

# Linguistic information visualization and web services

Chris Culy and Verena Lyding

European Academy Bolzano-Bozen

Bolzano-Bozen, Italy

<http://www.eurac.edu/linfovis>

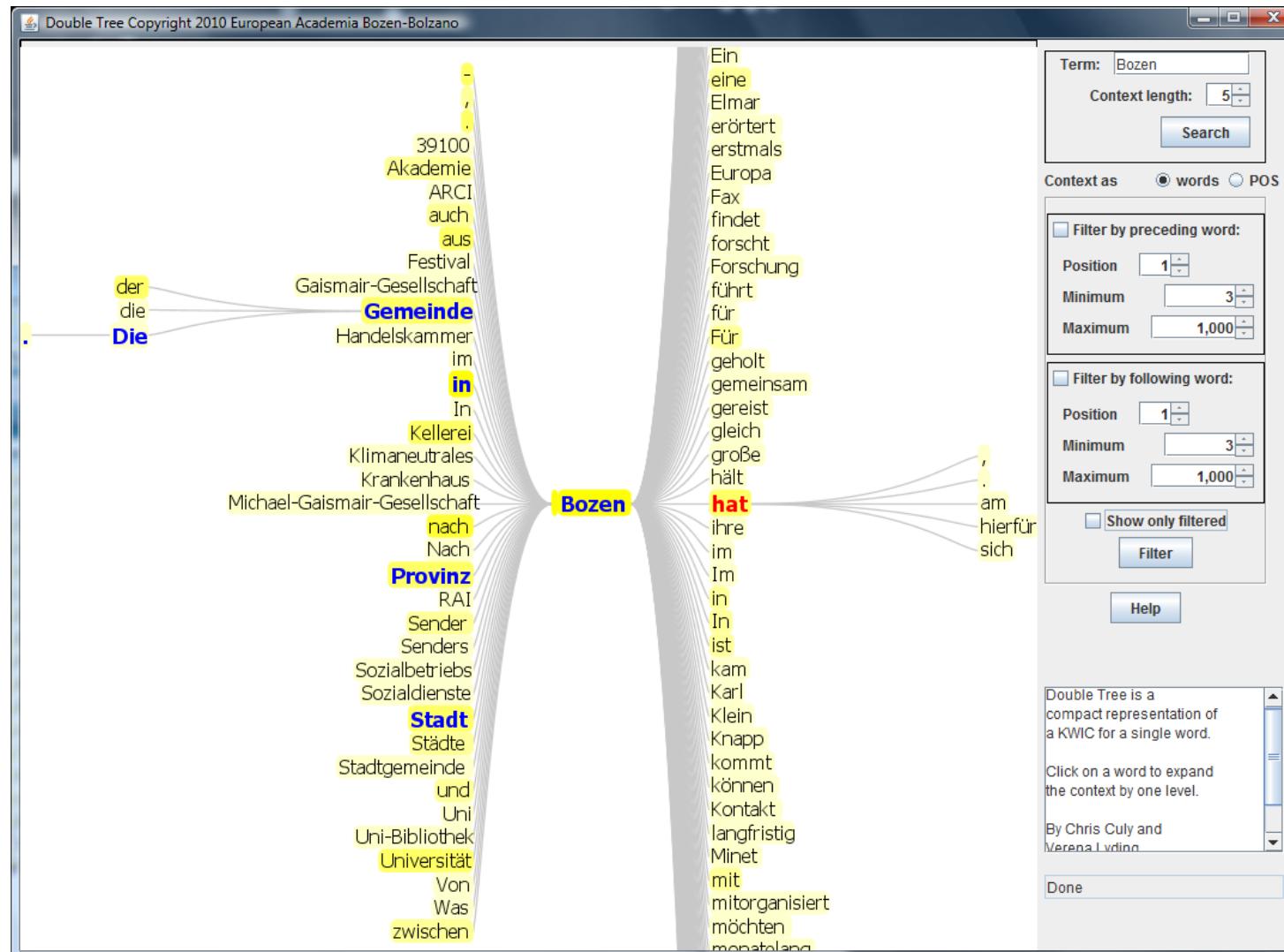
# LInfoVis (= Linguistic Information Visualization)

- LInfoVis = The graphical display, primarily on computer monitors, of any kind of information concerning language and its use.
  - Subfield of Infovis = Information visualization
- Visualizations of language are everywhere
  - Text, Tag clouds, ...
- The meaningful visualization of linguistic information is needed by language users (experts and non-experts): for presentation, manipulation and analysis
- LInfoVis provides tools for visualization
- WebLicht provides linguistic information via web services
- How do we connect LInfoVis to WebLicht (and web services more generally)?

