Robert Gutounig / Eva Goldgruber / Julian Ausserhofer / Keith Andrews / Thomas Traunmüller / Thomas Wolkinger

The Styrian Diversity Visualisation Project: Communicating Data Stories with an Open Data Visualisation Web App

109 - Data Science: Erfassung, Modellierung, Analyse und Visualisierung von Daten

Abstract

This article describes the Styrian Diversity Visualisation Project (SVV)[1] focusing on three main parts: the development of the prototype SVV web app which enables users to explore their own situation in relation to others within the wider community, a corresponding data blog demonstrating the journalistic potential of data visualisation, and, finally, the integration into a course for students producing data-intensive stories. Furthermore, it highlights the project results as a prototypical example of a data journalistic piece.

Keywords:

data visualisation, data-driven journalism, data journalism, data stories, diversity, styria, web app, data server, triple store, SPARQL.

1. Introduction

In accordance with the global trend towards open government data (Lathrop / Ruma 2010), numerous datasets have been published by local, regional, and national government authorities in Austria.[2] Despite the improved availability of data, there are hardly any (interactive) visualisations available, neither for the region of Styria nor for other Austrian regions, nor for the general public, in particular.

The Styrian Diversity Visualisation Project (SVV) makes use of some of the available open data about the region of Styria's diversity. A specific objective of the project is to demonstrate the benefits of a visualisation platform for open government data, both for the end user and for university-level

[1] In German, Steirische Vielfalt visualisiert (SVV).

[2] For an overview of participating institutions and available datasets see https://www.data.gv.at/.

journalism education. The SVV web app enables users to visualise their own situation within the wider community. The more specific information users reveal about themselves, the more corresponding information is visualised regarding similar other people. Accompanying communication measures, such as data-intensive blog posts, figure as an integral part of the project in order to demonstrate the narrative value of data visualisation. Some elements of the project are integrated into a class on data journalism, in which students learn to produce data-intensive stories. Finally, SVV contributes to the goals of the Styrian provincial government's Charter for Living Together in Diversity in Styria (Styrian Government 2011), which has been signed by a number of regional institutions declaring themselves to be *integration partners*. The conceptual components of the project are illustrated in Figure 1:

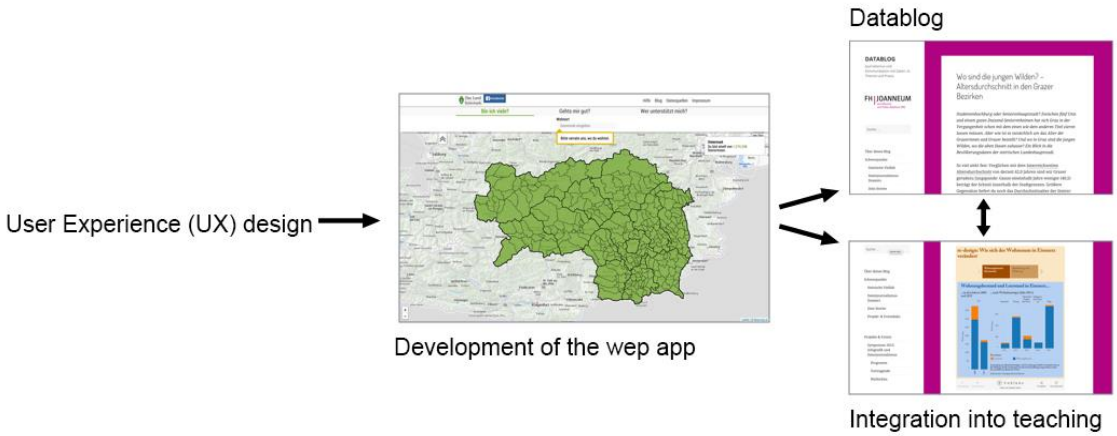

Figure 1: The conceptual components of the Styrian Diversity Visualisation Project (SVV).

