

# How a virtual appliance facilitates the data management: the SeaDataCloud project experience

Leda Pecci, Marcello Galli

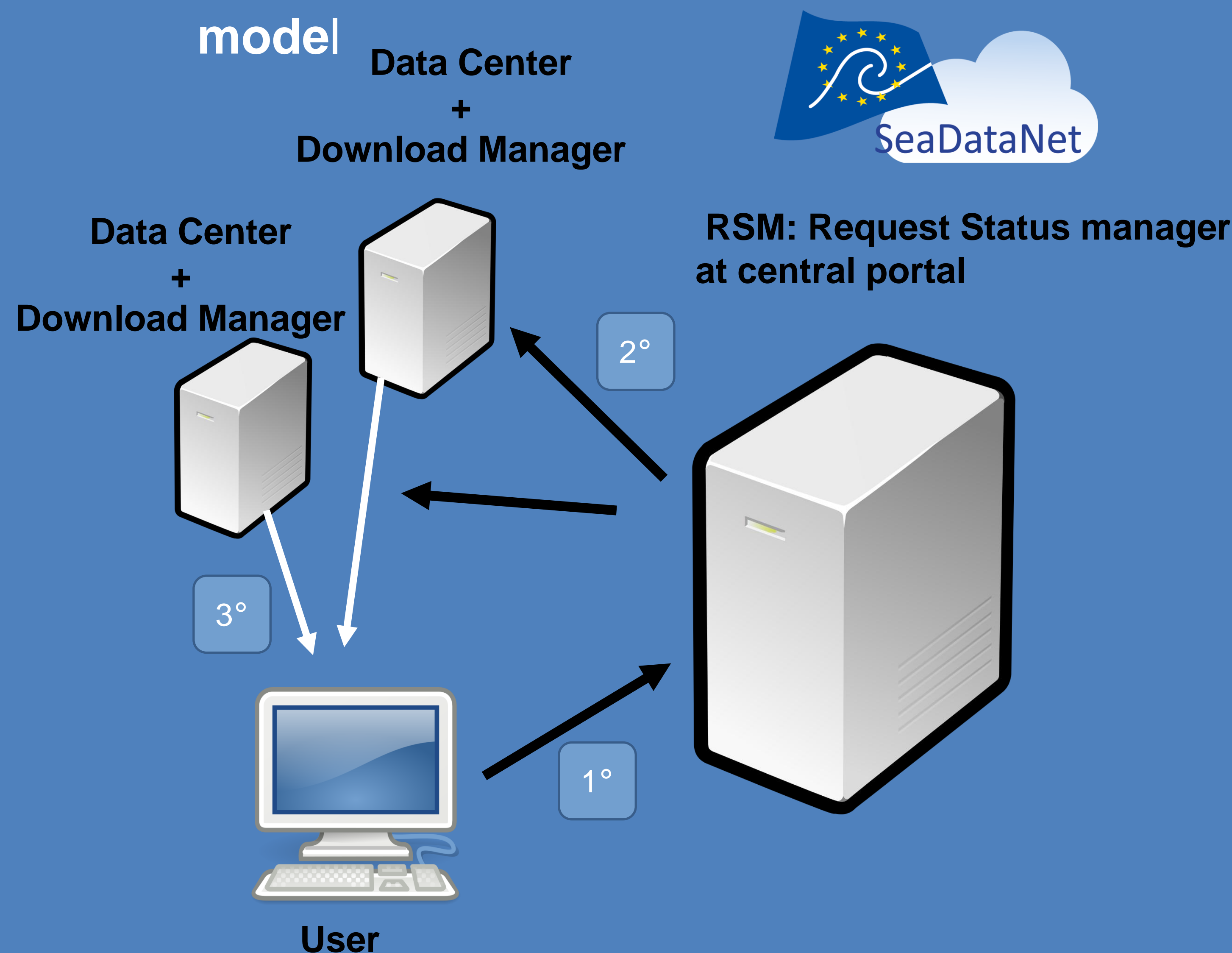
ENEA – Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile

The SeaDataNet infrastructure allows to access to a very large number of ocean and marine data sets stored by more than 110 data centres deployed in a territory that extends beyond Europe.

The SeaDataCloud project has in its main objectives the innovation of the SeaDataNet infrastructure, reached by means of the adoption of cloud computing as well as the High Performance Computing (HPC).

