

ATMOSPHERIC DATA ACCESS FOR THE GEOSPATIAL USER COMMUNITY (ADAGUC)

Nikée E. Groot^{a*}, John van de Vegte^b, Wim-Jan Som de Cerff^b, Gijsbertus H. J. van den Oord^b, Raymond Sluiter^b, Ian A. van der Neut^b, Maarten Plieger^b, Richard M. van Hees^c, Richard A.M. de Jeu^d, Michael E. Schaepman^a, Marc R. Hoogerwerf^a, Ben Domenico^e, Stefano Nativi^f, Olga V. Wilhelmi^g

^aWageningen University and Research Centre (WUR), Costerweg 50, 6701 BH, Wageningen, the Netherlands

^bRoyal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, 3732 GK, De Bilt, The Netherlands

^cNetherlands Institute for Space Research (SRON), Sorbonnelaan 2, 3584 CA, Utrecht, the Netherlands

^dFaculty of Earth and Life Sciences, Free University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands

^eUNIDATA, UCAR Foothills Lab 4, 3300 Mitchell Lane, Boulder, CO 80301 USA

^fInstitute of Methodologies for Environmental Analysis (IMAA), C.da S. Loja - Zona Industriale, I-85050 Tito Scalo (PZ), Italy

^gThe National Center for Atmospheric Research (NCAR), 3450 Mitchell Lane, Boulder, CO 80301, USA

KEY WORDS: ADAGUC, GIS, Atmospheric and Meteorological Datasets, OGC, WMS, WFS, WCS, RGI

ABSTRACT:

Historically the atmospheric and meteorological communities are separate worlds, with their own data formats and tools for data handling, which makes sharing of these data with the geospatial communities difficult and cumbersome. These information sources are becoming of increasing interest outside of these communities at present, because of the continuously improving spatial and temporal resolution of e.g. model and satellite data and the interest in historical datasets. New user communities that use geographically based datasets in a cross-domain manner are emerging. This development is supported by the progress made in Geographical Information System (GIS) software. The current GIS software is not yet ready for the wealth of atmospheric data, although the faint outlines of new generation software are already visible: support of HDF, NetCDF and an increased understanding of temporal issues are only a few of the hints. To aid in this trend, the ADAGUC project aims to reduce the need for scientists to invent their own converter tools. Selected space borne atmospheric datasets will be made accessible for GIS software. Both the geospatial and the atmospheric user communities are involved intensively in the project, to obtain a high fitness for use. The deliverables of this project are: Open Source conversion tools, selected atmospheric datasets in a GIS-friendly format and a web service to demonstrate the usability of the above to the geospatial and atmospheric community. Dissemination of the results is pursued through publications, workshops and (inter)national cooperation.

1. INTRODUCTION

The ADAGUC project (Atmospheric Data Access for the Geospatial User Community; see <http://adaguc.knmi.nl>) aims at reducing the need for users to invent their own converter and mapping tools. Selected space borne atmospheric datasets will be made accessible by GIS to allow easy data comparison, resampling, selection, manipulation and visualization. Representatives of the (inter)national user communities are strongly involved in the project to guarantee proper accommodation of the user requirements. Besides easy access to datasets for non-atmospheric and -meteorological communities, the international operational meteorology and meteorological research can profit from the introduction of GIS technology (see COST 719; <http://www.knmi.nl/samenw/cost719>). ADAGUC will focus on interoperability and harmonization of data resources such that a 'GIS-enabled' user can work with these datasets. ADAGUC serves as a pilot for applying GIS-technology in meteorological and climate datasets.

The ADAGUC project's expected results will be: Open Source conversion tools for conversion of selected atmospheric datasets into an Open Standard GIS format, published atmospheric

datasets in GIS format and a web service to demonstrate the usability of the above to the geospatial and atmospheric communities. Dissemination of the results is pursued through publications, workshops and international cooperation. The challenge is to create an environment that has the potential to become the next generation operational solution, dealing with international frameworks, standards, and cross domain end-users.

