

**From Measurement Data to Environmental Information:
MARQUIS - A Multimodal AiR Quality Information Service for the General Public**

Leo Wanner^{1,2}, Daniel Nicklaß³, Nadjat Bouayad-Agha², Bernd Bohnet⁴, Joachim Bronder⁵,
Franciso Ferreira⁶, Rainer Friedrich², Ari Karppinen⁷, Francois Lareau², Achim Lohmeyer⁸,
Alejandra Panighi⁹, Sergi Parisio¹⁰, Helmut Scheu-Hachtel¹¹, Joao Serpa¹²

¹Institució catalana de recerca i estudis avançats (ICREA) and

²Fundació Universitat Pompeu Fabra, Barcelona
leo.wanner@upf.edu

³Institute of Energy Economics and Rational Use of Energy (IER),
University of Stuttgart
dn@ier.uni-stuttgart.de

⁴Universität Stuttgart, Germany; ⁵Institute for Ecology of Industrial Areas, Poland;

⁶Universidade Nova de Lisboa, Portugal; ⁷Finnish Meteorological Institute, Finland;

⁸Ingenieurbüro Lohmeyer, Germany; ⁹Televisió Catalunya S.A., Spain; ¹⁰Servei Meteorològic de Catalunya, Spain; ¹¹Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg, Germany; ¹²YDreams, Portugal;

Abstract

This article introduces the Multimodal AiR Quality Information Service for General Public MARQUIS. The service, which is based on air quality measurements, is able to provide information on air quality in Europe in a way that is easily understandable for the general public. The goal of the project is presented below with a special focus on the role of the users. Furthermore the architecture of the system and the components are described.

1. Introduction

There is an objective need for high quality cross-border air quality information services for several reasons. First, air quality information related to topics that are relevant for health and well-

being (such as air and water quality, noise, and meteorological conditions) or to topics that are relevant for leisure time (such as snow quality, wind direction and strength, water temperature, etc.) possesses a great economic potential, which has hardly been exploited so far in the EU. Second, citizens are increasingly aware of the importance of environmental information, critically evaluating the impact of environmental conditions on both their every day life and long term health, and consulting environmental information sources for decision-making concerning their professional activities and leisure time. As a consequence, they are increasingly sensitive to the quality (and amount) of the information provided by public and commercial services on the subject. Third, air pollution does not stop at the border of the member states. In the case of unfavourable meteorological conditions, high concentrations of air pollutants in a border region of a member state directly affect citizens across the border. They must be informed but, in order to be useful, the information must be given in a way that takes into account their air quality perception model. Fourth, the EU member states are obliged to comply with environmental EU legislation such as the Directive 96/62/CE (European Commission 1996) on ambient air quality assessment and management and its daughter directives. According to this legislation, the authorities have to offer exhaustive environmental information to citizens in a comprehensive and self-explanatory form.

The traditional air quality information services as maintained by public institutions – usually environmental departments of regional governments and/or national or regional meteorological institutes – fail to provide adequate information. They primarily offer “raw” air pollutant concentrations in terms of tables and graphics. However, such tables and graphics are certainly not self-explanatory. Some display pictograms or colour scales for a qualitative (index-oriented) presentation of the concentrations of individual air pollutants. Although much more informative than tables and graphics, pictograms cannot provide the explanatory and advisory dimension needed in relation to air quality information. The need for new generation, user-tailored, air quality information services has recently been discussed in several scientific forums (Bøhler et al, 2000; Peinel et al., 2000; Johansen et al, 2001). However, so far, no such services have been available. MARQUIS’s operational air quality information service developed for a number of European regions is the first of its kind. The characteristic features of the service are:

- reference to a default user profile typology, with the option of a flexible individualization of each profile by the users;
- cross-border information;
- coverage of the major modern communication channels: web, email, mobile phone (SMS and WAP), TV, and printed media;
- coverage of the major air pollutant substances of each MARQUIS-region;
- use of advanced air quality assessment and interpretation models;
- use of cutting edge computational linguistics techniques for planning and generation of multimodal and multilingual material.