## Where we're going

- A few ideas about information visualization
- Some particular challenges of LInfoVis
- LInfoVis and WebLicht
- Conclusion / Future Directions
- But first ...

# Double Tree (Demo)



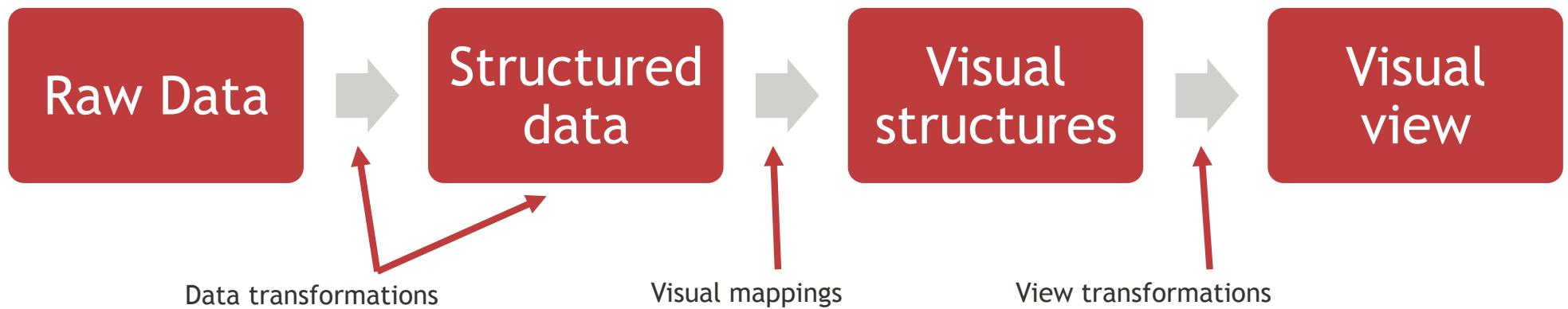
Search for “Bozen”,  
with focus word “hat”

# Some general ideas about information visualization

- Data is encoded by *visual variables* (color, shape, size, etc.)
  - It is well established which kinds of visual variables are suited for which kinds of data
    - Cf. Shades of yellow for frequency in Double Tree
- Interactive aspects are a fundamental part of many digital visualizations
  - Especially for data exploration
  - A variety of interaction techniques suitable for visualizations is well known
    - Cf. Hide/show data + animation to help track changes in Double Tree
- BUT, these are only the foundations for visualizations
  - Much more to do to design the visualizations
  - Much more research to do on finer levels of detail in both areas

# A reference model of LInfoVis

(Culy & Lyding 2008, cf. Card et al. 1999)

