Article

Multi-Source Data Processing Middleware for Land Monitoring within a Web-Based Spatial Data Infrastructure for Siberia

Jonas Eberle *, Siegfried Clausnitzer, Christian Hüttich and Christiane Schmullius

Department for Earth Observation, Institute for Geography, Friedrich-Schiller-University Jena, Loebdergraben 32, 07743 Jena, Germany; E-Mails: siegfried.clausnitzer@uni-jena.de (S.C.); christian.huettich@uni-jena.de (C.H.); c.schmullius@uni-jena.de (C.S.)

* Author to whom correspondence should be addressed; E-Mail: jonas.eberle@uni-jena.de; Tel.: +49-3641-948-889; Fax: +49-3641-948-882.

Abstract: Land monitoring is a key issue in Earth system sciences to study environmental changes. To generate knowledge about change, e.g., to decrease uncertaincy in the results and build confidence in land change monitoring, multiple information sources are needed. Earth observation (EO) satellites and *in situ* measurements are available for operational monitoring of the land surface. As the availability of well-prepared geospatial time-series data for environmental research is limited, user-dependent processing steps with respect to the data source and formats pose additional challenges. In most cases, it is possible to support science with spatial data infrastructures (SDI) and services to provide such data in a processed format. A data processing middleware is proposed as a technical solution to improve interdisciplinary research using multi-source time-series data and standardized data acquisition, pre-processing, updating and analyses. This solution is being implemented within the Siberian Earth System Science Cluster (SIB-ESS-C), which combines various sources of EO data, climate data and analytical tools. The development of this SDI is based on the definition of automated and on-demand tools for data searching, ordering and processing, implemented along with standard-compliant web services. These tools, consisting of a user-friendly download, analysis and interpretation infrastructure, are available within SIB-ESS-C for operational use.

1. Introduction

The availability of well-prepared geospatial time-series information for environmental research is limited. Individual processing, depending on the data source, as well as changing data formats, all need to be done by the user, posing additional challenges. Working with time-series data is especially time-consuming, due to the large amount of data. For environmental studies, time-series data is an important information source to identify impacts and changes. Several initiatives, such as the Global Observation of Forest and Land Cover Dynamics (GOFC-GOLD) and the Northern Eurasia Earth Science Partnership Initiative (NEESPI), focus on monitoring the impact of global climate changes on the Earth's surface [1,2]. This is important for land-cover and land-use change detection, as well as for disaster management, including fires, droughts and floods [3–8]. To do this, spatial time-series data from multiple data sources is needed. A lot of complimentary data is available for environmental research. Remote sensing satellites can provide this time-series data, as they are able to provide spatial and temporal views of environmental parameters, especially for large areas. However, *in situ* data from meteorological stations are also useful, supporting the analysis of remote sensing data and giving an overview of the climate and the environment being studied.

As many data distributors provide data through web-based systems and programming interfaces, research needs to find ways to automate the steps of finding, downloading and processing data. One example is the Giovanni tool for interactive time-series data exploration and analysis [9]. In addition to the processing needs of users, it is important to establish a system that provides access to multiple data sources. Both multi-source data access and additional processing steps can be integrated in an infrastructure based on data warehouse principles, introduced by Jones [10]. A data warehouse is "a foundation for decision support systems and analytical processing" [10] and provides a snapshot of multiple operational databases. Furthermore, spatio-temporal data warehouses exist, including data analysis, data processing and data storage [11,12]. Such a spatio-temporal data warehouse contains the basis structure for the developed multi-source data processing middleware. Therefore, spatial data is being extracted from the original source, transformed and loaded into the warehouse and, afterwards, provided as a middleware component. Similar concepts are described in several papers [13–16]. Another concept of structured data access to distributed systems is the approach for data brokering introduced for geospatial data by Nativi *et al.* [17,18]. As both technologies differ, the aims are the same: reduce the entry barriers for multidisciplinary applications, as well as provide functionalities that give data added value via steps implementing advanced data discovery, data pre-processing and data transformation. Both the broker concept and the proposed infrastructure have, further, the same approach to "shift complexity from users and resource providers to the infrastructure/platform" [17]; this quotation clearly states the aim of the system supporting users in data access, visualization and processing. An example of a processing middleware based on the warehouse concept in the field of Earth observation is the infrastructure that was developed for the European research project, ZAPÁS [19].

In the provided web portal, remote sensing datasets from multiple sources and information products for Siberia can be visualized and downloaded for further analysis.

The Siberian Earth System Science Cluster (SIB-ESS-C, http://www.sibessc.uni-jena.de) was developed with the aim of providing operational tools for multi-source data access, analysis and time-series monitoring for Siberia. The system comprises a metadata catalog allowing for data searching, as well as interoperable interfaces for data visualization, downloading and processing. Within the SIB-ESS-C, data from remote sensing satellites, climate data from meteorological stations and outcomes of research projects are stored. The aim is to provide a wide variety of operational information products free of charge. The advantage of representing different products within a single system is the integration of users' needs into web-based processing services. Concerning climate change and land monitoring, the SIB-ESS-C focuses on land-based information products.

There are several other web-based systems that provide tools to search, order, download (NASA Reverb Client [20], USGS Earth Explorer [21], NCDC Climate WebGIS [22]) and analyze data (NCDC Climate WebGIS [22], NASA Giovanni [9], Virtual laboratory of remote sensing time-series [23]). However, if the data is not provided in a format that can be handled by the user, the data needs to be transformed into another format. In this case, there is a need for data processing steps that can be automated using programming languages. As a great amount of data is available online, clients can use web services to access this data. If further processing is needed (e.g., conversion of formats or units), the data can be transferred and processed in another system, and the processed data should also be made available for users. Providing this data in a standard-compliant format is important, so that other clients (WebGIS or Desktop GIS software) can access the processed data. These activities support the core principle of the Global Earth Observation System of Systems (GEOSS) to establish a "global and flexible network of content providers allowing decision makers to access an extraordinary range of information" [24]. According to the GEOSS architecture principles [25], component systems can be scaled from national to global networks and from *in situ* to remote sensing data, which will also be implemented in the SIB-ESS-C. A further goal of GEOSS is to link existing and planned observation systems together; however, ultimately, the "success of GEOSS will depend on data and information providers accepting and implementing a set of interoperability arrangements, including technical specifications for collecting, processing, storing and disseminating shared data, metadata and products" [26].

Further research is needed in the management, processing and standard-compliant distribution of spatial time-series information. The challenge of generating standardized and operational multi-source data handling structures has not yet been completely addressed in geo-information science. Therefore, the provision of standard-compliant data is a key component of data distribution; however, the combination of data distributed from different sources is also important. Another need is to overcome the lack of still not existing up-to-date time-series data, operationally acquired, preprocessed and provided in common data formats. The interoperability between data providers, data application engineers and policy makers has to be strengthened to make the large amount of valuable information accessible to experts in diverse fields. Data availability, in general, is not an issue, but which kind of data is available for specific dates and areas is a frequently asked question.

The objective of the middleware within the SIB-ESS-C is to build up an operational web-based system where data from different sources are provided and updated. The middleware automatically

collects data from integrated resources to provide standard-compliant web services for data access and visualization. Datasets are then available for further analysis. This paper describes the development of a processing middleware to build up a multi-source database to support land-monitoring research by:

- establishing a multi-source data processing middleware for land observations,
- implementing additional and individual processing steps for integrated data,
- providing standard-compliant visualization, access and download services for time-series data,
- and fostering near real-time monitoring of land processes.

The description of the integrated data sources and their datasets is given in Section 2. Section 3 shows the framework that was developed for the middleware, including data integration and provision. Section 4 explains the SIB-ESS-C web portal as a client and administration component for the middleware, and Section 5 describes cases showing different applications of the middleware. Section 6 offers a conclusion based on the developments and experiences discussed in the previous sections and gives recommendations and predictions regarding further work in this area.

2. Data Sources

Finding suitable spatial data in the area of interest is an essential task in many environmental studies. Different types of data from Earth observation satellites are available, but there are some drawbacks for very-high-resolution satellite data; this data is cost-intensive and not available to the general public. However, some high- and medium-resolution satellite data is freely available, such as data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor from the National Aeronautics and Space Administration (NASA). Another important data source is climate data from meteorological stations. The National Oceanic and Atmospheric Administration (NOAA) provides global climate data, which can be used for non-commercial purposes, according to Resolution 40 (Cg-XII) of the World Meteorological Organization (WMO) [27].

Until now, datasets from three data sources were used for data integration and provision for the SIB-ESS-C (Table 1). Raster time-series data are based on MODIS products; *in situ* climate time-series data are based on the Global Surface Summary of the Day (GSOD) as a daily dataset and the Integrated Surface Database (ISD), which provides hourly data from meteorological stations.

Table 1. Data sources used for operational land monitoring.