2. The MARQUIS scenario

The base scenario of MARQUIS can be covered by the following: (i) citizens are interested in information concerning the air quality in their home area as well as in neighbour or holiday re-

gions and generally in the air quality in the whole of Europe. (ii) The kind of information which the citizens are interested in depends on the different regions. (iii) The background information and the information needs of citizens differ a lot (e.g. those subject to certain diseases) but only a few are able to interpret “raw concentrations” correctly without further information and explanations. (iv) There is no general presentation mode for air quality information but it depends on the special nature of the information.

The basic scenario gives the outline for the MARQUIS service with its four major features: cross-border (Europe), user-oriented, interpretation and multilingual/ multimodal natural language processing. These features make MARQUIS a next generation, environmental information service. As far as we know, there is only one other service which also combines interpretation and speech processing for the generation of environmental information - AutoText UIS (Bohnet et al. 2001a), which is used by the LUBW¹. AutoText UIS has some limitations: It is focused only on one pollutant – ozone, the interpretation is limited and the power of the speech processing component is reduced.

The European dimension in MARQUIS

MARQUIS covers five European regions: Baden-Württemberg (Germany), Catalonia (Spain) Finland, Silesian Voivodeship (Poland, with a focus on the region of Katowice) and Portugal (with a focus on the region of Lisbon). For each of these regions the existing measurement network is used. This means that to some extent different pollutants are measured in each region but in general at least the most prominent components such as ozone (O₃), particulate matter (PM₁₀) and nitrogen dioxide (NO₂) are measured.

The user orientation in MARQUIS

MARQUIS provides air quality information tailored to the users depending on their interests and previous knowledge, via different communication platforms. Thereby the interests of professionals, e.g. from the political or health-care sectors, as well as those of the general public are addressed by this service. Another target user group is commercial information brokers such as newspapers, web portals or mobile service providers.

The interpretation concept in MARQUIS

Unlike traditional air quality services, MARQUIS is not restricted to the presentation of raw data, forcing the users themselves to filter and interpret the relevant information. In fact MARQUIS evaluates the data and extracts user-relevant information such as exceedances of thresholds. Associated with this evaluation is also an assessment of health impacts and advice to the users on how to react to the situation in order to reduce their own health risk.

The multimodal/ multilingual concept in MARQUIS

Besides the user-orientated interpretation, the preparation of information in different modes is of major importance. MARQUIS uses computer linguistic techniques for on demand generation of information. The service is able to provide textual information for all regions in Catalan, English,

¹ AutoText UIS was funded by the state ministry of environment and transport of Baden-Württemberg and developed by the Intelligent Systems Group, University of Stuttgart in cooperation with the LUBW – both MARQUIS partners.

Finnish, French, German, Polish and Portuguese. If it seems to be valuable, the textual information is enriched by graphics and tables. The use of different modes to provide information results from the understanding that there is a specific favoured mode for the presentation of different kinds of information in the various media. Thus, for media such as internet, mobile services and printed media as well as for radio and television, specific output formats are generated.

3. The role of the users in MARQUIS

Air quality information is requested for different purposes and by persons with different backgrounds. It is used among other things:

- for decision support including in traffic, health and local affairs or in commerce (e.g. as a basis for traffic interdictions and other measures for emissions reduction or health warnings, etc.);
- for the health sector to advise patients with cardiovascular or respiratory diseases;
- by citizens planning their leisure time and sport activities or taking care of their health (especially patients with diseases which are caused or impaired by air quality);
- by media and other commercial information brokers as a product with a certain value.

Thus the application as well as the cultural and knowledge background of the user is crucial for deciding which kind and amount of air quality information should be provided to the user and what the best mode is to do so. Obviously a medical doctor should get a different kind of information than a politician or a citizen, even when talking about the same pollutant. Furthermore one citizen may be interested in background information and a detailed presentation of the current situation in contrast to another citizen who is often interested only in health warnings if a threshold is exceeded.

For these reasons the concept of user typology occupies a central position for MARQUIS. The user typology is a hierarchically organized classification of several user profiles. These profiles are compiled on the basis of a large number of interviews in all five MARQUIS regions with interested parties including information brokers (newspapers, TV, mobile services, web portals), decision-makers (municipal and regional politics, executive officers), health sector (doctors and pharmacists) and citizens (patients and general public).