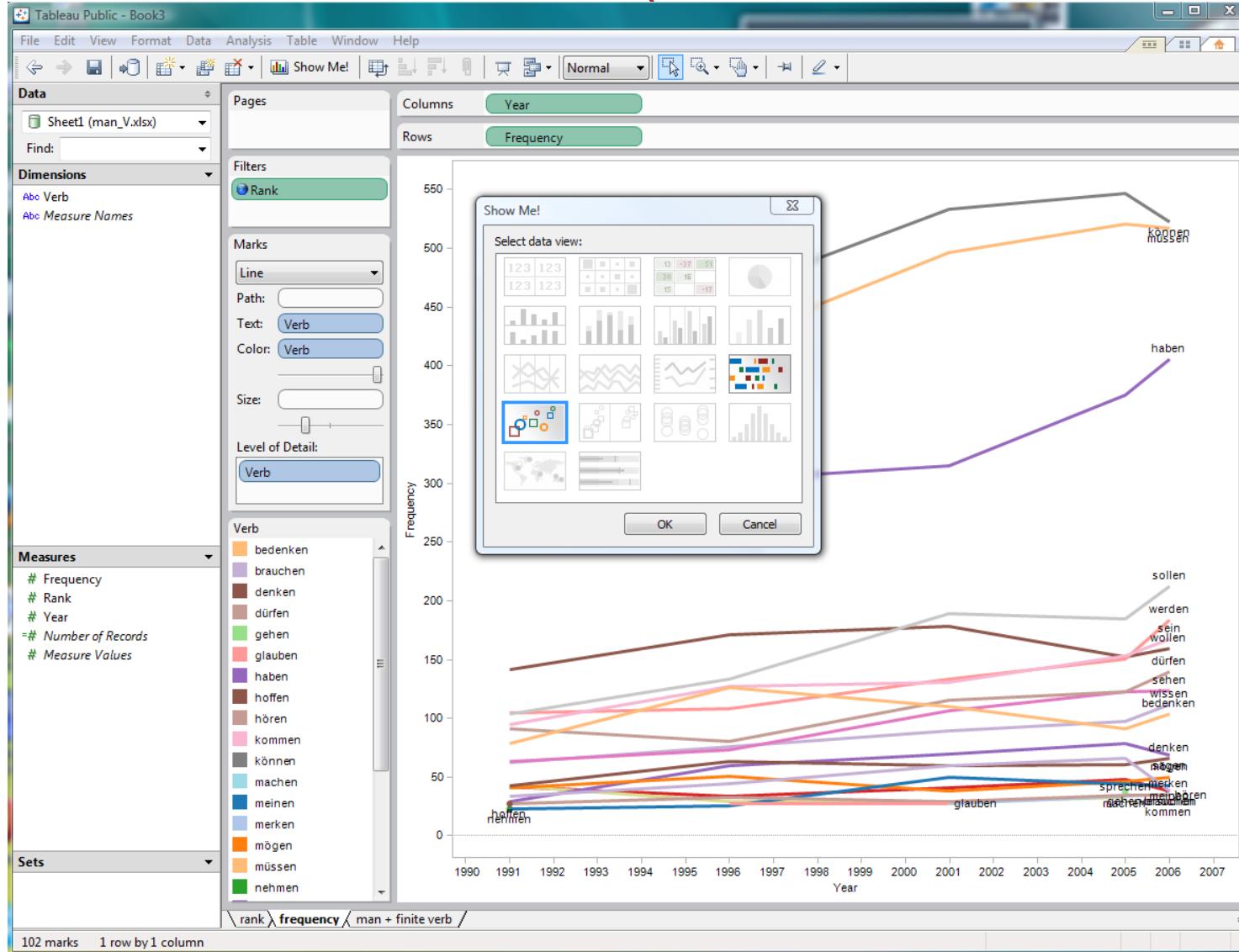


1. Raw Data, e.g. texts
2. Data transformations, e.g. counting, sorting, tagging
3. Structured data, e.g. document vectors, word/lemma/POS lists
4. Visual mappings = the *type* of visualization,
  - e.g. POS -> color, scatter plot, tree
5. Visual structures = the general visual form, e.g. chart, tree, text
6. View transformations = visual structures are assigned specific visual appearances
7. Visual view = the *visual appearance*, e.g. color, shape, size, position
  - Also includes filtering of data: which data is visible

## What makes LInfoVis special?

- LInfoVis has additional visual requirements, beyond what is known about visualizing numerical/quantitative information, because
- LInfoVis data is (usually, in some form or another) text and
  - textual items are not *mappable* (our term)
  - i.e. we (usually) cannot effectively represent textual items by something else meaningful (shape, color, position, etc.)
    - too variable
    - too complex
- LInfoVis is often concerned with structure, which is not easily be represented by compact visualizations