2. Method

The project was developed using methods of User Experience (UX) design, including user research, stakeholder interviews, content strategy, and user stories (Garrett 2010). In a perfect scenario, users find what they need, understand what they find, and are able to interact within a reasonable time (Redish 2013: 294). As shown in Figure 2, content strategy forms an integral part of UX and "helps organizations [to] provide the right content, to the right people, at the right times, for the right reasons" (Casey 2015: xxiv). Central elements of content strategy include the definition of the core topics and messages, the analysis of content gaps, and strategic recommendations for the generation, publication and administration of content (Halvorson 2008, Halvorson / Rach 2012).

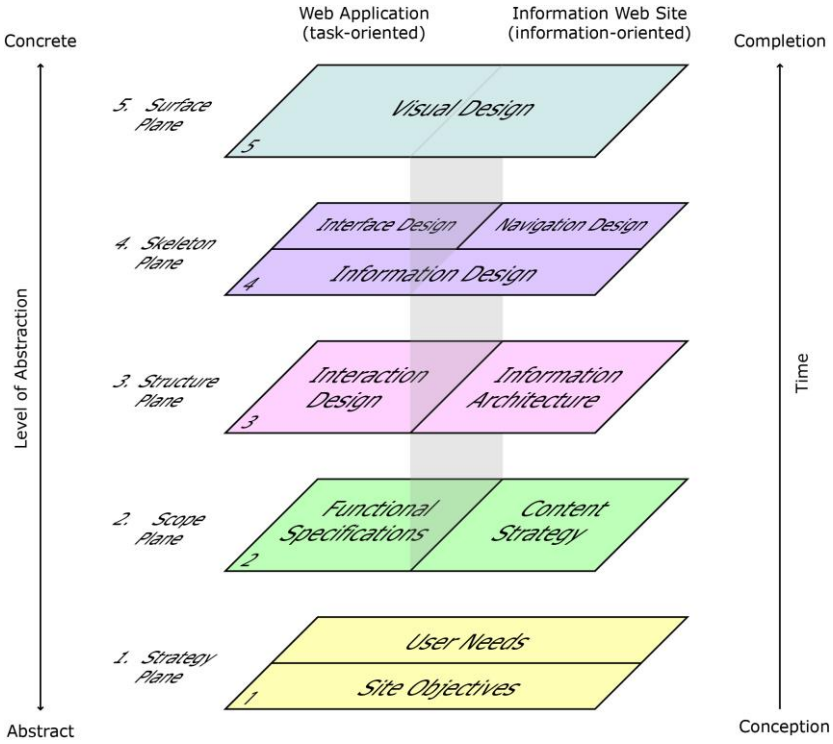

Figure 2: The five planes of user experience. Adapted by Andrews (2015), based on the original by Garrett (2010).

First, a competitive analysis was performed of international best practice examples of data visualisation projects in similar areas. These are projects which have similar goals: for example, the "Qui Sommes Nous? ("Who are we?") visualisation by the city of Rennes also tries to promote open data and to "spark a debate around the sense of belonging to the urban community" (Dataveyes 2014). This analysis generated ideas for user scenarios and ways of presenting and visualising demographic open data. Afterwards, stakeholder interviews were conducted, a content audit was performed, and web usage data from related online services of the regional government of Styria were analysed. This revealed that, in general, users are interested in demographic statistical data about Styria, but the relatively high bounce rate of relevant pages indicated a specific need for a user-friendly frontend like SVV.

The results were used to develop user stories, describing in a few words potential user types and their motivation for using the service (GDS 2016). User stories are one of the techniques of agile software development (Beck et. al 2001). The final three user stories which emerged from this process are shown in Table 1. Two additional user stories were considered, but not used in the first version of the

web app: *The Journalist*, interested in finding potential stories through the service and personal contact for further questions, and the *Data Enthusiast*, interested in complex data and sophisticated visualisations. Finally, a thorough audit was performed to determine the ownership, quality, and licence conditions of relevant statistical data and data concerning the integration partners.