A user profile is defined by a set of parameters. The most important parameters are outlined in Table 1. The number of parameters is not limited, thus new profiles can be generated by defining new parameters.

Table 1: Some of the most prominent parameters for MARQUIS

Personal parameters of the user	
group	information broker health care politics individual
residence	<the place of residence of the user>
regions	<regions to be covered by MARQUIS for the user>
locations	<locations to be covered by MARQUIS for the user>
profession	<different options depending on the group of the user>
age	<relevant for individuals>

health profile	respiratory cardiovascular ...
medium	email SMS WAP internet TV newspaper ...
...	...
Information related parameters of the user	
pollutants	<pollutants the user is interested in>
info trigger	on demand regular trigger thresholds
info period	current current day last 24 hours last 7 days (or a combination of the four)
info type	index concentrations
index info	value value + index composition ...
pollutant info	all prominent
forecast	none short detailed
...	...

4. The layout of the MARQUIS system

Before talking about the single modules and their interaction it is useful to introduce the general function of the MARQUIS service. Therefore the schematic architecture of the MARQUIS system is displayed in Figure 1.

The service has a “two pipe” architecture: the data processing pipe and the user request pipe. The data processing pipe has the following functions:

- Monitoring of air pollutant concentrations and of meteorological conditions in the five MARQUIS-regions and execution of data quality assurance and air pollution forecast models.
- Delivery of the measured and forecasted data from the local databases to the MARQUIS-server.
- Assessment and interpretation of the data delivered with respect to their relevance to any of the EU and regional environmental legislation issues and to any of the MARQUIS-users; determining the primary meteorological and contextual influence on the measured and forecasted air quality (for explaining / justifying them). Air quality forecast models may also be run in this step when required.

The user request pipe looks as follows:

- Receiving an information request from a user via the MARQUIS-Client interface (this can be an automated periodic request or a single request).
- Selecting the content that is relevant for the user in question from the structure produced in the third step of the data processing pipe.
- Generating the discourse structure of the content to be conveyed to the user, determining the appropriate mode for the individual chunks of the content, and starting the corresponding information generators.
- Generating the information by the table, graphic and multilingual text generator.
- Conveying the generated information to the user using his/her preferred communication channel.

This two pipe architecture is necessary to separate the user overlapping parts of the process from the user request specific parts. When the data processing is triggered by the arrival of new measured concentration data, so the generation of the user-tailored information can be effected on demand. The reasons for this separation are (a) the response to user requests is time-critical and (b) it is not possible to prepare all information beforehand because of the variety of user interests and communication platforms.

