1 INTRODUCTION

1.1. Scope

The IODE Ocean Data Portal (ODP) delivers a standards-based infrastructure for integration of marine data and services submitted by the NODCs, data centers (institutions) of IOC and other programmes (ODP participants), as well as for data discovery, evaluation (through visualization and metadata review) and access to them. The ODP participants provide the operations of the global, regional/specialized, technical and national nodes.

The ODP v2 toolkit is available to set up the ODP node by participants based upon the designated participant role and functionality required. The Data Provider is a component of the toolkit for submission of the data and services (resources) to the ODP distributed marine data system (DMDS) and their management. The Data Provider is used by the ODP nodes of all categories. The ODP participant can use the Data Provider, installed at its own server or available at another ODP node.

1.2 Overview of the document

The document includes the following sections:

Section 1 - Introduction

Section 2 – The Data Provider functions, and its hardware, software and other technical requirements.

Section 3 – A set of instructions to be followed when installing the software in a Data Center.

Section 4 – A guide to the use of the User Interface for technical support of the Data Provider in a Data Center.

Section 5 – A set of instruction what maintenance should be provided in a Data Center.

1.3 Terms and definitions

Data set (local) - set of data (or metadata) being served by a Data Center. It is characterized by a specific data model, data coding system, data storage system and format, and any conditions placed on data access.

Distributed Data Source System – set of data sources in Data Centers operating under the ODP technology.

Data Source - combination of the Data Provider (software component) and local data sets (information component). One Data Provider installed in a Data Center should interface with all local data sets of the Center.

Integration Server – ODP technology software performing Distributed Data System (DDS) data sources management and reception and execution of requests from external software applications addressed to the DDS data sources.

Resources - local data sets representing DSS data sources. A resource (single) is a local data set or part of it. A resource has a unique identifier and may consist of structural data units (resource instances). The content and structure of a resource are recorded in the resource description and discovery metadata of the DSS.
The Integration Server performs periodic harvesting of the DDS resource discovery metadata and uses them to access distributed data sources on request from the portal of the system. On request from the Integration Server, Data Provider performs data sampling out of the local data set by request criteria, data unification and presentation of sampled data in the form of a transport data file.

**Code** - names of objects and values used for unambiguous understanding of data in information systems. Code is unambiguously defined by a coding system (ISO, IOC, WIS, etc).

**System codes** – codes of resource description elements (data accessibility, spatial presentation, storage system, etc). System codes are maintained as a controlled list of codes and used for description of DDS resources.

**System dictionaries** – codes and structured descriptions of objects (organizations, ships, buoys, satellites, etc.). System dictionaries are used for unification of data coding by mapping local codes into system codes.

**Data element** – atomic portion of data with specific properties (code, name, units of measurements, data type, etc). Data elements can be:

- metadata (information on data such as identifiers of country, organization, platform; time period, spatial coverage, etc.
- values of marine environment parameters such as sea surface or subsurface ocean temperature and salinity.

Local data elements compose local data sets in Data Centers. System data elements are data elements managed by the DDS. The Data Provider performs mapping of local data elements into system data elements assuring data unification.

**Data exchange protocol** – agreements regulating data exchange between the Data Provider and Integration Server. Data exchange protocol consists of request-message, response-message and transport data file.

**Transport data file** – data set provided by a data source in response to a request from the Integration Server addressed to the local data system

1.4 Contact information

Sergey Sukhonosov, tel.: +7(48439) 74194, fax: (499) 795 22 25, serg@meteo.ru

Sergey Belov, tel.: +7 (48439) 74194, fax: (499) 795 22 25, belov@meteo.ru

Kristina Bulgakova, tel.: +7 (48439) 74194, fax: (499) 7952225, christy@meteo.ru
2 DATA PROVIDER OVERVIEW

2.1 Purpose

The Data Provider software is an ODP technology component responsible for integration of information resources and performing the following functions of the ODP data source:

- connection to local data sets maintained by the Data Center;
- generation and updating of description of resources in compliance with the current status of the local data sets;
- access to resources, with any data sampling based on the criteria of the request from the Integration Server, structural and code unification of data, and transport data file formation as a response to the request;
- monitoring of the Data Provider software and registered resources performance on the basis of the current status information generated by the software itself.