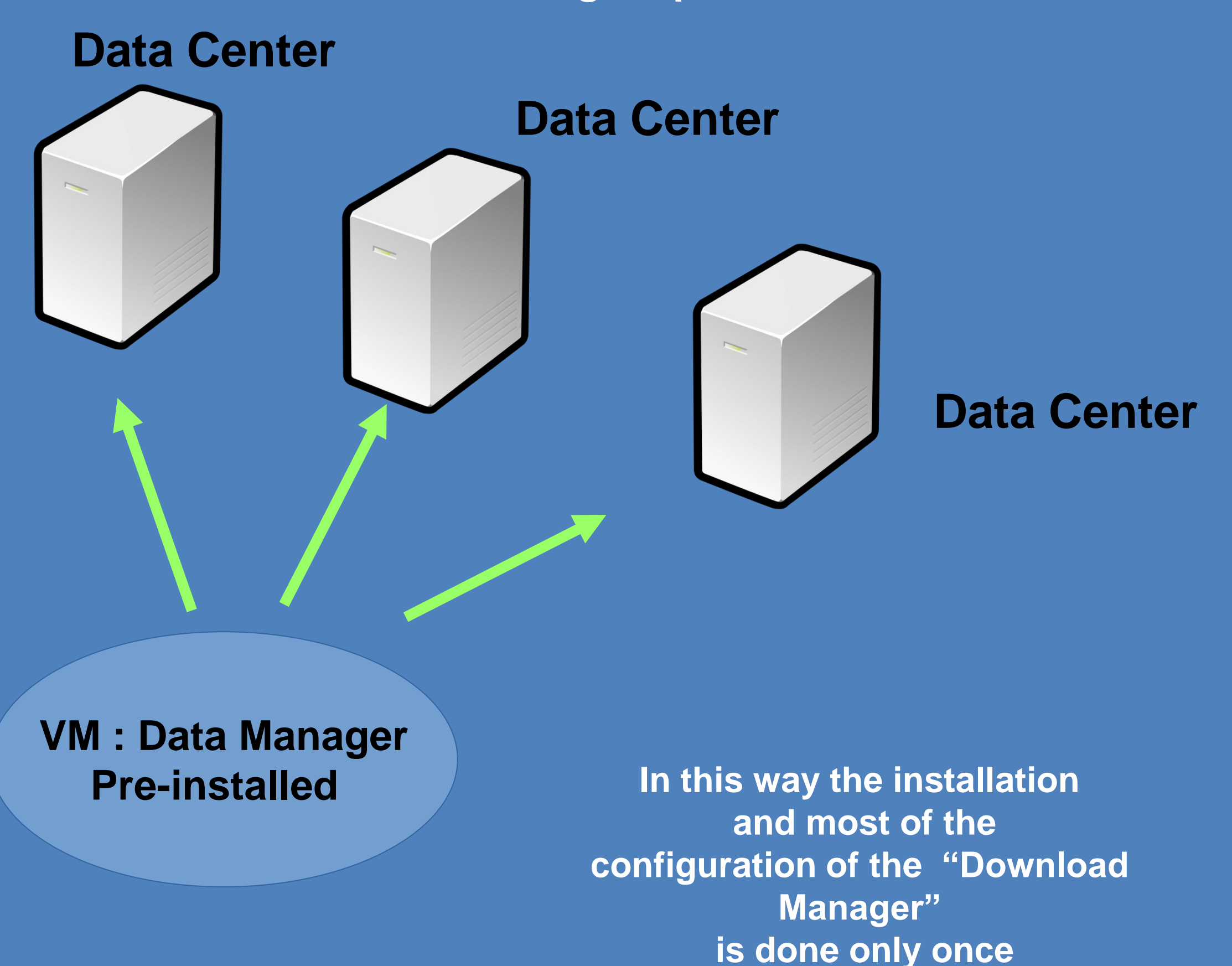
The SeaDataNet central portal (RSM) plays the role of providing data visualisation, discovery and access services, the latter service is based on the CDI (Common Data Index). A key component allows data centres to manage data sets requests, it is called Download Manager (DM), a Java application that communicates with the central portal and provides data to SeaDataNet users. As far as regards the project activity there is a challenge to simplify the management of the network nodes, especially for new data centres, of the semi-distributed system architecture, that is the base of the infrastructure.

## SeaDataNet : data flow model



## The installation and configuration of the "Data manager" on each one of the 110 "Data Centers" is a complex task

Our solution is an appliance (a Virtual machine) with the "Download Manager" pre-installed



**The Data Centers store the data and transmit to the RSM only the metadata, in standard format, for indexing and data search purposes. A data user asks RSM, which authenticates the user and informs the data centers of the request. Once data sets are ready the user downloads the requested data from the Data Centers. A software component: the "Data Manager" is installed on each "Data Center" to manage the data flow.**

A solution that helps the system managers of data centres from installation, configuration and maintenance of the SeaDataNet node has been carried out.

Data providers that intend to become a node of the infrastructure have to install, configure and maintain a server that contains the "Data Manager" application, that manages the delivery of the data files ordered, and the software required to run the application.

Virtualization technology can optimise the connection of data centres to the infrastructure and delivers an alternative over traditional hardware servers.

We make available a pre-built virtual appliance, containing the "Download Manager" component. This is a special purpose virtual machine, based on a Linux operating system and containing only the necessary software applications (e.g. web and application servers) and minimal operating system components. This makes the system simple, easy to maintain and less vulnerable to security breaches.

This virtual appliance requires only some easy configuration after being installed, which doesn't need an expert system administrator.

The SeaDataNet virtual appliance includes all necessary data management tools, it is easily deployable into a compatible virtualisation environment, ready to use with minimal setup, which eases the process of maintenance and management of the distributed nodes. A specific bash procedure has been developed by ENEA that allows a semi-automatic installation of the appliance software and some testing.

The virtual appliance is realised using a TurnKey Linux (TKL), based on Debian 8 (Jessie) with some updated components.



This is a stable operating system, suitable for the production environment, well known for its robustness and reliability. The operating system is updated and installed automatically as well as the software patches, contained in the SeaDataNet virtual appliance,. The virtual appliance contains only open source software.

## It is easier to become a SeaDataNet Data Provider

The Download Manager is easy to install and configure

Easy to maintain the node

The virtual appliance is available for different virtual machine environment

The SeaDataNet Virtual appliances eliminate the need for physical hardware because they run on virtual platform solutions. The advantages of adopting this solution are different, data centres save money on hardware, software, the availability and reliability of the global infrastructure increase, because the disaster recovery of the nodes is simplified, the operating system and the data can be easily recovered by duplicating images of the virtual appliance and migrating them in another host system.

Furthermore virtualization provides the ability to adapt the hardware to real needs, increasing the amount of RAM, the size of hard disks, etc., via software. On a single physical platform multiple virtual machines can run, this allows to have lower energy consumption, that means lower costs and benefits for the environment.

The appliance is usually managed via a ssh (secure shell) , but can also be managed by a web interface, which uses the tools shellinabox and webmin. Currently, the time required to install and configure the download manager is reduced and the maintenance of the system is easier.