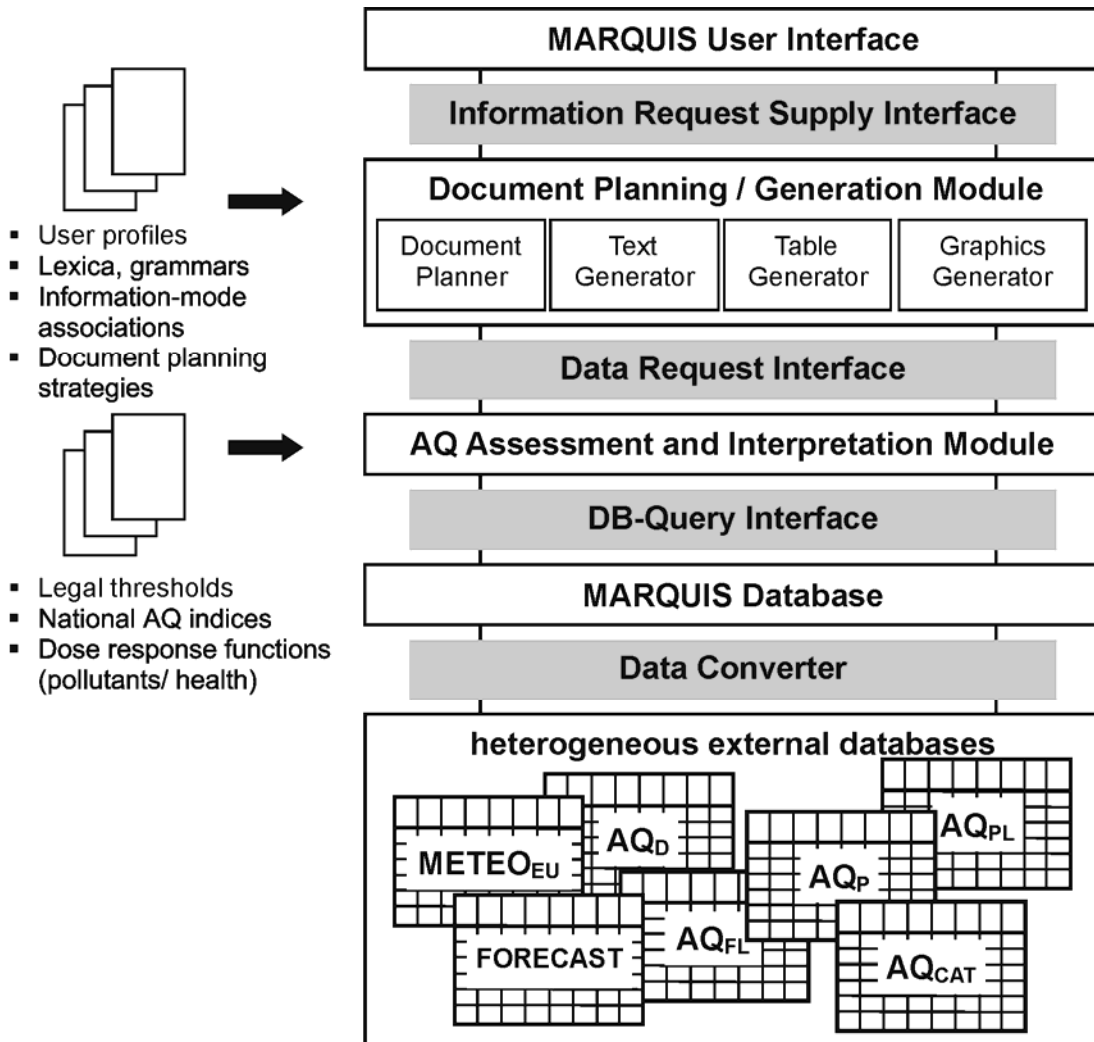


Fig. 1: Schematic architecture of the MARQUIS system

5. The input data and the MARQUIS database

Monitored data

The starting point of the MARQUIS system is measured air pollutant concentrations. These measurements are done by official agencies in the five European MARQUIS regions using automatic sensor networks that are already installed (see Fig.1 – AQ_{REGION}). Depending on the region, the sensor networks consist of up to 80 stations which measures air pollutant concentrations and in part also meteorological components.

The data measured vary among the regions and stations – as a rule there is ozone (O₃), particulate matter (PM₁₀), nitrogen dioxide (NO₂) and sulphur dioxide (SO₂). Some stations also include PM_{2.5}, carbon monoxide (CO) and ultraviolet solar radiation.

As well as the observed data, modelled data are also consulted within MARQUIS. These are the current and forecasted European-wide meteorology which is needed for the forecast module of MARQUIS in order to predict air pollutant concentrations for the next days and for interpretation reasons (see Fig.1 – METEO_{EU}).

All these data pass through a plausibility check and are afterwards transferred to the central server in a regionally and seasonally dependent frequency, every one to six hours. There they are harmonized and stored in the MARQUIS database. The MARQUIS database is thus the first European real time multi-pollutant database.

Forecast data

As well as current information, MARQUIS also provides forecasted air quality information. A considerable number of operational large scale models are available, describing and forecasting air pollutant concentrations all over EUROPE. The use of one large scale model for all MARQUIS regions would certainly have been very attractive. Even more so, since in contrast to local scale models, these models should be able to describe the influence of long range transport of pollutants. MARQUIS evaluated two of these models, EURAD (Jacobs, 2005), and THOR (Van Loon, 2004) but the evaluation was negative because the models are not able to provide sufficient data for fine grained air quality assessment.