In the ADAGUC project the ESA's PSS-05 Lite standard for software development is used. It recognizes the following phases: 1) use case definition, 2) user requirements definition, 3) detailed design, 4) implementation, 5) integration and tests. During phase 1 and 2 the user communities were consulted. With these users' results, the spatial data infrastructure, based on OGC compliant web services, is currently being developed, to provide the atmospheric datasets. These web services include Web Mapping Services (WMS) for online visualization, Web Feature Services (WFS) for downloading vector data and Web Coverage Services (WCS) for downloading raster data. The ADAGUC project will be finished at the end of 2008 and the results will be presented in the final workshop on December 4th and 5th at the VU University in Amsterdam, the Netherlands.

* nikee.groot@wur.nl; phone +31 317 481552; fax: +31 317 419000; <http://adaguc.knmi.nl>

2. THE WORLD OF GEO-ICT

Generally (geo)data usage is accomplished by tools that are provided by third parties, like companies and governmental institutes specialized in geospatial solutions. Other stakeholders work on the standardization of data formats, metadata descriptions and services that can be used to exchange geospatial information. The most relevant developments in tooling, standards and services are discussed below.

2.1 Tools

2.1.1 Google Earth: Google Earth has made the public aware of the importance of geospatial information. As a consequence, third parties, commercial and non-profit organizations, are exploring the possibilities to present their own products into the Google Earth system. A wealth of geo-information is becoming available via Google Earth, like holiday pictures, hotels, forest fires, earthquakes and meteorological forecasts. An example with Sciamachy data can be found at the SRON website (http://www.sron.nl/~krijger/google_earth/google_earth.html; Figure 1). Google Earth uses the KML (Keyhole Markup Language) standard. This standard is under control of Google and is turned into a propriety format, thereby limiting the development of an open user community. Moreover, the use of Google Earth for scientific and professional applications is limited despite its important role for GIS-awareness among the general public.

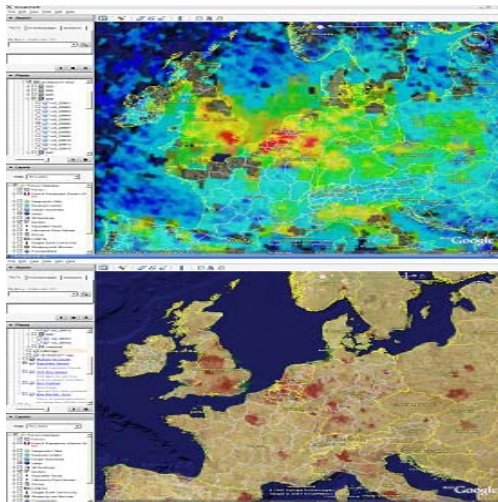


Figure 1. Google Earth: Monthly average of Sciamachy NO₂ tropospheric columns over Europe (above) and the population density for the same area (below).

2.1.2 Geographical Information Systems: A well established set of GIS tools is a necessary framework for dealing with different kinds of spatial data. Since the 1960's GIS has evolved from digital cartography into geo-ICT and offers standardized ways to store, process, analyze and visualize spatial data. The current commercial and free GIS software cannot deal with the large data volumes and the temporal aspects of atmospheric data yet, although these issues are being addressed because of pressure from user interest groups. A good example of this is the support of NetCDF (Network Common Data Form) COARDS (Cooperative Ocean/atmosphere Research Data Service) and CF (Climate and Forecast Metadata Convention) in ArcGIS (see: <http://www.esri.com>) 9.2. NetCDF is a set of software libraries that enable the (technical) information interchange unambiguously across different computer platforms. COARDS and CF are metadata definitions (descriptions) for promoting the exchange of information stored in NetCDF files by using a common standard. The incorporation of NetCDF in ArcGIS was promoted by the National Centre for Atmospheric Research (NCAR; see: <http://www.ncar.ucar.edu>) and ESRI's Atmospheric Special Interest Group (see: <http://www.gis.ucar.edu/sig>).

