Geophysical Research Abstracts
Vol. 15, EGU2013-9051, 2013
EGU General Assembly 2013

Service architecture challenges in building the KNMI Data Centre

Wim Som de Cerff (1), John van de Vegte (1), Maarten Plieger (1), Ernst de Vreede (1), Raymond Sluiter (1), Jan Willem Noteboom (1), Ian van der Neut (1), Hans Verhoef (1), Robert van Versendaal (1), Martin van Binnendijk (1), Henk Kalle (1), Arthur Knopper (1), Gijs Calis (2), Siu Siu Ha (2), WIm van Moosel (2), Henk-Jan Klein Ikkink (3), and Tuncay Tosun (3)

(1) Royal Netherlands Meteorological Institute, R&D ICT & Sensor Technology, De Bilt, Netherlands (wim.som.de.cerff@knmi.nl), (2) TriOpSys, Utrecht, the Netherlands, (3) Quality house, the Netherlands

One of the objectives of KNMI is to act as a National Data centre for weather, climate and seismological data. KNMI has experience in curation of data for many years however important scientific data is not well accessible. New technologies also are available to improve the current infrastructure. Therefore a data curation program is initiated with two main goals: setup a Satellite Data Platform (SDP) and a KNMI data centre (KDC). KDC will provide, besides curation, data access, and storage and retrieval portal for KNMI data. In 2010 first requirements were gathered, in 2011 the main architecture was sketched, KDC was implemented in 2012 and is available on: http://data.knmi.nl

KDC is built with the data providers involved with as key challenge: 'adding a dataset should be as simple as creating an HTML page'. This is enabled by a three step process, in which the data provider is responsible for two steps:
1. Provide dataset metadata: An easy to use web interface for providing metadata, with automated validation. Metadata consists of an ISO 19115 profile (matching INSPIRE and WMO requirements) and additional technical metadata regarding the data structure and access rights to the data. The interface hides certain metadata fields, which are filed by KDC automatically.
2. Provide data: after metadata has been entered, an upload location for uploading the dataset is provided. Also scripts for pushing large datasets are available.
3. Process and publish: once files are uploaded, they are processed for metadata (e.g., geolocation, time, version) and made available in KDC. The data is put into archive and made available using the in-house developed Virtual File System, which provides a persistent virtual path to the data.

For the end-user of the data, KDC provides a web interface with search filters on key words, geolocation and time. Data can be downloaded using HTTP or FTP and can be scripted. Users can register to gain access to restricted datasets.

The architecture combines Open Source software components (e.g. Geonetwork, Magnolia, MongoDB, MySQL) with in-house built software (ADAGUC, NADC) and newly developed software.

Challenges faced and solved are: How to deal with the different file formats used at KNMI? (e.g. NetCDF, GRIB, BUFR, ASCII); How to deal with the different metadata profiles while hiding the complexity of this to the user? How to incorporate the existing archives? KDC is a node in several networks (WMO WIS, INSPIRE, Open Data): how to do this?
In the presentation/poster we will describe what has been done for each of these challenges and how it is implemented in KDC.