Data	Provider	Available Time Ranges
Moderate Resolution Imaging Spectroradiometer	NASA	2000–present
Global Surface Summary of the Day	NOAA	1929–present (depending on station)
Integrated Surface Database	NOAA	1929–present (depending on station)

2.1. Moderate Resolution Imaging Spectroradiometer (MODIS)

The aim of the MODIS sensor is to monitor land surface, oceans and the atmosphere. MODIS data is received from two satellites, *Aqua* and *Terra*, both of which were launched through NASA's Earth Observation System program. Continuous observations are available for Terra since 2000 and for Aqua since 2002. Data from MODIS is provided as land, atmospheric and ocean products on a

systematic basis by different science teams [28]. Each science team has developed a number of standard products, which are applied to raw MODIS data by their data centers. These standard products contain information about atmospheric profiles, surface reflectance, clouds, land and sea surface temperature, thermal anomalies, vegetation indices, snow cover, sea ice cover, *etc.* [29].

The land products are available in several spatial (0.05°, 1 km, 500 m, 250 m) and temporal (monthly, 16-daily, eight-daily, four-daily, daily) resolutions. Data is provided in the Hierarchical Data Format-Earth Observing System (HDF-EOS) format. Based on the spatial resolution, data are available as 5 min swaths, as tiles with a width and height of 10 degrees (see Figure 1 for the tile grid) or as a Climate Modeling Grid (CMG) [30].

Up until now, SIB-ESS-C has provided land-based information datasets. Therefore, only MODIS products from the science teams of land and cryosphere are integrated into the middleware database. The data sources used are listed in Table 2; products for automated integration are listed in Section 5.

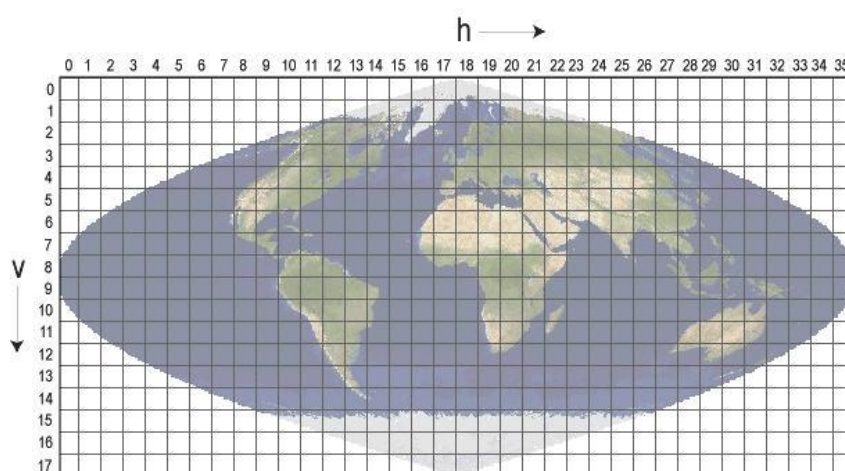**Figure 1.** MODIS sinusoidal tiling system [31].

Table 2. Data sources for MODIS land and cryosphere products.

Science Discipline	Archive Center	Data Access
Land	Land Processes Distributed Active Archive Center (LPDAAC)	http://e4ftl01.cr.usgs.gov/ [32]
Cryosphere	National Snow & Ice Data Center (NSIDC)	ftp://n4ftl01u.ecs.nasa.gov/SAN/ [33]

2.2. Global Surface Summary of the Day (GSOD)

NOAA distributes the GSOD dataset based on meteorological stations that meet WMO standards. These daily summaries are calculated based on the Integrated Surface Database (See Section 2.3). Users can download climate data by station and year, either from interactive web tools or from a server through the File Transfer Protocol (FTP). Data is available as text files in a character-delimiting format; the units are based on the standard American measurement system. As this system differs from the international standard, the units are converted during the processing to SI units (International System of Units). Available parameters for this dataset, such as daily mean/min/max temperature, mean pressure, visibility, wind speed and precipitation, are listed in Table 3, in both the original and

the converted SI-based unit. Indicators for the occurrence of fog, rain or drizzle, snow or ice pellets, hail, thunder and tornados or funnel clouds are also available. The GSOD dataset has been available, for most purposes, since 1929, with some differences based on the measurement period of the individual meteorological stations at each location [34].

Table 3. Available climate parameters within the dataset (original and converted units). SI, International System of Units.

Parameter	Original Unit	Converted SI-Based Unit
Mean temperature	Degrees Fahrenheit	Degrees Celsius
Mean dew point	Degrees Fahrenheit	Degrees Celsius
Mean sea-level pressure	Millibar	Pascal
Mean station pressure	Millibar	Pascal
Mean visibility	Miles	Kilometer
Mean wind speed	Knots	Meter per second
Max sustained wind speed	Knots	Meter per second
Max wind gust	Knots	Meter per second
Max temperature	Degrees Fahrenheit	Degrees Celsius
Min temperature	Degrees Fahrenheit	Degrees Celsius
Precipitation amount	Inches	Millimeters
Snow depth	Inches	Millimeters

2.3. Integrated Surface Database (ISD)

The Integrated Surface Database (ISD) dataset contains hourly measurements from global distributed meteorological stations that meet WMO standards. The US Air Force and Navy, as well as the NOAA National Climatic Data Center (NCDC) started contributing to this database in 1998. Each dataset contains information about temperature, precipitation, wind direction, wind speed, visibility, sky cover, station and sea-level pressure and snow depth. In addition to GSOD, it is possible to retrieve climate data for a specific time of day within a range of at least one to three hours. With such data, weather information can be provided to support the analysis of satellite data. The data is available via FTP and can be downloaded yearly as a text file. The file content was processed to extract all necessary information and converted to a text file with a comma-separated-value (CSV) format to make it easier for users to handle (Figure 2). The units are already SI-based, so no further conversion is needed for data integration [35].

Figure 2. Original and new CSV (with shortened header information) text file format for a single meteorological measurement.

```
# original format:
0150105510999991995121318004+50617+010667FM-12+050599999V0200501N00211220001CN009
0001N1-00141-00431102931ADDAA106000091AG14000GA1001+999999999GF100991001001999999
001001KA1120M+00111KA2120N-00141MA1999999100321MD1010001+9999REMSYN023 333 100 11
21014 55300;
# new format:
USAF,WBAN,datetime,flag,lat,lon,code,elevation,clident,qualitycontrol,winddirecti
on,qc,tc,windspeed,qc,skyceilingheight,qc,dc,skycavokcode,visibility,qc,vc,qvc,ai
rtemp,qc,dewpointtemp,qc,sealevelpressure,qc,precipperiod,precipitation,cc,qc
05510,99999,199512131800,4,50.617,10.667,FM-12,505,,V020,050,1,N,2,1,22000,1,C,N,
009000,1,N,1,-2,1,-5,1,1029,1,6,0,9,1
```

3. Framework for the Multi-Source Data Processing Middleware

The middleware service developed within the SIB-ESS-C (Figure 3 gives a system overview) integrates the time-series data from the Land Processes Distributed Active Archive Center (LPDAAC), the National Snow and Ice Data Center (NSIDC) and the National Climatic Data Center (NCDC) into a spatial database. Metadata information based on available datasets is stored in a second database. The integrated datasets are published via web services that are compliant with Open Geospatial Consortium (OGC) specifications. Data visualization and access is available through the Catalogue Service for Web (CSW; [36]), Web Map Service (WMS; [37]), Web Feature Service (WFS; [38]), Web Coverage Service (WCS; [39]) and Sensor Observation Service (SOS; [40]). The time-series plotting service is available as Web Processing Service (WPS; [41]). In addition to these web services, the SIB-ESS-C web portal acts as a client, accessing the data that are processed by the middleware service. The web portal contains a search engine, a dataset viewer, a time-series plotter and functions to initiate new, on-demand data integration requests for the middleware database.

Open source tools (Table 4 for a complete list) were used to develop the middleware services; PostgreSQL with the PostGIS extension provides the database with the ability to store raster and vector data. Data integration is done with Python scripting (e.g., to execute command-line tools for data downloading and for raster time-series data processing), and R script is used to plot the integrated time-series data. On the service level, MapServer (data visualization and downloading), istSOS (climate data provision), pycsw (metadata provision) and PyWPS (time-series plotting) are used to publish OGC-compliant services.

Figure 3. System overview for the framework developed within the Siberian Earth System Science Cluster (SIB-ESS-C). OGC, Open Geospatial Consortium; CSW, Catalogue Service for Web; WMS, Web Map Service; WFS, Web Feature Service; WCS, Web Coverage Service; SOS, Sensor Observation Service; WPS, Web Processing Service.
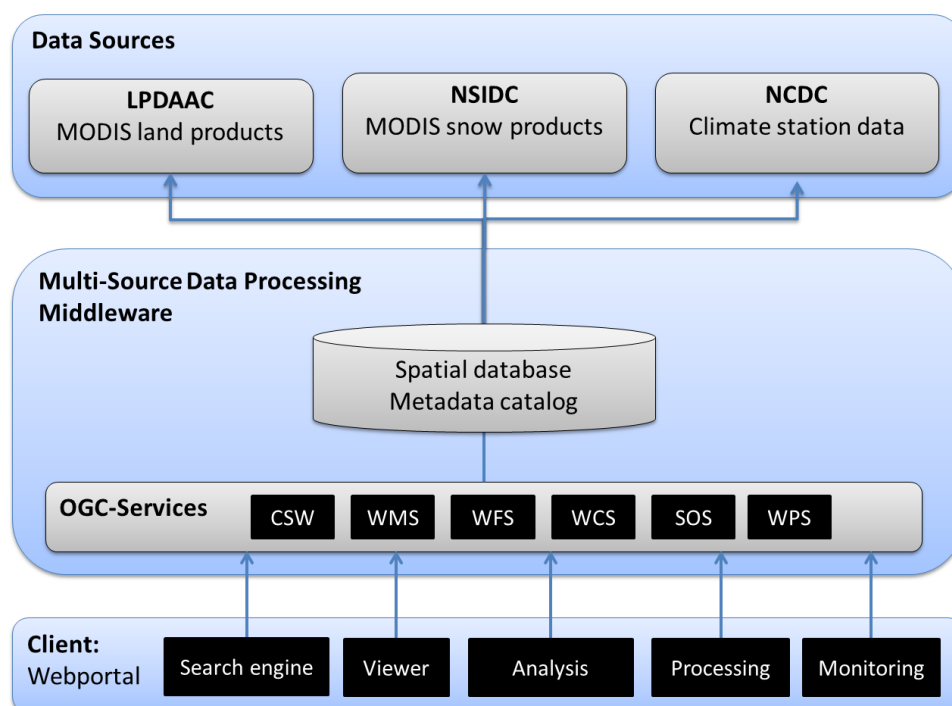
Table 4. Open-source tools used for the developed framework.