The Curious Visitor	The Searching Type	The Integration Partner
"I found the service through social media, I would like to find out if the promised benefit can be realised: to find out about socio-demographic diversity in an entertaining way."	"I am interested in the variety of people living together in Styria. I would like to know how diverse my neighbourhood is and how I fit into this social landscape. In order to find this out, I am ready to devote some time."	"I am well-informed about the initiatives of the Province of Styria concerning integration and am actively taking part in this process, I would like the role which my institution is playing to be well visible."

Table 1: The final three user stories.

3. Results

After several rounds of discussion, three scenarios began to crystallise and mock-ups were set up for each stage of the respective scenario. All scenarios have in common the idea of *progressive disclosure*: that the user is supposed to disclose a piece of information such as their place of residence or age range, then in return should be rewarded with a more specific visualisation of corresponding statistics as well as additional information on the chosen topic. The results are presented in the three SVV application scenarios shown in Figures 3 to 5. Each scenario takes form of a structured dialogue framed in relation to the user's input ("How many people are in the same situation as me?" "In which part of the region there are more/fewer people like me?" etc.)

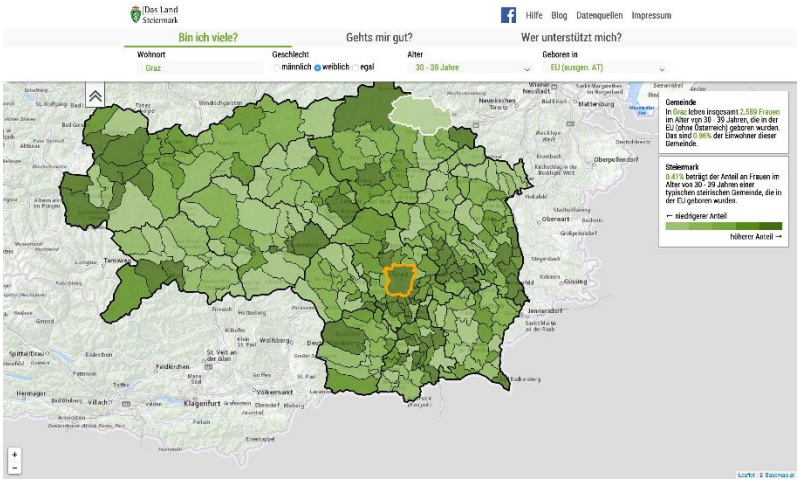

Figure 3: Scenario 1 - "Bin ich viele?" ("How Many Am I?")

For example, in the first scenario, the user is prompted to enter their place of residence, gender, age range, and place of birth, in that order. In return, a choropleth map of Styria is successively updated to reflect the distribution of other people in the same demographic (Andrews et al. 2015).

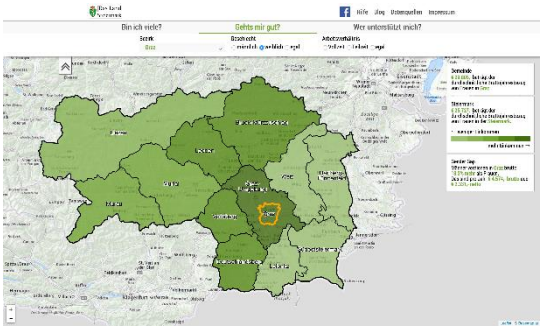

Figure 4: Scenario 2 - “Geht’s mir gut?” (“How Am I Doing?”)

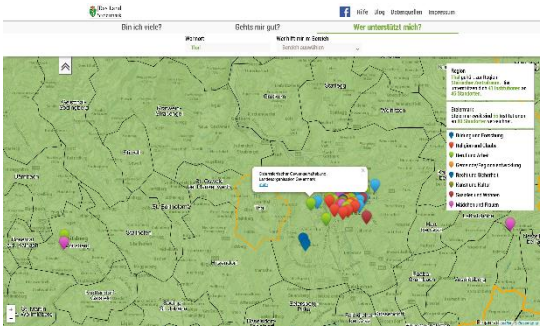

Figure 5: Scenario 3 -“Wer unterstützt mich?” (“Who Can Support Me?”)