The Data Provider provides interaction with the following types of data storage systems:

- DBMS database;
- structured data files in “flat” (non-hierarchical) data formats;
- object data files: electronic documents, images, structured data files in formats, which are not handled by the ODP technology software;
- Web-applications performing an access to the local data sets;

2.2 Components

The Data Provider distribution package is represented with 2 options – distribution as fully-functional virtual machine image and standalone distribution.

Virtual machine (VM) image has following contents:

- Linux CentOS 6 X_64 operating system
- JDK 1.7
- JBoss 4.2.3 application server
- Data Provider v.2, including main module, configurations, connectivity libraries for following data bases - MySQL, PostgreSQL, Oracle ODBC, Oracle 9i, Oracle 12g, web services for communication with upper-level node, JOSSO client configurations and libraries

Standalone distribution includes:

- J2SDK v.1.7;
2.3 Use requirements

2.3.1 Computing equipment

The Data Provider should be installed on a separate server with no less than 1 GHz processor and 1 GB of memory.

2.3.2 Software

The Data Provider is operated under Linux, Unix-based and Windows operating systems and uses the following general software:

- J2SDK v. 1.7;
- JBoss Application Server v. 4.2.3.GA.

All of the above software components are available in the Data Provider distribution package.

The Data Provider only receives requests from and exchanges messages with the Integration Server and Service Bus. With this in mind the rules for the firewall should be spelled out to allow HTTP incoming and outgoing connections through port 80 for the Integration Server IP address and port 18080 for the Service Bus address.

Where the JBoss application server is installed, it is necessary to have no less than 1 GB of free disk space. This space is used to maintain log-files and generate temporary files inside the application server.

If any other applications are installed on the server where the Data Provider operates then they should not occupy ports and names of the above application servers and Web servers. In case other or similar software components and libraries are installed, they should not cause conflicts when they are used jointly or when their versions are incompatible.

2.3.3 Interaction with the local data system

The Data Provider should have access to the local data sets, presented as the DDS resources.

2.3.3.1 Data managed in a DBMS

The Data Provider should have access to the local databases by the IP address in the local network of a Data Center, database name, password and other access tools specified during resource registration. The above access tools are not published and held in the Data Provider configuration files.
It is recommended to provide the Data Provider with the quickest possible access to the local databases.

If databases are located on a dedicated (external) database server it is necessary to arrange updating (replication) of the database from the main server (servers) of a Data Center, where data sets are maintained with the help of thematic technologies of the Center.

2.3.3.2 Structured data files

It is necessary to provide the Data Provider with access to the sets of structured (nonhierarchical) ASCII data files, which are handled in a way similar to databases.

In the current version of the Data Provider structured data files should be placed (updated on a regular basis) to the appropriate hard disk partition of the PC (or other host computer), where the Data Provider is installed.

The path to the data file directory is specified in the process of resource registration. The above information is not published and held in the Data Provider configuration files.

2.3.3.3 Object data files and Web-applications

It is necessary to provide the Data Provider with access to the sets of object data files (images, documents, software applications, etc.) or web-applications by the URI address or full path to files from the PC (or other host computer) for the Data Provider. The above information is not published and held in the Data Provider configuration files.

The Data Provider copies object data files to the appropriate hard disk partition.

2.3.4 Network communication with the Integration Server

2.3.4.1. Communication process

The Integration Server checks the availability of the DDS data sources, synchronizes system codes, elements and dictionaries, collects resource descriptions from Data Providers, receives an XML-request from external applications, identifies Data Providers that should be involved, and distributes elementary requests among Data Providers.

In response to an XML request the Data Provider sends an XML message to the Integration Server containing a reference to the transport (objective) file. If necessary the Integration Server independently extracts files by the references given in the XML message and sends both the response and data file or just the response to the external application.