Thus MARQUIS uses local scale models. For some of the regions, namely Portugal, Finland and Catalonia, MARQUIS was able to draw upon available local air quality forecast models either developed by members of the consortium or provided to them by external institutions. The range of approaches deployed is very largely region and component dependent. It goes from manual prediction (ozone – Baden-Württemberg), statistical models (Portugal, Baden-Württemberg) (Ferreira et al., 2000, 2004; Neto et al., 2005), physico-chemical Lagrange model (Catalonia) up to a kNN machine learning technique (Baden-Württemberg) (Lohmeyer et al., submitted 2007).

From measurement points to area data

Air quality is not a point phenomenon but a spatial one. In order to be able to provide air quality information not only for the measurement sites but for the whole MARQUIS area, an approach to transform point measurements into spatial data is needed. For areas with low pollution, the point measurements are transformed using some kind of interpolation into a raster map. Cities and more highly polluted areas are classified. This is done by subsuming areas with similar characteristics and consequently similar air quality situations. All of these classes are connected to representative measuring sites for this area. Thus city centres with similar volumes of traffic and industrial influences form one class which is represented by several measuring sites located in

such areas. For this process a strong detailed knowledge of this region is necessary which can be contributed by the regional data providers.

6. The air quality assessment and interpretation module of MARQUIS

In order to transform the raw data into user friendly information, some processing and interpretation are necessary. The first step is done by the air quality assessment and interpretation module. For this step additional information is necessary such as European and national directives concerning air quality pollutant concentration thresholds (European Commission, 2000) and examination of the effects of air quality on human health (World Health Organization, 2000). The thresholds thus extracted provide a basis for further interpretation.

The air pollutant concentration distribution curves are assessed for several criteria as follows:

- Exceedance of the above-named thresholds. This enables the inference if there is currently (was in the last hours/days) a health risk and how to act preferably to minimise it. The behaviour advisories are tailored to special user groups such as patients with a certain disease pattern or sportsmen.
- Air quality indices which are a simplified way of illustrating air quality on a more abstract scale than the actual pollutant concentrations; they are based on national or regional air quality guidelines.
- Significant extreme values and changes of pollutant concentrations require to get a qualitative overview about the situation in the last hours or days and to identify tendencies.
- Regional differences regarding the neighbouring measurement stations.
- Coherences between the meteorological or emission situation and air pollutant concentrations to give justifications for air quality.

7. The document planning and generation module of MARQUIS

The document planning and generation module of MARQUIS reacts to a user request and provides information tailored according to the language, background knowledge and requirements of the user. Therefore the module is separated into four sub modules with different functions: document planning, text generation, table generation and graphic generation.

The document planning module

The document planning module is responsible for two aspects (Bouayad-Agha et al., 2006). First it has to refine the data provided by the assessment and interpretation module by taking into account the user's requirements. This includes involving the user's language and chosen station or region as well as filtering the large amount of information provided by the assessment module for user relevant parts. Thematically linked information is grouped together.

Second it generates a "document plan". The document plan is a tree-like structure whose nodes display information units which are relevant for the user. Each content chunk is annotated with information concerning the appropriate realization mode. For example, a seven day trend of a pollutant concentration should be displayed as a graph, but the reason for a sudden drop of concentration value as a text. MARQUIS uses text as the main mode for providing information. The display medium is of course also relevant for the selection of the realization mode (e.g. an SMS

needs a different presentation than an article for newspapers) as well as for the selection of relevant information.

Information units that belong thematically together are linked in the document plan. The links are annotated with discourse relations such as elaboration, justification or reason. This annotation is mainly implemented automatically.

Information generation

The document generation module consists of three generators: the text generator, the graphics generator and the table generator. Given that the document plan contains the specification of the generation mode for the realization of the different fragments, in the first step of the generation, the document plan is split into mode-specific parts. Each generator processes its own part separately.

