Dashboards as strategy to integrate multiple data streams for real time surveillance

Fabian Eckelmann, Stephane Ghozzi, Alexander Ullrich

Robert Koch Institute, Berlin, Germany

Objective

Providing an integrative tool for public health experts to rapidly assess the epidemiological situation based on data streams from different surveillance systems and relevant external factors, e.g. weather or socio-economic conditions. The efficient implementation in a modular architecture of disease- or task-specific visualisations and interactions, their combination in dashboards and integration in a consistent, general web application. The user-oriented development through an iterative process in close collaboration with epidemiologists.

Introduction

The mission of the Infectious-Disease-Epidemiology Department at the Robert Koch Institute is the prevention, detection and control of infections in the German population. For this purpose it has a set of surveillance and outbreak-detection systems in place. Some of these cover a wide range of diseases, e.g. the traditional surveillance of about 80 notifiable diseases, while others are specialised for the timely assessment of only one or a few diseases, e.g. participatory syndromic surveillance of acute respiratory infections. Many different such data sources have to be combined to allow a holistic view of the epidemiological situation. The continuous integration of many heterogeneous data streams into a readily available and accessible product remains a big challenge in infectious- disease epidemiology.

Methods

The first step in the development of visualisation and analysis dashboards was the identification of relevant epidemiological questions. This was done through the review and analysis of existing epidemiological tools and workflows, among others through surveys and interviews. With the help of domain experts we identified the relevant data sources for specific tasks. We then chose data visualisations that are common in the field of infectious-disease epidemiology, e.g. disease maps, epicurves and age pyramids, as well as visualisations that were suggested by experts, e.g. time-series graph with severity thresholds. In an iterative process of propositions and expert feedback, we refined the user experience, adjusting variables, control parameters and the layout.

We have used two different technologies for the dashboard development. For tasks that needed extensive data integration and statistical computing we used the *Shiny* web-framework of the statistical programming language *R*, which allows for a seamless integration of data-wrangling, statistical methods and web design with interactive visualizations. For tasks where a more flexible and fluid user experience is desired and for the integration in a general web application, we used the more versatile single-page application (SPA) framework *AngularJS* in combination with *ASP.NET*. In both approaches we used standard open-source visualisation libraries such as *Leaflet* or *Plotly*. The dashboards were designed in a modular way, abstracting data sources and visualisations in order to reuse them and adapt them easily to other data sources. Where applicable, interfaces to live data bases and OLAP cubes where developed and implemented.

Results

We have developed a set of dashboards that allow the exploration of infectious-disease data, each designed for a specific epidemiological task. While still under active development, the dashboards are accepted and routinely used by epidemiologists of the Robert Koch Institute. The expansion to other user groups (e.g. local health agencies) is planned for the near future. Further dashboards will be developed as new epidemiological tasks are identified.

A general dashboard ("Signals Dashboard", see Figure 1 A) is displaying laboratory confirmed cases and their distribution across time, space, age and sex in linked widgets. Additionally it highlights anomalous clusters of cases in all widgets and lists the anomalies in an interactive table. The dashboard is available for all (approx. 80) notifiable diseases. The "Severity Dashboard" (Figure 1 B) integrates influenza-related syndromic data, virological information and laboratory confirmed cases. The indicators

transmissibility, seriousness and impact, as defined by the PISA guidelines of the World Health Organization, are displayed in time-series charts (absolute and cumulative) and tables; parameter-adjustable severity assessments are computed on the fly. This dashboard has then been adapted to monitor in real time the severity of rotavirus infections. One further dashboard focusses on vaccine-preventable diseases and allows the simultaneous exploration of incidences and vaccination rates through synchronized maps and histograms. Lastly, a "Context Dashboard" enables the exploration of possible connections between tick-related diseases such as TBE and Lyme disease on the one hand, and weather and environment as external factors on the other. It provides visual comparisons through maps and time-series charts, correlation analysis and statistical modeling. The user can choose a set of (lagged) variables to be included in a linear statistical model, which is immediately trained. The contributions and significance of the chosen factors, as well as the fit and prediction accuracies, are displayed in tables, scatter plots and time series. Both "Signals" and "Severity" dashboards serve the rapid assessment of the epidemiological situation and as such display live data as read from internal databases and cubes. The others are at present rather meant for retrospective analyses but will be connected to live data streams in the future.

Conclusions

Dashboards can provide a way to integrate different epidemiological data streams and statistical methods, offering experts a useful tool to assess the epidemiological situation. Close collaboration between epidemiologists and data scientists in the design and development is beneficial to the relevance and sustainability of such a tool.
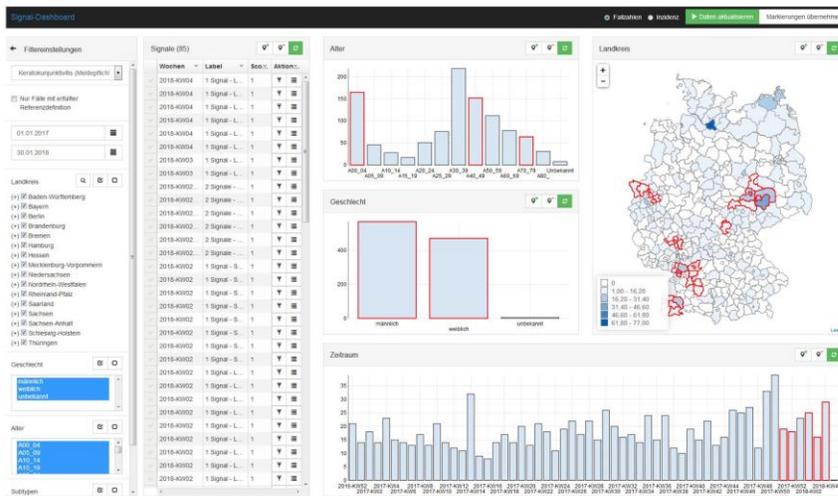
Acknowledgement

A

B

Figure 1. A) The "Signals Dashboard" giving an overview of the distribution of lab reported infectious disease cases in Germany as well as highlighting anomolous clusters of cases; B) The "Severity Dashboard" showing influenza indicators in comparison to severity thresholds computed from historical data.