2.3.4.2 Ports
HTTP incoming and outgoing connections should if possible be allowed through port 80 for the Integration Server IP address and port 18080 for the Service Bus address. 80 is a default port for the JBoss application server in Data Provider distribution package. If necessary however, it is possible to choose another port.

2.3.4.3 Security

The Data Provider receives requests only from the registered Integration Server and call web-services of Service Bus. The software does not allow direct and uncontrolled communication between the Data Provider in one Center and Data Providers in other Centers or between the Data Provider remote applications. To ensure security the IP address of the source, from which a request to the Data Provider arrived, is checked. If the IP address does not agree with the specified IP address of the Integration Server, a request is not handled and the Data Provider sends in response an HTTP status: 403 Forbidden. This allows data transmission to be protected from unauthorized requests.
2.3.5 Institutional support

The Data Provider should be operated by an information technology specialist of a Data Center, who is responsible for:

- installation of the Data Provider;
- maintaining the Data Provider operations with the help of the Administrator User Interface;
- interaction with the developers on the issues of Data Provider software upgrades, on alarm conditions, and when help from the developers is needed.
3 INSTALLATION

3.1 Data Provider installation – standalone distribution.

Installation process consists of the following steps:
− downloading installation package to the server;
− fine-tuning configuration files.

Installation package contains following components:
− jboss-4.2.3.GA – application server JBoss 4.2.3.GA (preconfigured for Data Provider needs);
− dpms.sar-2.0-<Data Centre ID>-<release date>.zip – web module with settings for current ODP Data Centre;
− dpms-2.0-<Data Centre ID>-<release date>.zip – directory with configuration files. It’s actual for all ODP data centres;
− init_soi_gen.sh – script for automated web-service stubs generation. This is mandatory for sending signals to Service Bus about new metadata and data;
− dpdb.ear, dpdb-ds.xml – database connection module

Identification information about ODP Data Centres is shown in table 3.1.

Table 3.1 – ODP Data Centres identification information

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-NODC</td>
<td>US National Oceanographic Data Center(NODC)</td>
</tr>
<tr>
<td>NODC</td>
<td>National Oceanographic Data Centre,Russia</td>
</tr>
<tr>
<td>ISDM</td>
<td>Integrated Science Data Management</td>
</tr>
<tr>
<td>IBSS</td>
<td>Institute of Biology of the Southern Seas National Academy of Sciences of Ukraine</td>
</tr>
<tr>
<td>MHI</td>
<td>Marine Hydrophysical Institute National Academy of Sciences of Ukraine</td>
</tr>
<tr>
<td>NMDIS</td>
<td>National Marine Data and Information Service</td>
</tr>
<tr>
<td>BGODC</td>
<td>Bulgarian National Oceanographic Data Centre</td>
</tr>
<tr>
<td>MIRC</td>
<td>Marine Information Research Center</td>
</tr>
<tr>
<td>GDNA</td>
<td>Tbilisi State University, Georgian DNA</td>
</tr>
<tr>
<td>NODEC/NIMRD</td>
<td>National Institute for Marine Research and Development</td>
</tr>
<tr>
<td>METOFFICE</td>
<td>Met Office</td>
</tr>
<tr>
<td>IMOS</td>
<td>Integrated Marine Observing System</td>
</tr>
<tr>
<td>IODEPO</td>
<td>International Oceanographic Data and Information Exchange</td>
</tr>
<tr>
<td>CENPAT</td>
<td>Centro Nacional Patagonico – CONICET</td>
</tr>
</tbody>
</table>
3.2 JDK installation.

Download JDK installation package v.1.7 from Oracle web-site [Oracle](http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html) and put it to directory `/usr/java`.

Notice that you have to switch to “root” user to run following commands. To switch the user execute command: `su –`.