Software	Version	Function	Source
PostgreSQL	9.1	Database	postgresql.org
PostGIS	2.0	Spatial extension for PostgreSQL for raster and vector data	postgis.refractions.net
Wget	1.13.4	FTP data retrieval	gnu.org/software/wget
Python	2.7	Data integration script	python.org
GDAL	1.9	HDF to GeoTIFF, data clipping	gdal.org
R	2.14.1	Time-series plotting	r-project.org
MapServer	6.0.3	Data access services	mapserver.org
MapCache	1.0	Caching WMS services	mapserver.org
istSOS	2.0	OGC SOS services	istgeo.ist.supsi.ch/software/istsos
PyWPS	3.2.1	Data processing services	pywps.wald.intevation.org
pycsw	1.4.0	Metadata services	pycsw.org
Apache	2.2.22	Web server	httpd.apache.org
Drupal CMS	7	Web content management system	drupal.org
jQuery	1.8	Frontend JavaScript library	jquery.org
OpenLayers	1.12	Frontend web mapping library	openlayers.org

3.1. Data Integration

The middleware service aims to enable the integration of datasets from external data sources. Therefore, downloading and processing steps have to be automatic. All necessary methods for the complete processing chain—download preparation, downloading, processing and publishing—were implemented with the programming language, Python. The following sections will describe the integration and processing of MODIS products from LPDAAC and NSIDC, as well as the climate data from GSOD and ISD datasets.

### 3.1.1. Integration and Processing of MODIS Products

Multiple steps are necessary to process MODIS products: downloading the file, dataset extraction, GeoTIFF export, tile merging and clipping to the specified region (Figure 4). The workflow for dataset publishing is described in Section 3.2. To integrate a specific MODIS product, further information is needed; this information is also stored within the middleware database for each product:

- data server and directory where files are stored (e.g., FTP)
- raster type of product (Swath, Tile, CMG)
- whether 5 min swaths or tiles are needed (if raster type is equal to swath/tile)
- dataset names to be extracted
- no-data and scale information (if necessary for processing)

Prior to conducting the download process for tiled products, it is necessary to request the specific tiles needed for the specified region (Figure 4(1a)). The bounding box of the gridded tiles is available as vector data [42], which can be intersected with a polygon representing the region. A second step for preparing the download is to identify which dates from the dataset are available and will be processed (Figure 4(1b)).

The next processing steps (Figure 4(3)) must be done for each dataset and timestamp. First, the specified dataset is extracted from the original HDF-EOS file. This is done with commands from the Geospatial Data Abstraction Library (GDAL) [43]. GDAL is a library for reading and writing various spatial raster data formats. It also provides command-line tools for data translation and processing. GDAL commands used in data integration are *gdal_translate* (data translation), *gdal_merge.py* (mosaic building) and *gdalwarp* (transforming an image into a new coordinate system). For each tile or CMG-based file, the necessary dataset is extracted and stored as a GeoTIFF file (Figure 4(3a)). After extraction, the GeoTIFF files from the tiles are merged and re-projected to EPSG:4326 projection (Figure 4(3b)). Using the global CMG format—a single GeoTIFF file for the whole world—it is possible to extract, re-project and export to a GeoTIFF format in one step. At the final stage of processing, the output of the conversion is clipped to the specified region and compressed to save disk storage space (Figure 4(3c)).

Figure 4. Processing chain for the integration of MODIS products.
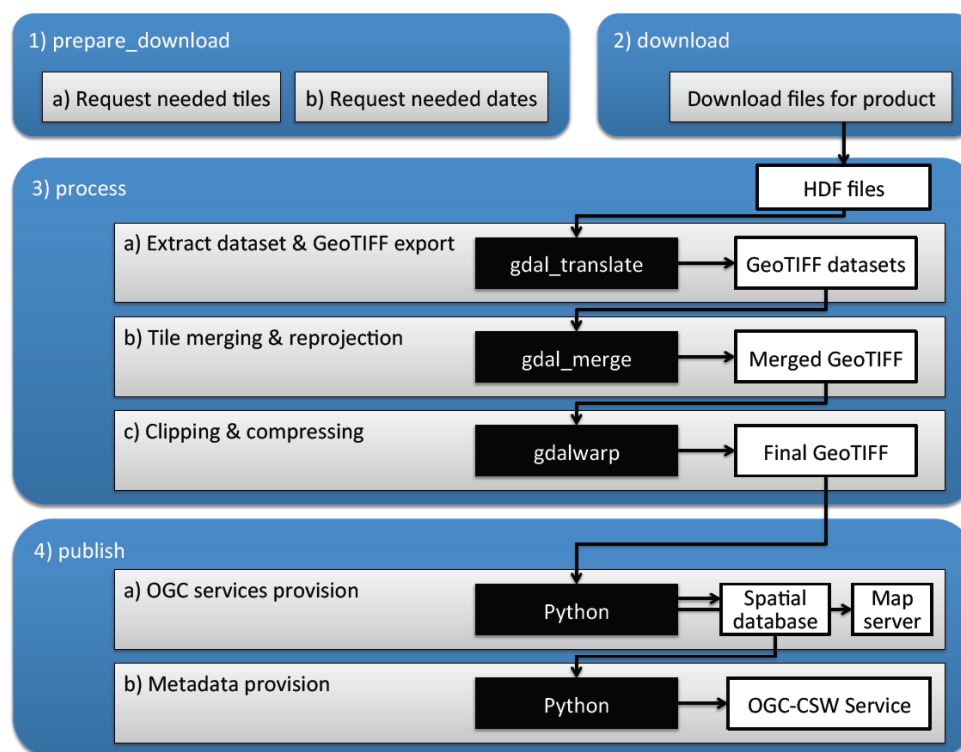
A database for the integrated products was established to provide easy access to the data (Figure 4(4a)). Therefore, any time-series item from MODIS is inserted into the database (Table 5) with the dataset name, the absolute path to the GeoTIFF file, a MultiPolygon as bounding box and the date of the time-series item. The structure of the tables is derived from the raster tile index structure of GDAL, which can be created for GeoTIFF files with the command-line tool, *gdaltindex*. This also automatically creates a bounding box, necessary for OGC-compliant provision. The output is a shapefile containing features of each GeoTIFF file with the generated bounding box. Additional columns for dataset name and the date of the time-series item are added to this structure. In addition to a local link to the GeoTIFF file, it is possible to store the raster data directly in the database. To do this, a further column (raster) is needed, which is provided by the new raster column type of PostGIS 2 [44].

Table 5. Database table structure for MODIS time-series data.

Column Name	dataset	date	location	geom	raster
Column Type	String	Date	String	MultiPolygon	Raster

The subsequent steps in the integration chain are data (Figure 4(4a)) and metadata (Figure 4(4b)) provision with OGC-compliant services (as described in Sections 3.2.1 and 3.2.3). The MODIS processing chain can be extended with additional steps, such as the conversion of units (e.g., degrees Kelvin to degrees Celsius for temperature data), the removal of the scale factor from data values or the resampling to a coarser resolution.

3.1.2. Integration and Processing of the GSOD Dataset

The automatic integration and processing of the GSOD dataset is based on a Python script. Figure 5 shows the processing chain, with the individual steps divided into the methods used for integration into the middleware database.