2.1.3 ESRI ArcGIS: ESRI provides professional solutions to deal with geospatial problems. Their software products serve the 'earth bound' communities like governmental organizations, cadastre, road and traffic organizations, petrochemical industries and the soil and vegetation related institutes. ESRI's next step is to fulfill the wishes of the atmospheric data users; importing of atmospheric data formats, satellite view transformations and time resolutions from centuries to seconds are being investigated by ESRI for future releases.

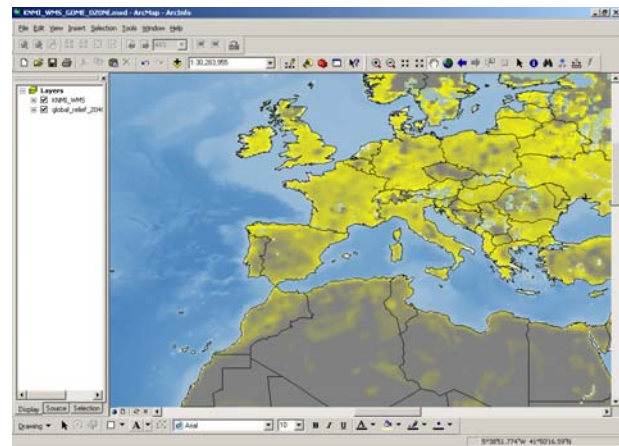


Figure 2. Viewing Soil Moisture derived from AMSR-E measurements by the VU in ArcMap using an OpenGIS WMS.

2.2 Standards

2.2.1 The Open Geospatial Consortium (OGC): The OGC (see: <http://www.opengeospatial.org>) deals with standardization of all geo related issues; their focus is on defining specifications by open working groups in order to create a common ground for interoperability. The interpretation of geographical information thus becomes identical for all OGC compliant software, making it easier to handle and exchange datasets. OGC is for the interoperability of GIS what the World Wide Web Consortium (W3C) is for internet technology.

2.2.2 Infrastructure for Spatial Information in the European Community (INSPIRE): On 25 April 2007 the European Parliament adopted the INSPIRE directive (see: <http://www.ec-gis.org/inspire>), which aims at establishing coordination between the users and providers of spatial information to ensure that the information and knowledge from different sectors can be combined. The directive applies to spatial data held by or on behalf of public authorities and to the use of spatial data by public authorities in the performance of their public tasks. In the forthcoming years INSPIRE will define the standards on data, metadata and services to which the European spatial information infrastructure must comply. INSPIRE implementation rules should be based, where possible, on international standards; in fact, these rules shall take account of relevant, existing international standards and user requirements, in particular with relation to validation metadata. The guidelines of all the standardizing bodies (e.g. WMO) require an overall vision and adequate participation to evaluate the potential impacts for meteorological institutes, space agencies and data providers in general.

2.2.3 Metadata, ISO19100 series: Optimal search and order capabilities rely on the use of homogeneous and uniform descriptions of metadata. ISO (<http://www.iso.org>) has developed a series of standards (the 19100 series) for defining, describing and managing geographic information that includes a standard for metadata. This standard has been adapted by organizations like the WMO and the OGC. INSPIRE recommends the use of this standard for metadata. In the Netherlands GeoNovum (<http://geonovum.nl>) has defined the Dutch profile for the Metadata Standard for Geography based on the ISO standards ISO19115 and ISO19139. Within the meteorological community ISO formatted datasets are becoming scantily available as most datasets are uniquely structured. This includes time series (in-situ measurements) from before the 1900s until now, model runs for climate and weather analysis, satellite measurements since the 1970s, and all sorts of other datasets that are created from meteorological phenomena. Therefore it takes time and effort to implement these standardizations.