The user interface of the SVV web app was then designed and improved over a dozen or more iterations. The user enters the URL of the web app into their web browser, and it is downloaded into the browser to run. Whenever the browser needs data, it requests it on demand by sending a SPARQL query to the data server. The data requested (and only that subset) is then returned to the web app, which displays it in an appropriate visualisation. This architecture is shown in Figure 6. The web app uses the principles of responsive web design (Marcotte, 2014) to adapt to the particular characteristics of the display device.

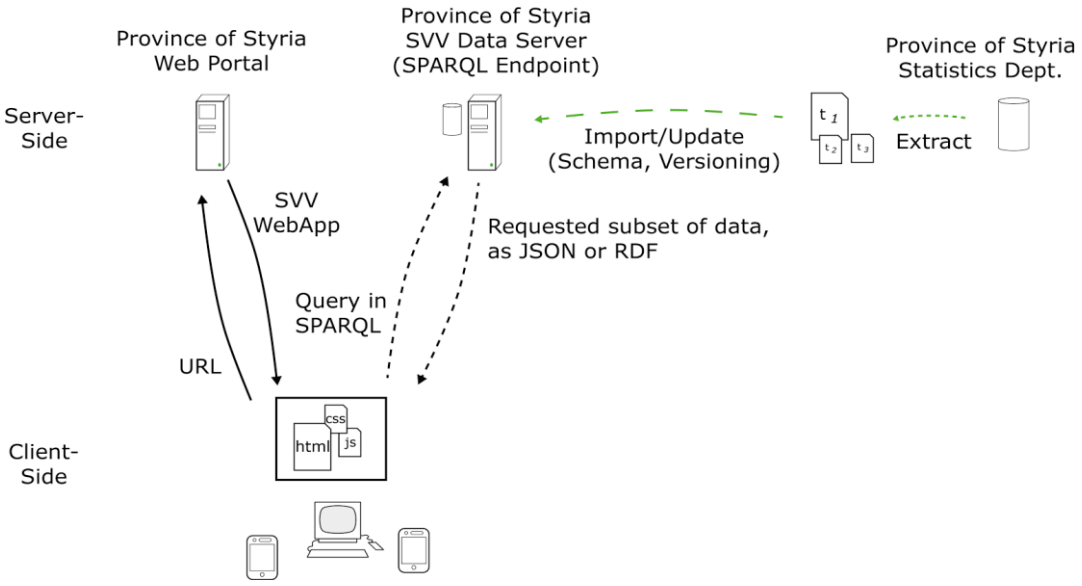

Figure 6: The client-server architecture of the Styrian Diversity Visualisation (SVV) web app (Andrews et al. 2015).

SVV can be seen as an example of *data journalism*, which Howard (2014: 4) defines as "gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism." Data journalists are introducing computational methods and the use of open data into newsrooms (Ausserhofer, 2015). Although not developed in a newsroom context, SVV is a 'typical' data-journalistic piece, fulfilling the criteria Loosen et al. (2015: 19-20) extracted from a content analysis of 120 data journalism items:

- covers a political topic,
- relies on public data from official sources
- builds its story on financial and […] geodata […],
- is based on a simple unit of analysis such as individual persons,
- compares values in order to show differences and similarities between different objects of study (e.g., people of different gender, neighbourhoods)
- combines two types of visualisations – preferably pictures with maps or simple charts,
- allows the user to zoom into a map, request details and/or to filter data.

In order to demonstrate the journalistic value of data visualisations like SVV, blog posts were developed highlighting certain aspects of the platform. The production of data stories was part of a class on data reporting in an undergraduate journalism course. Students used the platform to explore aspects of diversity in Styria and then further investigated chosen topics by researching additional data and exploring hotspots that became evident through the web app. At this stage of development, it is not possible to embed maps from SVV into other websites and end users cannot yet visualise their own datasets. Hence, the students used third-party applications such as CartoDB to produce their own diversity maps, employing additional data. The data stories were then written in an accessible style and readers were invited to contribute and share.