Extract archive and install script alternatives in directory `/usr/bin` for java, jar and javac commands:

```
[root@local java]# tar -xf jdk-7u17-linux-x64.gz
[root@local java]# alternatives --install /usr/bin/java java /usr/java/jdk1.7.0_17/bin/java 100
[root@local java]# alternatives --install /usr/bin/javac javac /usr/java/jdk1.7.0_17/bin/javac 100
[root@local java]# alternatives --install /usr/bin/jar jar /usr/java/jdk1.7.0_17/bin/jar 100
[root@local java]# alternatives --config java
There are 3 programs which provide 'java'.
Selection  Command
-----------------------------------------------
 1          /usr/lib/jvm/jre-1.6.0-openjdk.x86_64/bin/java
 2          /usr/lib/jvm/jre-1.4.2-gcj/bin/java
 3          /usr/java/jdk1.7.0_17/bin/java
Enter to keep the current selection[+], or type selection number: 3
```

After this check java version:

```
[root@local]# java -version
java version "1.7.0_17"
```
Update the value of environment variable $JAVA_HOME in file /etc/profile:

```
export JAVA_HOME="/usr/java/jdk1.7.0_17"
```

Execute following command to apply changes:

```
/etc/profile
```

### 3.3 Application server JBoss installation.

Download JBoss directory to the path /var/opt. To reconfigure memory and other JVM settings open configuration file /var/opt/jboss-4.2.3.GA/bin/run.conf. Find record like

```
JAVA_OPTS="-client -Djava.encoding=UTF-8 -Dfile.encoding=UTF-8 -Xms512m –Xmx8096m –XX:MaxPermSize=2048M"
```

- `Xms` – initial heap memory size;
- `Xmx` – maximum heap memory size;
- `XX:MaxPermSize` – maximum size of permanent generation memory segment.


**WARNING:** Application server JBoss should run under proper user (non root!). It’s a good practice to have user “jboss” for this purpose. Directory /var/opt/jboss-4.2.3.GA should have “jboss” as owner and rights 755. To run JBoss under proper user open file /var/opt/jboss-4.2.3.GA/bin/jboss_init_redhat.sh in a text editor and find record

```
JBOSS_USER=${JBOSS_USER:-"RUNASIS"}
```

Specify “jboss” instead of "RUNASIS" value.

### 3.4 Data Provider web application installation.

1) Extract dpms.sar-2.0-<Data Centre ID>-<release date>.zip to the directory /var/opt/jboss-4.2.3.GA/server/default/deploy;

2) Extract dpms-2.0-<Data Centre ID>-<release date>.zip to the directory /var/opt/jboss-4.2.3.GA/server/default/conf/;

3) Extract DPDB module dpdb.ear and database configuration file dpdb-ds.xml to the directory /var/opt/jboss-4.2.3.GA/server/default/deploy;
4) Create new directory /var/opt/jboss-4.2.3.GA/server/default/data/dpms and copy files dpmsdb.properties and dpmsdb.script there;
5) Copy libraries dpdb-api.jar and dpdb-ejb.jar to the directory /var/opt/jboss-4.2.3.GA/server/default/lib;
6) Copy script init_soi_gen.sh to the directory /opt/dp;
7) Run script to make web-service stubs for Service Bus see item 3.5;
8) Configure concurrent jobs configuration: open file /var/opt/jboss-4.2.3.GA/server/default/conf/dpms/config/quartz.properties in the text editor. Value for the parameter org.quartz.threadPool.threadCount should be 150. If not so, specify this value to 150 and save configuration file;
9) Configure user roles: open file /var/opt/jboss-4.2.3.GA/server/default/conf/dpms/config/dpms-roles.properties in the text editor. The role name should meet the rule:

```plaintext
<nodeld>:DP:ADMIN – Data Provider Administrator
<nodeld>:DP:USER – Data Provider Operator
```

where nodeld is ODP Data Centre identificator. To find out what is the proper nodeld for Data Centre please contact Regional ODP Node Administrator or Partnership Centre for the IODE ODP;

10) Open file /var/opt/jboss-4.2.3.GA/server/default/conf/dpms/dpdb.properties in the text editor. Update local pathes PATH_TO_IR and ELEMENTS_MD_FILE if needed;
11) In case of update version to 2.x it is necessary to update config.properties manually. Add new record to the end of the file:

```plaintext
nodeld = RNODC
```

where RNODC is Data Centre identificator. Each Data Centre has unique identificator.