The multilingual text generator is capable of generating text in eight languages: Catalan, English, Finnish, French, German, Polish, Portuguese and Spanish. For all communication media, except SMS, the main text in these languages is generated by a full-fledged generation process, i.e., from scratch, using generation grammars. It is based on the MATE platform (Bohnet et al., 2001b; Bohnet, 2006).

Health risk warnings and legally predefined recommendations are represented as canned text chunks that are provided in the document plan and are introduced into the text as appropriate. For the generation of SMS messages, a template-based generator is used. This is because of the telegraphic style and restricted maximal length of SMSs.

The graphics and table generators are able to provide maps, graphs and tables for different media such as TV, internet or newspapers. These graphics and tables are referenced in the texts to increase the understandability.

8. The user interface of MARQUIS

In order to be able to reach various users, MARQUIS realised a large number of possible ways to access the information produced. The assessment and interpretation module and the document planning and generation modules are able to generate information in different levels of detail. Thus internet and mobile services can be covered as well as printed media and television solutions. As the number of users dictates the number of possible interfaces and this goes beyond the scope of the article, it will not be gone into detail here.

9. Conclusions

MARQUIS is the first next generation information service in the field of air quality. It combines for the first time advanced air quality assessment tools with state-of-the-art multimodal multilingual information generation – based on the first European real time multi-pollutant database.

The MARQUIS service is extendable in many ways. The database can be enhanced by new stations or complete sensor networks of other countries. New languages or pollutants can be introduced. Finally the service can be applied to new fields of environmental information such as water quality, noise or pollen.

The project was funded as EDC-11258 by the eContent programme of the European Community.

References

- Bøhler, T., K. Karatzas, G. Peinel, T. Rose, and R. San José (2000): Providing Multi-Modal Access to Environmental Data – Customizable Information Services for Disseminating Urban Air Quality Information in APNEE. In: Proceedings of the 22nd Urban and Regional Data Management Symposium.
- Bohnet, B. et al. (2001a): Autotext UIS: Automatische Produktion von Ozonkurzberichten im Umweltinformationssystem Baden-Württemberg. In: Proceedings of the Workshop Hypermedia und Umweltschutz.
- Bohnet, B., Langjahr, A., Wanner, L. (2001b): A development environment for an MTT-based sentence generator. In: Proceedings of the First International Conference on Natural Language Generation.
- Bohnet, B. (2006): Textgenerierung durch Transduktion linguistischer Strukturen. Köln: DISKI 298.
- Bouayad-Agha N., L. Wanner and D. Nicklaß (2006): Discourse Structuring of Dynamic Content. In: Proceedings of SEPLN'2006. Zarragoza, Spain.
- European Commission (1996): Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management. Official Journal (L 296): 55-63.
- Jakobs, H., M. Memmesheimer, A. Ebele (2005): Daily Forecast of Air Quality over Europe with the EURAD Model System. Geophysical Research Abstracts, Vol. 7, 10739, 2005 SRef-ID: 1607-7962/gra/EGU05-A-10739.
- Johansen, P.H., K. Karatzas, J.E. Lindberg, G. Peinel, T. Rose (2001): Citizen-Centered Information Dissimination with Multimodal Information Channels. In: Sustainability in the Information Society, Environmental Informatics 2001. 15th International Symposium, Informatics for Environmental Protection. Zürich, Switzerland.
- Lohmeyer, A., I. Düring, M. Giereth, T. Hoffmann, D. Nicklaß, M. Klein, H. Scheu-Hachtel, C. Sörgel und L. Wanner (submitted 2007): Kurzfrist-Feinstaub-Immissionsprognose für MARQUIS und LÜSA Sachsen-Anhalt. Gefahrstoffe und Reinhaltung der Luft.
- Peinel, G., T. Rose and R. San José (2000): Customized Information Services for Environmental Awareness in Urban Areas. In: Proceedings of the 7th World Congress on Intelligent Transport Systems.
- Van Loon, M., M.G. M. Roemer and P. J. H. Bultjes (2004): Model Inter-Comparison in the framework of the review of the Unified EMEP model. TNO Report R2004/282.
- World Health Organization (2000): Air Quality Guidelines – Second Edition. WHO Regional Publications, European Series (91).