Figure 5. Processing chain for the integration of Global Surface Summary of the Day (GSOD) and Integrated Surface Database (ISD) data.
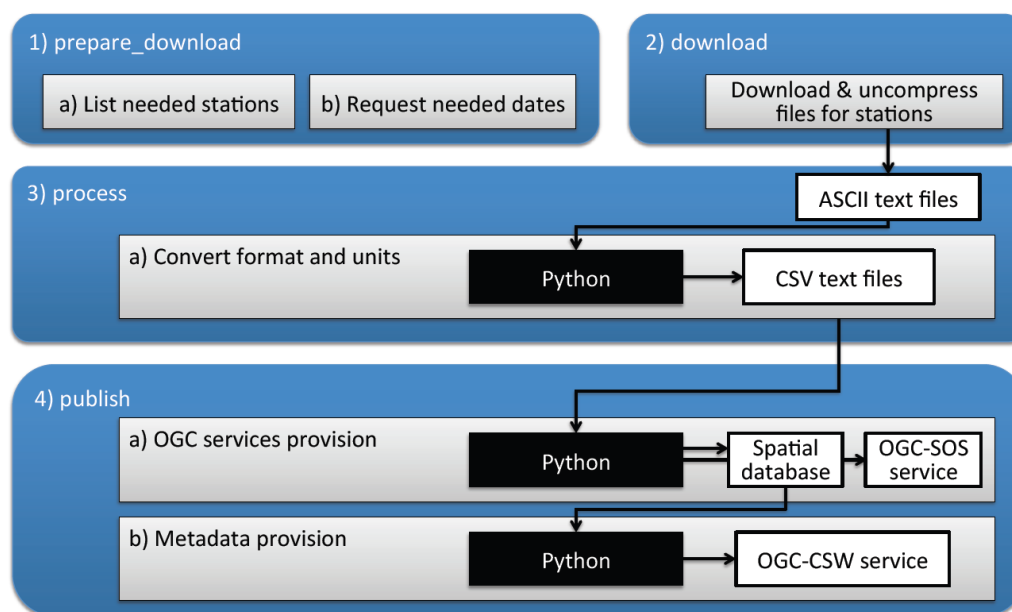
All available stations are listed in a spatial overview table with information about the station ID (USAF, WBAN), station name, coordinates and country, as well as start and end dates for the data (Table 6). The station ID is an important piece of information for automatic integration; it contains the USAF (Air Force Datsav3 station number), WBAN (historical Weather Bureau Air Force Navy number).

Table 6. Fields of database table with overview of integrated meteorological stations.

Gid	USAF	WBAN	Call (ICAO)	Country	Fips	US state	Station name
Lat	Lon	Elevation	Date begin	Date end	Geometry	Location	File identifier

To integrate the data into the middleware database, a list of station IDs and the start and end dates for each station are needed. This information is generated in the preparation step (Figure 5(1a,1b)). After downloading, the data is unzipped. As mentioned in Section 2, the data is then converted from standard US measurement units (degrees Fahrenheit, inches, *etc.*) to SI-based units; the format is also changed from a character-delimiting to a comma-delimiting format. This is done using a separate Python script (Figure 5(3a)). Through this process, all available data are converted and no-data values are specified. As dates with no data are not available within the original file, the dates are added with no-data values to provide a consistent time-series. In addition to the CSV file, the processed data is stored in the PostgreSQL database; its structure is the same as that of the CSV file (Figure 5(4a)).

3.1.3. Integration and Processing of the ISD Dataset

The integration and processing of the ISD dataset is done on demand based on user requests. The processing chain is the same as that used for the GSOD processing (Figure 5). An additional Python script is only needed for data processing. The text files from requested stations are downloaded for the requested time range. After downloading, the files are merged and processed to convert the character-delimiting format to a comma-separated-values (CSV) format. This is done within a separately developed Python script, extracting the most important data. In addition to the data file, the script generates an HTML-based table of the data to provide a quick overview, which is displayed within the web portal. After processing, the files are returned to the web portal, where the user can access the data.

3.2. Data Provision

Services based on the standards and specifications of the Open Geospatial Consortium (OGC), such as Web Map Services (WMS; [37]), Web Feature Service (WFS; [38]), Web Coverage Services (WCS; [39]), Sensor Observation Services (SOS; [40]) and Catalogue Service for Web (CSW; [36]), are used to provide access to the middleware database. In accordance with OGC specifications, these services for visualization and downloading can be implemented and then published.

3.2.1. MODIS Data Visualization and Download Services

OGC-based services for data visualization and raster data downloading can be accomplished using different software packages, such as the open-source software MapServer (http://mapserver.org). MapServer provides OGC-compliant services for data visualization (WMS) and download (WFS, WCS). With time-series data, these services have to be published using the TIME parameter. Using this parameter, users can visualize and download data from a specific date within the time-series.

As all integrated time-series data are inserted into a database table (Table 5), MapServer can use this table as a data source. However, further information is needed to provide a layer as WMS and WCS:

- dataset (layer) name
- time extend (start, end, interval)
- default time
- time positions
- styling and legend information

Two layers are needed for each dataset within the configuration file (map file) of MapServer: a tile-index layer, providing information about the connection to the database and the filtering (according to TIME parameter and dataset name), and the main layer, which contains the dataset that links to the tile-index layer (Figure 6). Information for the database connection and the TIME parameter settings is needed for the tile-index layer. In addition to some general layer settings, the raster layer is configured with information about the database table column where the GeoTIFF file is linked (tile item), the referenced tile-index layer (tile index) and further styling and legend information. For a WCS, a TIME parameter needs to be specified (rangeset name/label) and every available time position for this layer is named in the configuration (time positions).

Figure 6. MapServer configuration for tile-index and raster layer.

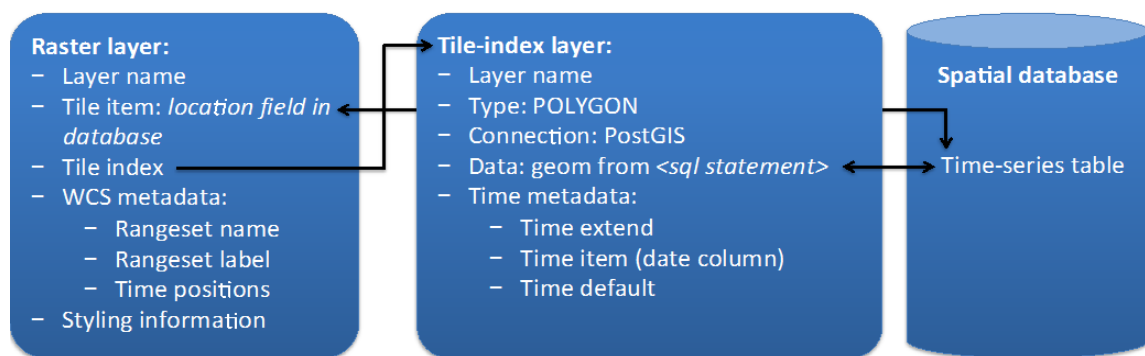**Figure 7.** MapCache configuration XML: source and tileset layer properties (e.g., for MOD13C1 normalized vegetation difference index (NDVI)).

```xml
<source name="mod13c1_ndvi" type="wms">
  <getmap>
    <params>
      <FORMAT>image/png</FORMAT>
      <MAP>/home/mapserver/modis/map.map</MAP>
      <LAYERS>mod13c1_ndvi</LAYERS>
      <TRANSPARENT>TRUE</TRANSPARENT>
    </params>
  </getmap>
  <http>
    <url>http://artemis.geogr.uni-jena.de/cgi-bin/mapserv?</url>
  </http>
</source>

<tileset name="mod13c1_ndvi">
  <source>mod13c1_ndvi</source>
  <cache>disk</cache>
  <grid>WGS84</grid>
  <grid>g</grid>
  <format>PNG</format>
  <metatile>5 5</metatile>
  <metabuffer>10</metabuffer>
  <expires>3600</expires>
  <dimensions>
    <dimension name="TIME" type="values" default="2013-01-01">2000-02-18,[...],2013-01-01</dimension>
  </dimensions>
</tileset>
```

To increase the performance of the data visualization, the caching software MapCache (http://mapserver.org/trunk/mapcache/) was installed and configured for the available layers. Some general properties have to be set, such as the caching storage type (disk or database), default format and error reporting. Service properties can also be set, stating forwarding rules for non-cacheable

requests from the download services, WFS and WCS, as well as for specific WMS request types (GetFeatureInfo and GetLegendGraphic), which will be automatically forwarded to the original service endpoint. In addition to these general and service properties, each cached layer has to be configured with a *source* and a *tileset* object (Figure 7). Within the *source* object, the connection to the original service endpoint is set by URL and default parameters (e.g., format, map file, layer name and transparent flag). The *tileset* object refers to the *source* and the caching storage object and defines properties for the tiling mechanism: grid type, format, expiration time and dimensions. Within the *dimensions* tag, every time position has to be listed.

After updating a time-series dataset, the configuration of MapServer and MapCache has to be updated, as the new time positions have to be added to the configuration. However, as the entire configuration is done using text-based files, it can be updated with any programming language on the server.

3.2.2. Climate Data Services

In addition to providing a CSV file for the climate data, the data can be served as a web service following OGC Sensor Observation Service (OGC-SOS) specifications. Such a web service was set up using the Python-based open-source software istSOS (http://istgeo.ist.supsi.ch/software/istsos/) [45,46].

The administrative interface of istSOS provides functions to create SOS services with required components, such as observed properties, procedures and offerings. The istSOS tutorial [47] proposes a structure whereby any meteorological station is defined as an SOS procedure linked with observed properties (sensor parameters). The procedures are grouped automatically with an SOS offering ("temporary"). Offerings provide the possibility for group procedures. Observations can be added either by using transactional SOS with XML code or an included Python script that parses text files with observations.