4. Concluding Remarks

The SVV project is a showcase for the visualisation of open government data. It demonstrates the use of a dedicated data server when realising open data apps. The project confirmed that while identifying data sources and building visualisations, both a solid content strategy and iterative development are necessary to reach a satisfying and sustainable end product. This process also included negotiating internal workflows regarding data sources and updates, especially when dealing with unstructured content (for Scenario 3). All data sets being used in the project are currently or scheduled to be available on the open government data portal of the Province of Styria[3], so that advanced users (*Data*

[3] http://data.steiermark.at/

Enthusiasts) can perform more detailed analysis. Further user testing is planned to both improve usability and suggest potential ideas for future development.

Acknowledgements

We would like to acknowledge the contribution and support of our project partners at the Province of Styria (Land Steiermark) and the support by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the Future program via VALiD project no. 845598, which enabled the integration of a data-journalistic perspective into this paper.

References:

Andrews, K. (2015). Information Architecture and Web Usability: Course Notes. http://courses.iicm.tugraz.at/iaweb/iaweb.pdf (28 Jan 2016).

Andrews, K., Traunmüller, T., Wolkinger, T., Gutounig, R., and Ausserhofer, J. (2015). Building an Open Data Visualisation Web App Using a Data Server: The Styrian Diversity Visualisation Project. In Proc. 15th International Conference on Knowledge Technologies and Data-Driven Business (I-Know 2015), Graz, Austria, pages 48:1–48:4, ACM. doi:10.1145/2809563.2809596 [Best Demo Award]

Ausserhofer, J. (2015). „Die Methode liegt im Code": Routinen und digitale Methoden im Datenjournalismus. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Eds.), Digitale Methoden in der Kommunikationswissenschaft (pages 87–111). Berlin. doi:10.17174/dcr.v2.5

Beck, K. et al. (2001). Manifesto for Agile Software Development. http://agilemanifesto.org/ (28 Jan 2016).

Casey, M. (2015). The Content Strategy Toolkit: Methods, Guidelines, and Templates for Getting Content Right. San Francisco, California: New Riders.

Dataveyes. A Data-Driven Snapshot of the Population of Rennes. Blog post. Jun 2014. http://dataveyes.com/#!/en/case-studies/rennes-metropole (28 Jan 2016).

Garrett, J. (2010). The Elements of User Experience. 2nd Edition. Berkeley, California: New Riders.

GDS (2016). Writing User Stories. Government Service Design Manual, Government Digital Service https://gov.uk/service-manual/agile/writing-user-stories.html (28 Jan 2016).

Halvorson, K. (2008). The Discipline of Content Strategy. http://alistapart.com/article/thedisciplineofcontentstrategy (28 Jan 2016).

Halvorson, K./Rach, M. (2012) Content Strategy for the Web. 2nd Edition. Berkeley, California: New Riders.

Howard, A. B. (2014). The Art and Science of Data-Driven Journalism. New York: Tow Center for Digital Journalism. http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf (26 Jan 2016).

Lathrop, D./Ruma, L. (Eds.). (2010). Open Government: Collaboration, Transparency, and Participation in Practice. Sebastopol, California: O'Reilly.

Loosen, W./Reimer, J./Schmidt, F. (2015). When Data Become News: A Content Analysis of Data Journalism Pieces. Talk at The Future of Journalism Conference 2015, Cardiff, UK.

http://slideshare.net/jpub20/when-data-become-news-a-content-analysis-of-data-journalism-pieces (28 Jan 2016).

Marcotte, E. (2014). Responsive Web Design. New York: A Book Apart.

Redish, J. (Ginny) (2013). Content as Conversation in Government Websites. In Proc. Second International Conference on Design, User Experience and Usability (DUXU 2013), Las Vegas, USA, pages 294–303. Springer. doi:10.1007/978-3-642-39253-5_32

Styrian Government (2011). Charter for Living Together in Diversity in Styria. http://soziales.steiermark.at/cms/dokumente/11562700_103650128/a502daed/Charta_EN.pdf (28 Jan 2016).