12) Double-check that all files and directories inside JBoss Application Server has owner “jboss”. To specify owner for all files inside use following command:

```bash
chown –R jboss.ftp_users jboss-4.2.3.GA/
```

where ftp_users is linux group name for user “jboss” by default.


1) Stop JBoss Application Server;
2) Open file /var/opt/init_soi_gen.sh in the text editor;
3) Specify proper Service Bus address in URLs:

```plaintext
http://<ODP_REGIONAL_NODE> :18080/webServiceBus?wsdl
http://<ODP_REGIONAL_NODE> :18080/loggingws/LoggingWSImpl?wsdl
```
where ODP_REGIONAL_NODE is URL of ODP node this Data Provider connect to.

4) Save file and run script init_soi_gen.sh.

As a result, following libraries will be created:
/var/opt/jboss-4.2.3.GA/server/default/lib/bus-ws-client.jar
/var/opt/jboss-4.2.3.GA/server/default/deploy/dpms.sar/lib/logging-ws-client.jar

5) Start up Jboss Application Server.

3.6. Installing and configuring user authentication.

ODP is implementing unified user access policy to all components, including the Data Provider. The policy is based on a uniform classification of user roles and provide single sign-on system (SSO). Authentication and authorization JOSSO client configured via a configuration file /var/opt/jboss-4.2.3.GA/server/default/conf/josso-agent-config.xml. It’s necessary to specify proper URLs to JOSSO server. In this example client configured for ODP regional node of Argentina (http://portal.mincyt.gob.ar:8080/josso/signon/login.do):

```xml
<gatewayLoginUrl>http://portal.mincyt.gob.ar:8080/josso/signon/login.do</gatewayLoginUrl>
<gatewayLogoutUrl>http://portal.mincyt.gob.ar:8080/josso/signon/logout.do</gatewayLogoutUrl>
<protocol:ws-service-locator endpoint="portal.mincyt.gob.ar:8080" />
<agent:partner-app id="dpms" context="/dpms"/>
<automaticLoginStrategies>
  <agent:default-automaticlogin-strategy mode="OPTIONAL">
    <property name="ignoredReferrers">
      <list>
        <value>http://portal.mincyt.gob.ar:8080/IDBUS</value>
      </list>
    </property>
  </agent:default-automaticlogin-strategy>
</automaticLoginStrategies>
```
4 USER INTERFACE

Purpose

The User Interface allows a user to choose the Data Provider (DP) functions and to switch to the relevant section of the menu.

4.1 Main configuration.

Open URL http://<Data_Provider_URL>/dpms in a web-browser. Click “Private login” link and type in login and password. To get user account, it’s necessary to create new account at the ODP Portal of regional node and ask portal administrator to assign “Data Provider Administrator” role. Click “Configure settings” link from the menu on the left. On the opened page click “Edit” link beside text “Configuration file of the Data Provider” (see Figure 4.1).

Figure 4.1 – Administrator configuration interface

The page with all parameters is shown on figure 4.2.

