- The access criteria that are to be used to search for data.

The second step consists in developing a specific web user interface, by adapting the 'basic-looking' web user interface provided by the SIPAD-NG. This interface is easily adaptable by modifying layout, menu, header, footer, CSS file and texts.

These two steps need iterations with the Data or Mission Center teams. With very little software development (only the adaptation of the web user interface may need a development), these steps lead to a prototype of the data access system.

CONCLUSION

SIPAD-NG is currently involved in several thematic data centers:

- Plasmas physics with CDPP (CNES/CNRS "Centre des Données de la Physique des Plasmas"),
- Material behavior in micro-gravity environment with DECLIC (CNES/NASA instrument installed on the International Space Station),
- Clouds and aerosols with ICARE thematic Data Center,
- Altimetry with the CNES AVISO Data Center.

New implementations are foreseen for:

- Ocean salinity and soil moisture with SMOS satellite CNES/IFREMER Mission Center,
- Simulation data produced by CNES Launcher Directorate.

Experience feedback shows that SIPAD-NG is very powerful to quickly develop a prototype of a catalog access system adapted to the needs of a Data or Mission Center. In operation, SIPAD-NG is able to archive, reference and make accessible large amounts of data (more than one million entries in the catalog). Difficulties are due to migrations from existing catalogs to new ones, when former catalogs do not use SIPAD-NG. But it is a general problem. We hope that choices made for SIPAD-NG (especially the representation of all metadata in XML format) will facilitate migrations of SIPAD-NG catalogs to future systems.