In addition to OGC-SOS specifications, istSOS provides a RESTful interface that allows users to communicate with and administer the service using the JSON data format. This interface also allows users to insert meteorological stations as procedures linked to observed parameters, which are geolocated with a point feature. The station IDs described earlier (USAF and WBAN) are also added as an *interface* value. The observation data is formatted in the proposed CSV format with column headers titled the same as the registered sensors. Using the included Python script *cmdimportcsv.py* and the system ID inserted in the prior procedure (meteorological station), the data can be inserted from the command-line.

3.2.3. Metadata Services

The available time-series data, meteorological stations and additional data layers are described using metadata records. These records list information about the data, visualization and download services, as well as the available time positions and time interval. Based on this, a client can build requests for accessing the data. The middleware service generates a metadata record for each time-series dataset following ISO 19115 [48] specifications. The Python-based open-source software, pycsw (http://pycsw.org) is used to provide metadata services. pycsw provides OGC-compliant Catalogue Service for Web (CSW), as well as transactional CSW for metadata insertion and updating. Within the configuration file, transactions can be activated only for specific IP addresses. The inserted

metadata are stored for the SIB-ESS-C in a SQLite-based database. PostgreSQL and MySQL databases are also possible.

Table 7. Excerpt of information stored in the metadata records (example of MODIS Terra land surface temperature time-series Data).

| General Metadata	
File Identifier	MODIS_MOD11_C3_LST_Day_Series
Title	Monthly Daytime Land Surface Temperature from MODIS Terra
Abstract	Time-series of monthly Terra MODIS daytime land surface temperature in Kelvin at 0.05 degrees spatial resolution. To retrieve actual values in Kelvin, a scale factor of 0.02 has to be applied. The unscaled no-data value is encoded as 0. Original MODIS data retrieved from the Land Processes Distributed Active Archive Center (http://e4ftl01.cr.usgs.gov/MOLT/)
Keywords	MODIS, Terra, Siberia, Temperature, Global, Monthly, Series, Daytime
Lineage	MODIS HDF Level 2 product was converted to GeoTIFF with *gdal_translate* (Version 1.9)
Data Information	
Description	Land Surface Temperature
Data Type	RASTER
Coverage Content Type	Physical Measurement
SRS	EPSG:4326
BBOX	57.1301270 81.2734985 179.8292847 42.2901001
Columns	2,454
Rows	780
Resolution	0.05
Scale Factor	0.02
No Data Value	0
Time Begin	2000-03-01
Time End	2012-09-01
Time Interval	P1M
Dates	2000-03-01, 2000-04-01, 2000-05-01,…, 2012-08-01, 2012-09-01
Services	
WMS URL	http://artemis.geogr.uni-jena.de/sibessc/modis
WMS Protocol	WebMapService:1.3.0:HTTP
WMS Description	MODIS Terra LST Day Monthly
WMS Name	mod11c3_lst_day
WCS URL	http://artemis.geogr.uni-jena.de/sibessc/modis
WCS Protocol	WebCoverageService:1.1.0:HTTP
WCS Description	MODIS Terra LST Day Monthly
WCS Name	mod11c3_lst_day

The metadata records are the main entry point for accessing the multi-source data middleware database. In addition to general information, such as the title, abstract, keywords and lineage, a wide range of metadata information can be used to describe the time-series data (Table 7). To link to OGC-compliant services for visualization and download, *DigitalTransferOptions* are provided within

the metadata record. A client can retrieve metadata records based on a specific identifier or a search result, then can parse the information and visualize or download the data through the provided services. With the metadata information, the client knows which time positions are available and which services (e.g., Web Map Service, Web Coverage Service or any other http link) can be used, in accordance with user needs. Parsing the metadata record, the client can further distinguish between time-series raster data as physical measurements or as classifications. This distinction is important for aspects, such as providing the correct analysis processes, which differ for classification results (e.g., burned area), as opposed to continuous data, such as land surface temperature, vegetation indices and snow cover.

4. SIB-ESS-C Web Portal as Client for Middleware Services

The middleware services are contained within the SIB-ESS-C web portal. This web portal (Figure 8) includes functions allowing users to administer and manage the middleware services; it also allows easy access to the integrated datasets. Users are supposed to interact closely with the data to receive the best information. Visiting the web portal, the user can go through the data catalog that contains the available data of the middleware database. The metadata catalog can be searched, and the resulting records can be investigated. The data can then be visualized and downloaded.

Figure 8. Screenshot of the SIB-ESS-C web portal (http://www.sibessc.uni-jena.de).
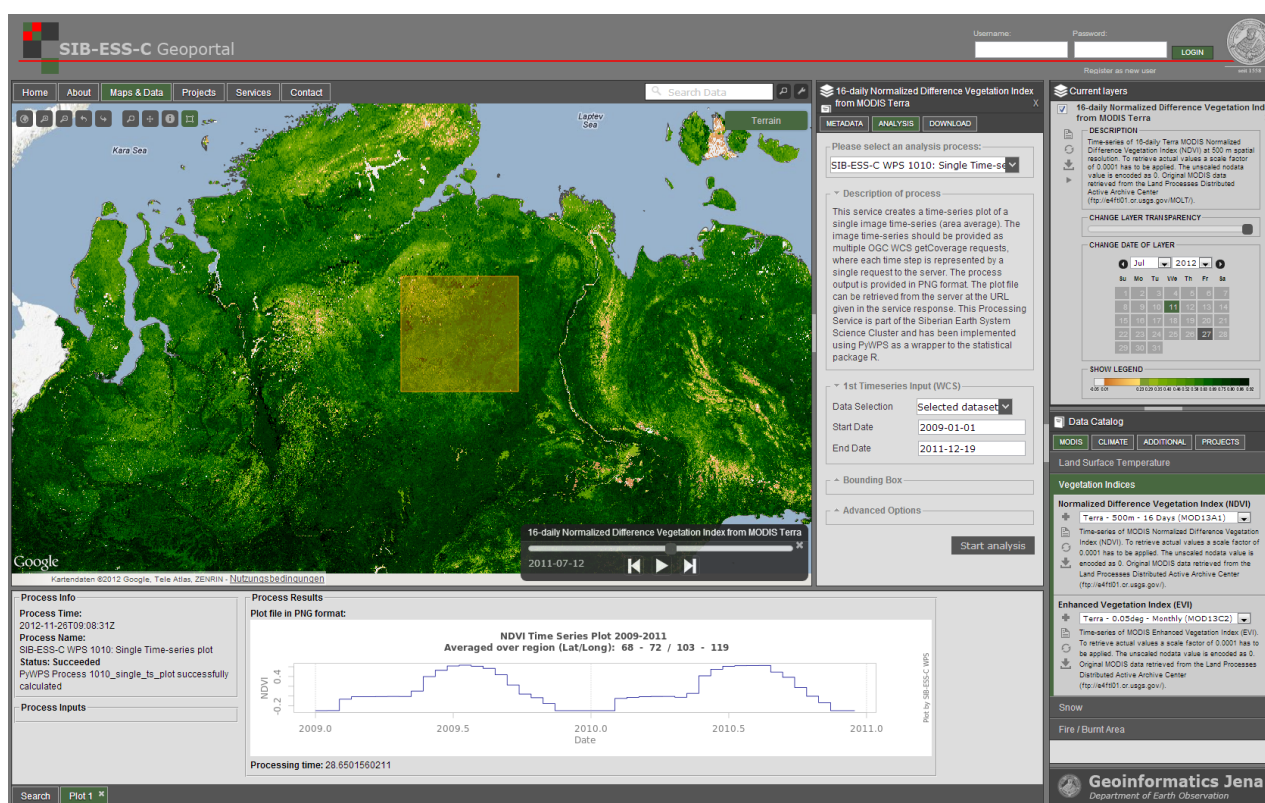
In addition to providing visualization and download tools for the developed middleware services, the web portal also controls the data-integration process, with each step logged within the system. This feature was integrated for on-demand processing. Developments are currently leading in the direction

that allow users to define their own study areas, select datasets and run integration themselves. With the implemented logging functions within the web portal, the user is informed when integration is completed.

Open-source software has been used to develop the web portal. In the backend, Drupal CMS (http://drupal.org) provides a proxy to external web services, converts XML code to JSON code for better processing within the web-frontend programming language, JavaScript, and provides RESTful services for user registration and authentication. The frontend has been developed using the jQuery library (http://jquery.org) and extensions of jQuery. The map viewer for visualizing the data has been created using the OpenLayers library (http://openlayers.org).

5. Applications for the Multi-Source Data Processing Middleware

The developed middleware services can be used for different applications supporting research initiatives, such as GOFC-GOLD and NEESPI, or for research projects, providing access to a wide range of datasets. Some data for Siberia is being integrated automatically (Section 5.1). Data for smaller study areas can be integrated on demand, especially data necessary for higher resolution products (Section 5.2).