2.2.4 GIS compliant (meta-)data formats: The exchange of information is more than moving data from A to B. One has to ensure that the data is interpreted in a correct manner by all users and is interpreted independent of the user community. Data formats like NetCDF (<http://www.unidata.ucar.edu/software/netcdf>) and HDF (<http://hdf.nsa.uiuc.edu>) are self-explanatory in terms of file content, which is in sharp contrast with the older binary files. The latter suffer from the fact that unfamiliar (remote or new) users cannot easily relate content and description, rendering them useless if additional information is not available. A similar aspect applies to the meta-data of a data collection. When meta-data descriptions are not uniform in terms of content and format, searching and ordering of data is cumbersome and often results in incomplete and erroneous results.

2.3 Services

2.3.1 Service Oriented Architecture (SOA): Service Oriented Architecture is a concept in which many relatively small, loosely coupled, distributed services interact to create a larger infrastructural service that is maintainable and scalable: a change in one service has low impact on the overall service / production chain. SOA focuses on the interfaces and not on the implementation inside a service which is the responsibility of the owner. SOA is the opposite of the traditional monolithic architectures in which scalability and additional functionality can have considerable (cost) impact. The OGC Web Feature Service (WFS) and Web Coverage Service (WCS) are good examples of the SOA concept.

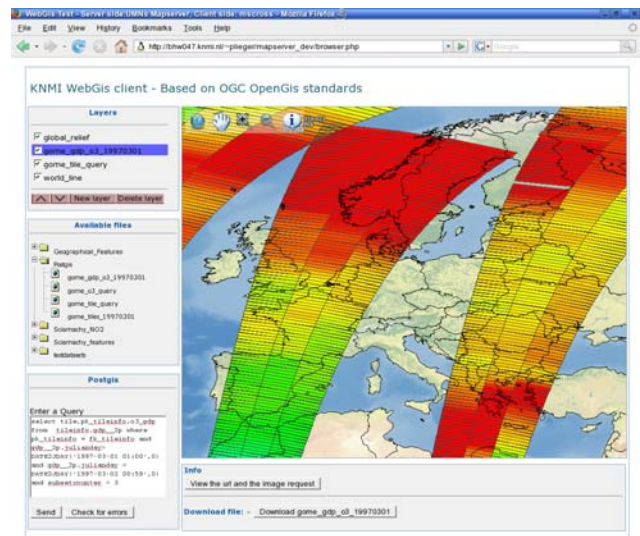


Figure 3. Example of an OGC based web service displaying features of ozone measurements from GOME, which have been stored in a PostgreSQL spatial database.

3. ADAGUC RESULTS

3.1 User Requirements and Use Cases

The first ADAGUC workshop was organized at KNMI on 3-4 October 2006. Four distinct user communities, based on their specific needs for reliable and flexible access to atmospheric data, were identified:

- Policy Makers are interested in long term averaged/forecast products and maps that are archived for easy access. Many (inter)national stakeholders provide atmospheric data to their governments for developing policies with respect to the social and political impact of i.e. climate change.
- Earth System Scientists and Atmospheric Scientists are interested in data of the highest possible quality and resolution, both archived and real-time. The atmospheric users are particularly interested in aerosols, clouds and trace gases like NO₂ and SO₂.
- GIS Users are interested in high quality and averaged data, including archived (offline) datasets. GIS users are characterized as being non-experts in the field of atmospheric research; their focus is on earthbound features and earth-atmosphere interactions.

- The risk assessment community is interested in mapped data, preferably in (near-)real time. These users are interested in products at an urban scale that are focused on industrial or natural calamities, like gas leakage or chemical fires. The combination of inter-comparable data sources provides vital input to crisis management teams.

These user communities have identified their needs in seven use cases, which have been translated to user requirements: several atmospheric data products (NO₂, CH₄ and cloud fraction) and weather model products (precipitation, wind, boundary layer height). These data products should be delivered on a high processing level: gridded and reanalyzed data (level 3/4). The user communities also indicated that they need an easy to use online viewing tool, including access by Google Earth (KML). Both user requirements document and use case document can be found on the ADAGUC website (<http://adaguc.knmi.nl>).