It’s necessary to check parameters concerning to table 4.1.
Table 4.1 – List of Data Provider configuration parameters.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Default value</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration Server URL</td>
<td><a href="http://is.mincyt.gob.ar:80/80/iserv2">http://is.mincyt.gob.ar:80/80/iserv2</a></td>
<td>URL to the Integration Server entry point, that Data Provider connect to. Integration Server is situated at ODP region node</td>
</tr>
<tr>
<td>Integration Server user name</td>
<td>jcportal</td>
<td>User name at Integration Server. This value don’t change as a rule</td>
</tr>
<tr>
<td>Integration Server user password</td>
<td>portal</td>
<td>User password. This value don’t changed as a rule</td>
</tr>
<tr>
<td>Internal server IP-address[:port]</td>
<td>127.0.0.1:8080</td>
<td>Internal server IP address.</td>
</tr>
<tr>
<td>External server IP-address[:port]</td>
<td>dp.mincyt.gob.ar</td>
<td>External server IP address <strong>Notice:</strong> It’s mandatory rule that IP address must be changed from localhost to DNS name or IP address that accessible to Integration Server address.</td>
</tr>
<tr>
<td>Proxy-server address[:port]</td>
<td></td>
<td>IP address of proxy server. This option</td>
</tr>
<tr>
<td><strong>uses in rare case of network topology and don’t used as a rule</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Country code** | AR | Country code concerning ISO 3166-1-alpha-2 ([http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm](http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm)) |
|**Date format** | dd-MM-yyyy HH:mm | Date and time template for user interface. |
|**Unique Data Provider ID** | MINCYT | Unique Data Provider identifier. Value is specified in Partnership Centre for the IODE ODP |
|**Data Provider name** | MINCYT | Data Provider name. Value is specified in Partnership Centre for the IODE ODP |
|**Node identifier** | MINCYT | Identifier specified on a type of the ODP node. |
|**Interface language** | English | User interface language |
|**Last Data Provider software update date** | | Last Data Provider software update date. Specified by software developers. |
|**Data Provider software version** | | Specified by software developers. |
|**Documentation web site (optional)** | [http://www.oceandataportal.org](http://www.oceandataportal.org) | Documentation web site |
|**SOAP web-service mode enabled** | true | Event-driven communication model. |
|**Logging mode (SOAP web-service invocation)** | true | Logging mode. Data Provider logs to Service Bus |
|**Number of records in test request response** | 100 | Record count received in response from data source during test request |
|**The maximum number of concurrent threads** | 5 | Number of concurrent threads running at Data Provider server. Number depends on CPU and RAM. |
|**Default federation code** | 1.3 | Don’t change |
|**Vocabulary dictionary standard** | BODC | Don’t change |
|**Quartz-service name** | DPMSQuartzScheduler | Don’t change |

After completing push button “Save”. Notice, that saving process can take too long because of large amount of vocabulary parameters. The changes will be saved after message appears at the top of the page.

4.2 Downloading and updating system vocabularies.

After first time Data Provider installation it’s necessary to update system vocabularies. Click “Administrator” tab and click “Configure schedulers” link.

Click on the icon beside item “Updating system parameters” and specify 2 minutes after current time in “Quartz job expression” field.

4.3 Data Provider maintenance.

JBoss Application server logs all events into files boot.log and server.log in the directory `/var/opt/jboss-4.2.3.GA/server/default/log/`.

To test Data Provider open following URL in a web browser: [http://[servername]:[port]/dpms/](http://[servername]:[port]/dpms/), where servername – domain name or server IP address, port – port number associated with JBoss.
Application Server (80 by default). Home page will appear in a browser if installation and configuration have been performed correctly (see Figure 4.3).

Figure 4.3 – Data Provider home page

Click “Private login” link and specify login and password. If authorization has not been successful, double check user role at the ODP Portal. User role should meet following rules:

\<nodeId\>:DP:ADMIN – Data Provider Administrator
\<nodeId\>:DP:USER – Data Provider Operator

where nodeID - ODP Data Centre identifier in accordance with Table 3.1

5 DATA PROVIDER MAINTENANCE

Maintenance of ODP Data Provider contains follows tasks:

− submitting data to ODP using Data Provider interface;
− server monitoring and submitted data and metadata availability. Recovery procedures after failures;
− maintenance work.

All maintenance work is performed by operator with the help of system administrator.

These actions should be performed on a daily basis:

− data sources availability and actual data inside;
− making tests after recovery procedures;
− put in new records to maintenance journal;

Following actions should be performed weekly:

− backup Data Provider configurations, metadata and data files;
− audit system logs

5.1 Maintenance work

System administrator performs backup to avoid information loss.

There is a list of directories to backup:

/var/opt/jboss-4.2.3.GA/server/default/deploy/dpms.sar – weekly
/var/opt/jboss-4.2.3.GA/server/default/conf/dpms – weekly
/var/opt/jboss-4.2.3.GA/server/default/conf/data/dpms – daily