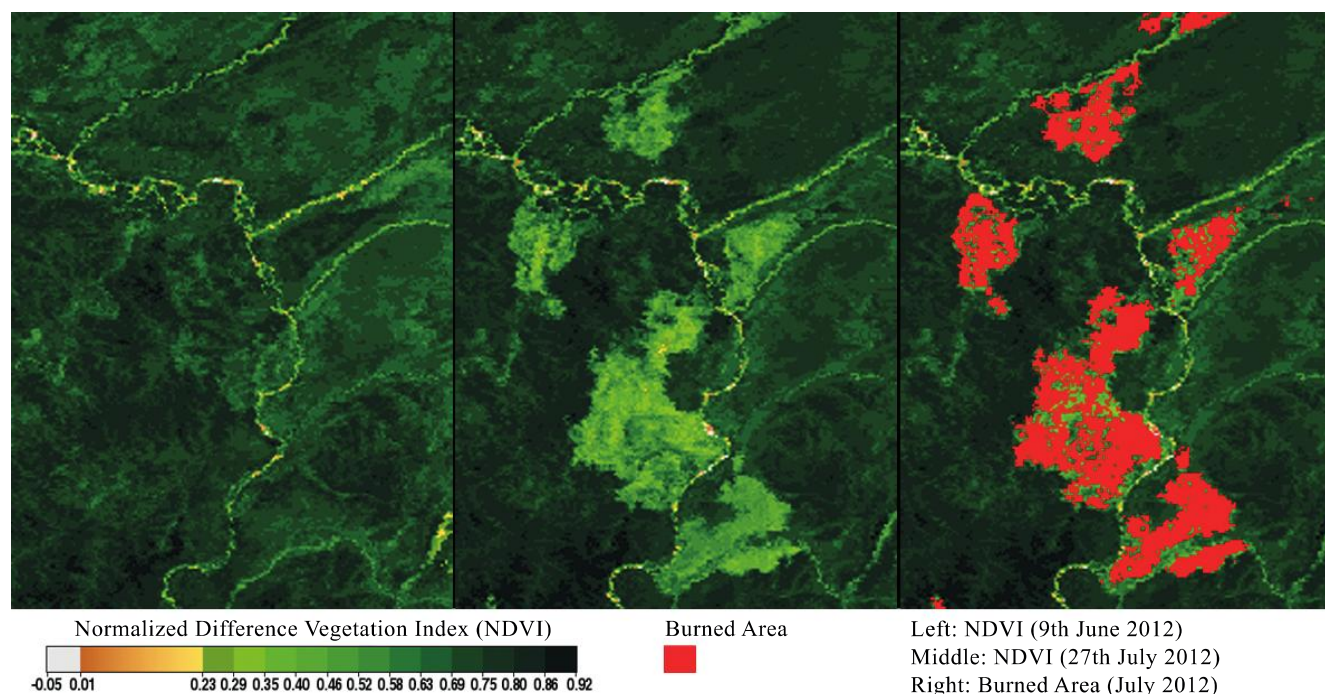**Table 8.** Automatic integrated products from MODIS sensor. EVI, enhanced vegetation index.

Short name	Platform	Resolution	Temporal coverage	Data period
Snow Cover				
MOD10CM	Terra/Aqua	0.05 °	2000/2002–present	Monthly
MOD10C2	Terra/Aqua	0.05 °	2000/2002–present	8-day
Land Surface Temperature				
MOD11C3	Terra/Aqua	0.05 °	2000/2002–present	Monthly
MOD11A1	Terra	1,000 m	2010–present	Daily
Vegetation Indices (NDVI, EVI)				
MOD13C2	Terra/Aqua	0.05 °	2000/2002–present	Monthly
MOD13C1	Terra/Aqua	0.05 °	2000/2002–present	16-day
MOD13A1	Terra	500 m	2010–present	16-day
Burned Area				
MCD45A1	Terra + Aqua	500 m	2000–present	Monthly

5.1. Automatic Data Integration

Depending on the available disk storage, data products can be integrated automatically for large regions. For Siberia, a few coarse resolution products from MODIS, as well as daily summaries of the climate data (GSOD), are integrated automatically. Based on the area of Siberia stored in the database, specific products are integrated and updated on a daily basis. Table 8 lists the available datasets from MODIS that can be visualized, downloaded and plotted within the developed web portal. The aim of the automatic integration is to get an overview of the whole region, for example, providing data for disaster management, such as fire and burned area data. This information, in combination with vegetation indices, can show the effect of burned areas on vegetation (Figure 9). Based on this information, specific areas can be identified for further research. The GSOD is integrated automatically for 1,000 stations in Siberia. Around 400 stations are still active and updated daily by the

middleware services. This data can either be compared to the integrated MODIS land surface temperature or to additional climate information, depending on the area under investigation.

Figure 9. Combination of MODIS NDVI and MODIS burned area datasets (**Left**: previous-fire NDVI; **Middle**: post-fire NDVI, **Right**: burned areas) near Yakutsk, Russia.

Normalized Difference Vegetation Index (NDVI)

-0.05 0.01   0.23 0.29 0.35 0.40 0.46 0.52 0.58 0.63 0.69 0.75 0.80 0.86 0.92

Burned Area

Left: NDVI (9th June 2012)
Middle: NDVI (27th July 2012)
Right: Burned Area (July 2012)

5.2. On-Demand Data Integration

On-demand processing was developed for the integration of higher resolution data that is only needed for specific regions and that can be used to support research projects through the provision of data. An administrator of the system (in the future, the users themselves) can register study areas and required datasets from the web portal. Based on the registered areas and datasets, the middleware service integrates the requested datasets into the middleware database, which were also clipped to the specified region. The processed datasets can then either be downloaded or directly used by other clients using the provided visualization and download services. For each processing request, the user can download a logfile. In this logfile, all downloaded files are linked, and further processing commands are listed to follow the processing steps afterwards (Figure 10).

The European FP7 research project, ZAPÁS, uses SIB-ESS-C for the analysis of vegetation indices from MODIS Terra with 250 m (spatial) and 16-day (temporal) resolutions for the entire available time-period (since 2000) in their study area. Another project, Forest Dragon 2, was interested in using the MODIS normalized vegetation difference index (NDVI). They used SIB-ESS-C to process MODIS NDVI with a 1 km spatial resolution on a monthly basis for all of 2005. This dataset was then referenced and integrated into their own geoportal (Forest Dragon Geoportal) via the OGC-WMS service provided by the SIB-ESS-C. Figure 11 shows the screenshot of both web-based systems.

Figure 10. Excerpt of a logfile of on-demand processing (processing of a single time step).

```
#processing MOD13A3 for study area: Forest Dragon

#download files
#needed tiles for all study areas: h28v04, h25v03, h24v03, h25v04, h27v04, h26v03, h26v04
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h28v04*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h25v03*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h24v03*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h25v04*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h27v04*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h26v03*.hdf*
wget -c ftp://e4ftl01.cr.usgs.gov/MODIS_Composites/MOLT/MOD13A3.005/2005.01.01/*.h26v04*.hdf*


#dataset extraction
#processing MOD_Grid_monthly_1km_VI:1 km monthly NDVI
gdal_translate -a_nodata -3000 -of GTiff -a_srs "+proj=sinu +lon_0=0 +x_0=0 +y_0=0 +a=6371007.181
+b=6371007.181 +units=m +no_defs"
HDF4_EOS:EOS_GRID:"MOD13A3.A2005001.h26v04.005.2007355120626.hdf":"MOD_Grid_monthly_1km_VI:1 km
monthly NDVI" MOD13A3.A2005001.h26v04.005.2007355120626.NDVI.tif
[... gdal_translate is being executed for each tile ...]

#merge for selected study area
gdal_merge.py -of GTiff -o merge.tif *.h28v04.*.NDVI.tif *.h27v04.*.NDVI.tif *.h25v04.*.NDVI.tif
*.h24v03.*.NDVI.tif *.h25v03.*.NDVI.tif *.h26v03.*.NDVI.tif *.h26v04.*.NDVI.tif

#set projection
gdalwarp -t_srs EPSG:4326 merge.tif MOD13A3.A2005001.NDVI.6.tif

#clip data to selected study area
gdalwarp -of GTiff -cutline PG:"dbname=sibessc" -csql "select ST_SetSRID(geom, 4326) from
userdata.study_areas WHERE uid='6'" -cblend 0 -crop_to_cutline MOD13A3.A2005001.NDVI.6.tif
MOD13A3.A2005001.NDVI.6.clipped.tif

#compress data
gdal_translate -co COMPRESS=PACKBITS MOD13A3.A2005001.NDVI.6.clipped.tif MOD13A3.A2005001.NDVI.
6.clipped.compressed.tif

#move to final destination
mv MOD13A3.A2005001.NDVI.6.clipped.compressed.tif /data2/userdata/user_1/task_239
```
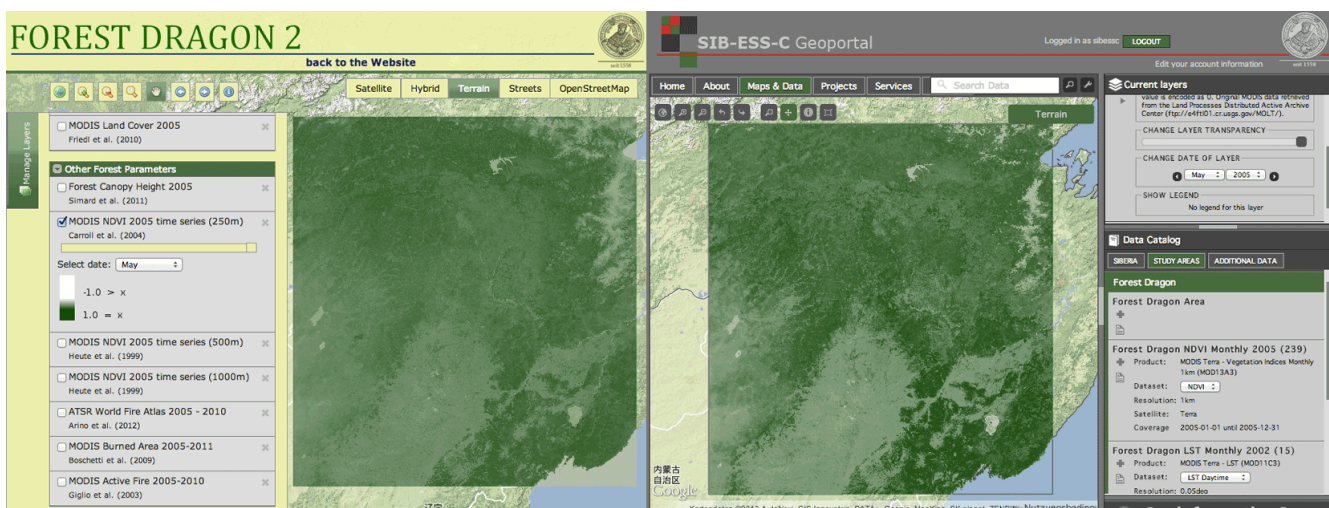
Figure 11. Forest Dragon 2 Geoportal (**Left**) shows the processed datasets from the SIB-ESS-C (**Right**).