3.2 Web service infrastructure

For providing the temporal atmospheric and land dataset products, OGC compliant web services are used: WMS for visualization, WFS for download of vector data and WCS for downloading raster data. The UMN MapServer is used as a basis and other Open Source solutions like THREDDS when needed (Figure 4). The main challenge encountered is: the ADAGUC datasets are dynamic, temporal and huge, while most OGC solutions are optimized for small static datasets.

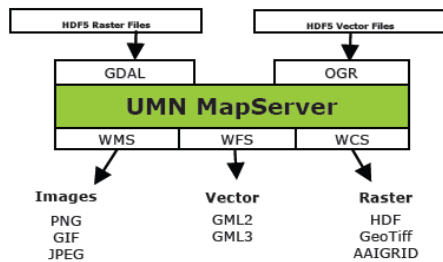


Figure 4. Conversion of atmospheric datasets to GIS

3.3 ADAGUC Data Products Standard

Besides ‘just’ providing data, an important requirement is a standardized, clearly defined and structured format. The HDF-5 ADAGUC data products standard has been defined for this purpose, which includes structured metadata (ISO 19115). All ADAGUC data will be archived in this format. Converters for converting these data into GIS friendly formats are developed (e.g. GeoTIFF, GML) within the project as are conversion tools for the most commonly used formats: Google’s KML, ESRI’s Grid/Shapefile, GeoTIFF and netCDF-3 (Figure 5).

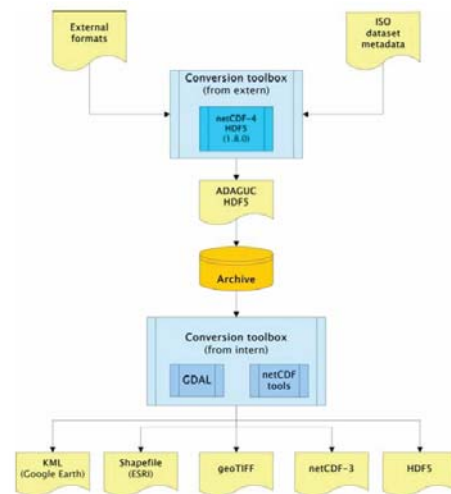


Figure 5. ADAGUC data flow: how data of an input product is written in ADAGUC file format using the netCDF-4 library. These products are stored in the NL-SCIA-DC archives. On user request, these products can be converted to various formats using the GDAL applicator.

4. CONCLUSIONS

Being able to exchange information according to international standards will provide an important scientific, social and economic benefit for different communities. In the forthcoming years new standards for data exchange and data services will be legally imposed by means of the INSPIRE directive. The ADAGUC project is a pilot that aims at preparing selected atmospheric and meteorological datasets for these upcoming changes. The project has a set of well-defined deliverables. Moreover the atmospheric community becomes more visible as a stakeholder in the GIS community and knowledge-sharing networks. The use of GIS also introduces a number of science-related challenges: mapping, gridding and projecting data must be done in such a way that the scientific integrity of the data is preserved. The ADAGUC project will identify these problems and attempts to resolve them.

ACKNOWLEDGEMENTS

The ADAGUC consortium consists of the following partners:

- Royal Netherlands Meteorological Institute (KNMI, lead),
- Netherlands Institute for Space Research (SRON),
- Wageningen University and Research Centre (WUR),
- VU University Amsterdam (VU),
- Institute for Marine and Atmospheric research Utrecht (IMAU),
- European Space Agency (ESA),
- National Center for Atmospheric Research (NCAR),
- Harvard University, Institute of Environmental Physics Heidelberg (IUP),
- UNIDATA programme,
- Institute of Methodologies for Environmental Analysis (IMAA)

ADAGUC is funded by the Space for Geo-Information Programme (RGI; see: <http://www.rgi.nl>)