

## Chapter XI

# Explorative Data Analysis Support for Environmental Epidemiology

V. Kamp  
University of Oldenburg, Germany

*Applications like environmental information systems define various scenarios of multidimensional data analysis, and they require special efforts concerning data representation, storage and processing. The project CARLOS (Cancer Registry Lower-Saxony) developed the Epidemiological and Statistical Data Exploration System (CARESS) to support multidimensional analysis of health data. The system is based on an architecture that focuses on extensive interoperability between a database management system and several analysis and visualisation tools. As spatial and statistical aspects of the data play an important role, CARESS provides special support for the integration of both.*

## INTRODUCTION

The objective of this chapter is to present an advanced approach to support explorative data analysis in an application scenario of environmental epidemiology. Requirements like a short response time are determined by the extremely explorative nature of the analysis process. It is characterised by multiple interactive and iterative analysis steps of complex query specifications mixed with the application of advanced statistical and epidemiological computations and result visualisations. Exploration means above all the integration of previous analysis results as input to consecutive analysis tasks (Berthold and Hand, 1999) which exceeds, especially in combination with the needs for specialised epidemiological methods, the capabilities of commercial database systems and geographic information systems.

Applications like environmental information systems define various scenarios of multidimensional data analysis, and they require special efforts concerning data

representation, storage and processing. OLAP- (On-line Analytical Processing) (Ruf, 1997; Shoshani, 1997; Lehner et al., 1997) and KDD- (Knowledge Discovery in Databases) techniques are relevant parts of information analysis and presentation. For example the projects EOSDIS (Earth Observing System Data and Information System) and CARLOS (Cancer Registry Lower-Saxony) (Appelrath et al., 1996) show characteristics of the intended application area. In this context, complex data types and adaptive processing techniques are central aspects to provide interoperability and cooperation between database systems and visualization and analysis tools.

CARLOS aims at providing software support for all important aspects of cancer registration. These are data collection, data security and environmental epidemiological research. The latter is based on exploratory analysis of cancer data (Rothman, 1986; Jensen et al., 1991), where the objective is to detect possible health risks by analysing the spatiotemporal distribution of health data. For example, in order to detect an increased cancer rate around a nuclear power station, geographic information is needed. In general, health data has to be related with various forms of environmental data. The CARESS software system supports the necessary multidimensional analysis of all these forms of data by advanced analysis and database technology (Appelrath et al., 1994). Its architecture focuses on extensive interoperability between a database management system and several analysis and visualisation tools. Thus, our approach is application-driven, but the developed concepts should not and will not be restricted to the specific application. In contrast, they will fit for a huge class of scientific data processing and analysing applications having similar requirements in common.

## **BACKGROUND**

To overcome cancer is not only still an immense medical task but also an increasing interdisciplinary task. The main aim of investigating causality is to improve prevention. To achieve this, computer supported population-based cancer registries form an important foundation. They improve methodical founded cancer research providing the possibility to manage huge and differentiated data about cancer diseases, death and complementary health-related and environmental data (Jensen et al., 1991). Besides, they offer new chances and methods to support epidemiological research for monitoring the development of diseases, especially cancer, by modern analysis and database technology (Appelrath et al., 1994) and appropriate interactive visualisation mechanisms.

### **Explorative analysis of cancer data**

Indispensable for modern health reports are descriptive epidemiological studies used to evaluate the data. Descriptive epidemiology is population-based and not related to individual cases. It deals with the occurrence of diseases, their respective accumulations, especially their spatial and/or temporal distributions in comparison to standardised populations. A continuous description of the development of cancer diseases in the population depends on an operating epidemiological cancer registry and is based on the periodical computation of quantified indicators. Indicators like mortality describe the health state of the population and are strongly related to the

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/explorative-data-analysis-support-environmental/18535](http://www.igi-global.com/chapter/explorative-data-analysis-support-environmental/18535)

## Related Content

---

### Expanding the Hydroinformatics Agenda: Information and Inequality behind Water Problems

Antonio A. R. Ioris (2011). *Handbook of Research on Hydroinformatics: Technologies, Theories and Applications* (pp. 1-21).

[www.irma-international.org/chapter/expanding-hydroinformatics-agenda/45437](http://www.irma-international.org/chapter/expanding-hydroinformatics-agenda/45437)

### A Comprehensive and Practical Green ICT Framework

Graeme Philipson (2011). *Green Technologies: Concepts, Methodologies, Tools and Applications* (pp. 186-200).

[www.irma-international.org/chapter/comprehensive-practical-green-ict-framework/51697](http://www.irma-international.org/chapter/comprehensive-practical-green-ict-framework/51697)

### Spatial Modeling and Geovisualization of Rental Prices for Real Estate portals

Harald Scherthanner, Hartmut Asche, Julia Gonschorek and Lasse Scheele (2017). *International Journal of Agricultural and Environmental Information Systems* (pp. 78-91).

[www.irma-international.org/article/spatial-modeling-and-geovisualization-of-rental-prices-for-real-estate-portals/179585](http://www.irma-international.org/article/spatial-modeling-and-geovisualization-of-rental-prices-for-real-estate-portals/179585)

### Prediction of Soybean Price Trend via a Synthesis Method With Multistage Model

Zhiling Xu, Hualing Deng and Qiufeng Wu (2021). *International Journal of Agricultural and Environmental Information Systems* (pp. 1-13).

[www.irma-international.org/article/prediction-of-soybean-price-trend-via-a-synthesis-method-with-multistage-model/274049](http://www.irma-international.org/article/prediction-of-soybean-price-trend-via-a-synthesis-method-with-multistage-model/274049)

### Ecological Areas and Challenges of Urbanization: A Stride Towards Environmental Restoration

Garima Toor and Tarush Chandra (2022). *Addressing Environmental Challenges Through Spatial Planning* (pp. 219-236).

[www.irma-international.org/chapter/ecological-areas-and-challenges-of-urbanization/290882](http://www.irma-international.org/chapter/ecological-areas-and-challenges-of-urbanization/290882)