A further on-demand link exists for the Integrated Surface Database. This dataset can be retrieved either by the selection of a specific meteorological station or by running a task for a registered study

area. With the selection of a station, the user can define a specific day. Only the data from this day will then be returned to the user, and an HTML table preview of the data is shown through the web portal (Figure 12).

Figure 12. HTML preview of the retrieved data from Integrated Surface Database of the selected station, Ust-Maja (near Yakutsk), for the 20 July 2012.

	Air Temperature Observation				Atmospheri
Date	Temperature [°C]	Air Temperature Quality Code	Dew Point Temperature [°C]	Dew Point Temperature Quality Code	Sea Level Pressure [hPa
20120720 0000	19	1	13	1	1013
20120720 0300	21	1	14	1	1012
20120720 0600	25	1	14	1	1010
20120720 0900	26	1		9	
20120720 1200	22	1	15	1	1010
20120720 1500	20	1	15	1	1009
20120720 1800	17	1	15	1	1010
20120720 2100	18	1	15	1	1009

USAF: 249660   WBAN: 99999   Lat: 134.45   Lon: 60.383   Elevation: 170 m

6. Conclusion and Outlook

In accordance with the aims of the study, an operational multi-source data processing middleware for land monitoring was established with standard-compliant services for data visualization and distribution. Data from MODIS Land Team and from NSIDC, as well as two climate datasets from NCDC, were connected with the middleware. Additional processing steps were integrated to generate common data formats, such as GeoTIFF and standardized SI-based units for climate data. Since it provides standard-compliant services, the data can be used with other GIS clients or within the developed web portal. The present paper showed applications as to how the middleware services within the SIB-ESS-C can be used to foster the monitoring of land processes using remote sensing and *in situ* climate data. Without any processing by the user, datasets can be investigated according to land changes; additional information, such as climate data, can easily be added, making data middleware services powerful in combining datasets from multiple sources. Different datasets can then be compared visually within the SIB-ESS-C web portal. Further applications could be the provision of Earth observation data for users who are not familiar with using such kinds of data. They can just visit the web portal and explore the data visually. A main further aspect is the integration of web-based analysis of the time-series data with the possibility to identify changes of the environment in the past.

Standard-compliant data provision is no longer a critical topic, as software packages exist that can handle most common data formats. In combination with a spatial database, sub-selections of data based on parameters, such as time-series data with timestamps, is no longer prohibitively complex to implement. The open-source tools used facilitate the development of such spatial data infrastructures, especially for handling spatial data and making sure that it complies with standards. A more complex task is the provision of styling information for the visualization of data, as this information is not stored in the original data file and has to be generated manually for each dataset. Technical limitations did not occur during implementation, but further work would be helpful in investigating possible improvements in data storage: for example, raster databases [49,50] or no-SQL databases for storing and analyzing time-series data [51]. Data access could also be improved in ways, such as the implementation of the OGC Web Coverage Processing Service specification [52], which provides

processing during data access [53]. The automated data integration could be implemented, as long as all information required for data downloading and processing is stored in the database. This information is required for datasets that need to be integrated by the middleware service.

Advantages of the implemented methods in the system are the flexibility of how data is provided to the user by using OGC-based services and data in GIS-common formats and the possibility to integrate further time-series analysis tools. For the visualization of integrated data, it is necessary that data is stored in the system, as well. However, this leads to a main problem: data has to be downloaded from external systems and needs disk storage at the owner system. The speed of downloading depends also on the external server, and for big datasets, this needs the most time of data integration and processing. As an example, the data download for the ZAPÁS project took around two days; the processing, just around four hours. To overcome this main issue, the only solution could be that these processing steps (dataset extraction, format conversion, clipping, OGC-compliant data provision, *etc.*) are implemented on systems with direct access to the data. In an optimal case, the user or the data analysis system just has to download a time-series file and can, then, for example, provide further analysis tools. This is especially needed when such an infrastructure is adapted to higher resolution data, as data from the upcoming ESA Sentinel satellites. Otherwise, the time for data download would increase significantly.

State-of-the-art web technologies make it possible to develop web systems to support science, consulting and policy making, especially in the area of additional data processing. Such tools also support experts in diverse fields with easy data access and visualization tools. Easy-to-use web systems can provide data processing and visualization procedures; in this case, users do not need to download any datasets if they are only interested in getting an overview of land observations. This is made possible as data is provided through interfaces that allow for automated processing. With the automated data integration, time-series databases can be kept up to date. Data that is provided free of cost and is allowed to be distributed through other systems is a major driver of this movement. Further datasets, such as Landsat, NPP VIIRS, Spot Vegetation or the upcoming ESA Sentinel data, need service-based access to integrate them into the developed middleware database. This can be realized by web-based search, order and download services or just providing the data with FTP. Additional metadata and processing information—depending on the dataset—have to be added to the system to extract the requested datasets and to provide OGC-compliant services.

Future enhancements and improvements are planned for the multi-source data middleware and the SIB-ESS-C. For example, datasets from Landsat and Geoland 2 services [54] might possibly be used for the on-demand integration. Further research on the implementation of processing and analysis tools, such as climate data interpolation, calculation of time-series statistics and the integration of a GIS software backend with Grass GIS or ESRI ArcGIS, have to be conducted. The developed web portal will be extended, e.g., offering tools for users to export time-series information in different formats. At the end of the SIB-ESS-C development, the ideal outcome is an operational monitoring service that observes time-series values, providing information and alerts as defined by the user. For example, users could request an alert in the event of fire or burned or flooded areas or receive an automatic notification regarding the monthly burned area within a user-defined region. The technical development of SIB-ESS-C generally allows the technical transfer and implementation of data middleware services for other areas, in addition to Siberia, as the integrated datasets are globally and

freely available, and the above presented developments, databases and services are not specific to a particular area on Earth.

Acknowledgments

References

1. Townshend, J.R.; Brady, M.A.; Csiszar, I.A.; Goldammer, J.G.; Justice, C.O.; Skole, D.L. *A Revised Strategy for GOFC-GOLD*; GOFC-GOLD Report No. 24; Food and Agriculture Organization of the United Nations: Rome, Italy, 2006.
2. Groisman, P.Y.; Bartalev, S.A. Northern Eurasia earth science partnership initiative (NEESPI), science plan overview. *Global Planet. Change* **2007**, *56*, 215–234.
3. Lentile, L.B.; Holden, Z.A.; Smith, A.M.S.; Falkowski, M.J.; Hudak, A.T.; Morgan, P.; Lewis, S.A.; Gessler, P.E.; Benson, N.C. Remote sensing techniques to assess active fire characteristics and post-fire effects. *Int. J. Wildland Fire* **2006**, *15*, 319–345.
4. Franklin, S.E. *Remote Sensing for Sustainable Forest Management*; Lewis Publishers: Boca Raton, FL, USA, 2001.
5. Asner, G.P.; Alencar, A. Drought impacts on the Amazon forest: The remote sensing perspective. *New Phytol.* **2010**, *187*, 569–578.
6. Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Prog. Plann.* **2004**, *61*, 301–325.
7. Justice, C.O.; Vermote, E.; Townshend, J.R.G.; Defries, R.; Roy, D.P.; Hall, D.K.; Salomonson, V.V.; Privette, J.L.; Riggs, G.; Strahler, A.; *et al.* The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1228–1249.
8. Giri, C.P. *Remote Sensing of Land Use and Land Cover: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2012.
9. Acker, J.G.; Leptoukh, G. Online analysis enhances use of NASA earth science data. *Eos Trans. Amer. Geophys. Union* **2007**, *88*, doi: 10.1029/2007EO020003.
10. Jones, K. An introduction to data warehousing: What are the implications for network? *Int. J. Network Manage.* **1998**, *8*, 42–56.
11. Wang, J.; Li, C. Research on the Framework of Spatio-Temporal Data Warehouse. In Proceedings of the 24th International Cartographic Conference, Santiago de Chile, Chile, 15–21 November 2009.
12. Bédard, Y.; Han, J. Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery. In *Geographic Data Mining and Knowledge Discovery*; Miller, H.J., Han, J., Eds.; CRC Press: Boca Raton, FL, USA, 2009; pp. 45–68.
13. Di, L. GeoBrain—A Web Services Based Geospatial Knowledge Building System. In Proceedings of NASA's Earth Science Technology Conference, Palo Alto, CA, USA, 22–24 June 2004.

14. Jiang, J.; Yang, C.; Ren, Y.; Zhu, X. A Middleware-Based Approach to Multi-Source Spatial Data. In Proceedings of 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010; pp. 1–4.

15. Tang, J.; Ren, Y.; Yang, C.; Shen, L.; Jiang, J. A WebGIS for Sharing and Integration of Multi-Source Heterogeneous Spatial Data. In Proceedings of 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, Canada, 24–29 July 2011; pp. 2943–2946.

16. Mohammadi, H.; Rajabifard, A.; Williamson, I.P. Enabling Spatial Data Sharing through Multi-Source Spatial Data Integration. In Proceedings of GSID 11 World Conference, Rotterdam, The Netherlands, 15–19 June 2009.

17. Nativi, S.; Craglia, M.; Pearlman, J. The brokering approach for multidisciplinary interoperability: A position paper. *Int. J. Spat. Data Infrastr. Res.* **2012**, *7*, 1–15.

18. Nativi, S.; Khalsa, S.; Domenico, B.; Craglia, M.; Pearlman, J.; Mazzetti, P.; Rew, R. *The Brokering Approach for Earth Science Cyberinfrastructure*. Available online: http://semanticcommunity.info/@api/deki/files/13798/=010_Domenico.pdf (accessed on 1 March 2013).

19. Hüttich, C.; Schmullius, C.; Thiel, C.; Pathe, C.; Bartalev, S.; Emelyanov, K.; Korets, M.; Shvidenko, A.; Schepaschenko, D. ZAPÁS: Assessment and Monitoring of Forest Resources in the Framework of EU-Russia Space Dialogue. In *Let's Embrace Space Vol. 2*; Schulte-Braucks, R., Breger, P., Bischoff, H., Borowiecka, S., Sadiq, S., Eds.; European Union: Brussels, Belgium, 2012; Volume II, pp. 164–171.

20. Cechini, M.F.; Mitchell, A.; Pilone, D. NASA Reverb: Standards-Driven Earth Science Data and Service Discovery. In *AGU Fall Meeting Abstracts*; American Geophysical Union: San Francisco, CA, USA, 2011; Volume 1, p. 1406.

21. Aundeen, J.L.F.; Anengieter, R.L.K.; Uswell, M.D.B. U.S. geological survey spatial data access. *J. Geospatial Eng.* **2002**, *4*, 145–152.

22. *National Climatic Data Center Climate.gov: Data & Services*. Available online: http://www.climate.gov/#dataServices (accessed on 6 March 2013).

23. De Freitas, R.M.; Arai, E.; Adami, M.; Ferreira, A.S.; Sato, F.Y.; Shimabukuro, Y.E.; Rosa, R.R.; Anderson, L.O.; Friedrich, B.; Rudorff, T. Virtual laboratory of remote sensing time series: Visualization of MODIS EVI2 data set over South America. *J. Comput. Interdiscipl. Sci.* **2011**, *2*, 57–68.

24. Group on Earth Observations GEO. *What is GEOSS?* Available online: http://www.earthobservations.org/geoss.shtml (accessed on 1 March 2013).

25. Christian, E.J. GEOSS architecture principles and the GEOSS clearinghouse. *IEEE Syst. J.* **2008**, *2*, 333–337.

26. Group on Earth Observations. *The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan*. Available online: http://www.earthobservations.org/documents/10-Year Implementation Plan.pdf (accessed on 1 March 2013).

27. *World Meteorological Organization Resolution 40 (Cg-XII)*. Available online: http://www.wmo.int/pages/about/ Resolution40.html (accessed on 2 March 2013).

28. Justice, C.; Townshend, J.R.; Vermote, E.; Masuoka, E.; Wolfe, R.; Saleous, N.; Roy, D.; Morisette, J. An overview of MODIS Land data processing and product status. *Remote Sens. Environ.* **2002**, *83*, 3–15.

29. Xiong, X.; Chiang, K.; Sun, J.; Barnes, W.L.; Guenther, B.; Salomonson, V.V. NASA EOS Terra and Aqua MODIS on-orbit performance. *Adv. Space Res.* **2009**, *43*, 413–422.

30. Wolfe, R.E.; Roy, D.P.; Vermote, E. MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1324–1338.

31. *NASA Land Processes Distributed Active Archive Center (LP DAAC) MODIS Overview*. Available online: https://lpdaac.usgs.gov/products/modis_overview (accessed on 1 March 2013).

32. *NASA Land Processes Distributed Active Archive Center (LP DAAC) Data PoolI*. Available online: https://lpdaac.usgs.gov/get_data/data_pool (accessed on 1 March 2013).

33. *National Snow and Ice Data Center (NSIDC) Order Data.* Available online: http://nsidc.org/data/modis/order_data.html (accessed on 1 March 2013).

34. Lott, N. *Global Surface Summary of Day*. Available online: http://www.ncdc.noaa.gov/cgi-bin/res40.pl? page=gsod.html (accessed on 1 March 2013).

35. Lott, N.; Vose, R.; Del Greco, S.A.; Ross, T.F.; Worley, S.; Comeaux, J. The Integrated Surface Database: Partnerships and Progress. In Proceedings of the 24th Conference in IIPS, New Orleans, LA, USA, 20–24 January 2008.

36. Nebert, D.; Whiteside, A.; Vretanos, P.P. *OpenGIS Catalogue Services Specification*; Open Geospatial Consortium Inc.: Wayland, USA, 2007.

37. De Beaujardiere, J. *OpenGIS Web Map Server Implementation Specification*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2006.

38. Vretanos, P.A. *Web Feature Service Implementation Specification*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2005.

39. Baumann, P. *OGC WCS 2.0 Interface Standard—Core*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2010.

40. Bröring, A.; Stasch, C.; Echterhoff, J. *OGC Sensor Observation Service Interface Standard*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2012.

41. Schut, P. *OpenGIS Web Processing Service*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2007.

42. Neteler, M. *MODIS Sinusoidal GIS SHAPE Files*. Available online: http://gis.cri.fmach.it/modis-sinusoidal-gis-files/ (accessed on 4 March 2013).

43. Warmerdam, F. The Geospatial Data Abstraction Library. In *Open Source Approaches in Spatial Data Handling*; Hall, G.B., Leahy, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 2, pp. 87–104.

44. Obe, R.O.; Hsu, L.S. *PostGIS in Action*, 1st ed.; Manning Publications Co.: Stamford, CT, USA, 2011.

45. Cannata, M.; Antonovic, M. istSOS: Investigation of the Sensor Observation Service. In Proceedings of 1st International Workshop on Pervasive Web Mapping, Geo-processing and Services, Como, Italy, 26–27 August 2010.

46. Cannata, M.; Molinari, M.; Pozzoni, M. istSOS Sensor Observation Management System: A Real Case Application of Hydro-Meteorological Data for Flood Protection. In Proceedings of International Workshop of "The Role of Geomatics in Hydrogeological Risk", Padua, Italy, 27–28 February 2013.

47. Cannata, M.; Antonovic, M. *Tutorial: Using istSOS*. Available online: http://istgeo.ist.supsi.ch/software/istsos/ tutorial.html (accessed on 8 March 2013).

48. *ISO 19115:2003 Geographic Information—Metadata*; International Organization for Standardization: Geneva, Switzerland, 2003.

49. Maduako, I. PostGIS-Based Heterogeneous Sensor Database Framework for the Sensor Observation Service. In Proceedings of FOSS4G Prague 2012, Prague, Czech Republic, 21–25 May 2012; Čepek, A., Ed.; Czech Technical University in Prague: Prague, Czech Republic, 2012; pp. 55–71.

50. Misev, D.; Baumann, P.; Seib, J. Towards large-scale meteorological data services: A case study. *Datenbank-Spektrum* **2012**, *12*, 183–192.

51. Gallaher, D.; Grant, G. Data Rods: High Speed, Time-Series Analysis of Massive Cryospheric Data Sets Using Pure Object Databases. In Proceedings of 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 5297–5300.

52. Baumann, P. *Web Coverage Processing Service (WCPS) Language Interface Standard. OGC 08-068r2*; Open Geospatial Consortium Inc.: Wayland, MA, USA, 2009.

53. Baumann, P. The OGC web coverage processing service (WCPS) standard. *GeoInformatica* **2009**, *14*, 447–479.

54. Smets, B.; Lacaze, R.; Pacholczyk, P.; Makmahra, H.; Baret, F.; Camacho, F.; Claes, P. Towards an Operational Delivery of a Time-Series of Essential Climate Variables for Global Land Monitoring. In Proceedings of 31st EARSeL Symposium 2011; Halounová, L., Ed.; Czech Technical University in Prague: Prague, Czech Republic, 2011; pp. 205–212.