## Illustration: Tableau (tableausoftware.com)



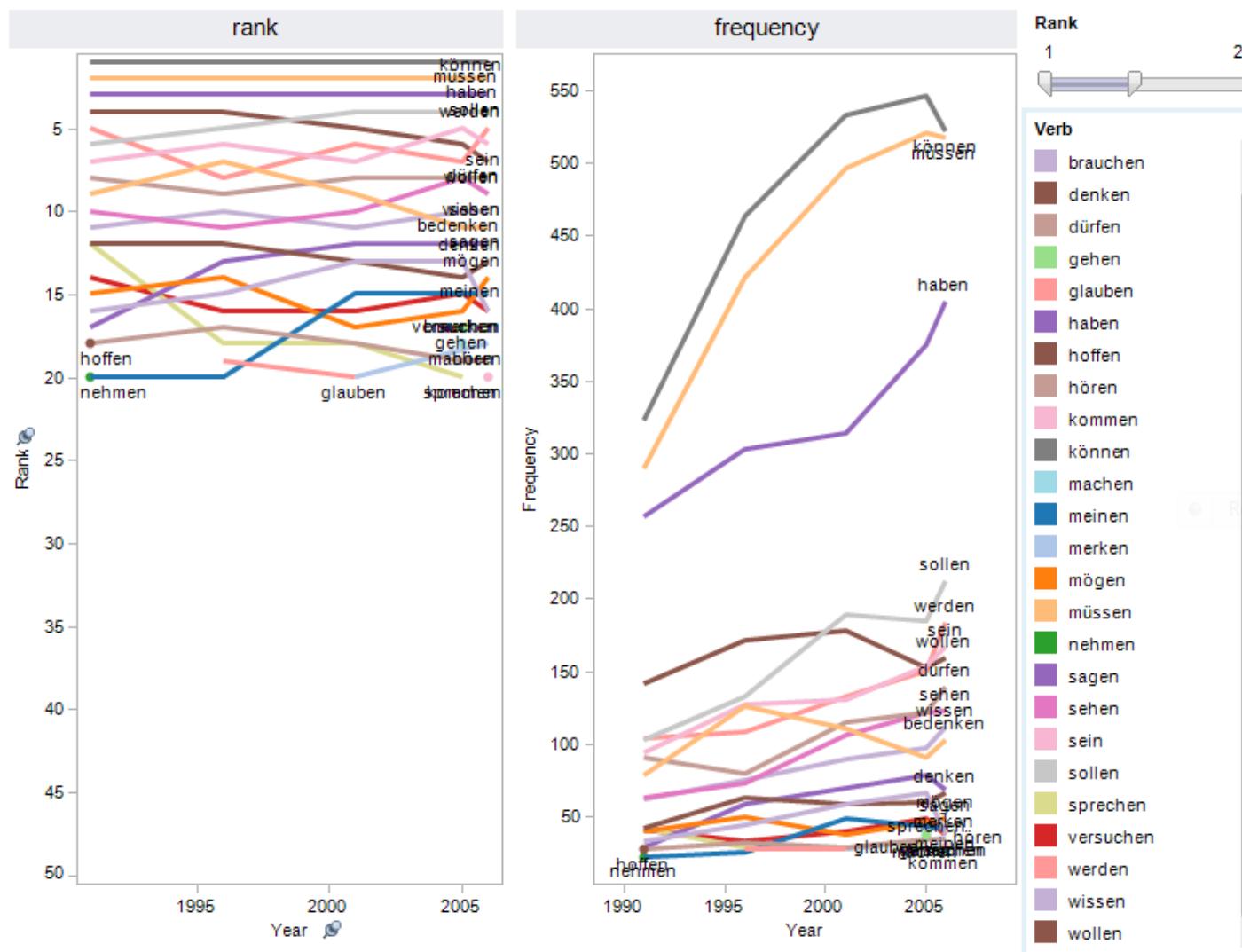
Show Me!

suggests appropriate graphs

Data:

Frequencies of  
„man“ + finite verb  
in the *Dolomiten*  
by year

## Illustration: Tableau (tableausoftware.com)



Dashboards show  
and coordinate  
multiple graphs

Data:  
Frequencies of  
„man“ + finite verb  
in the *Dolomiten*  
by year, filtered to  
the top 20 in each  
year

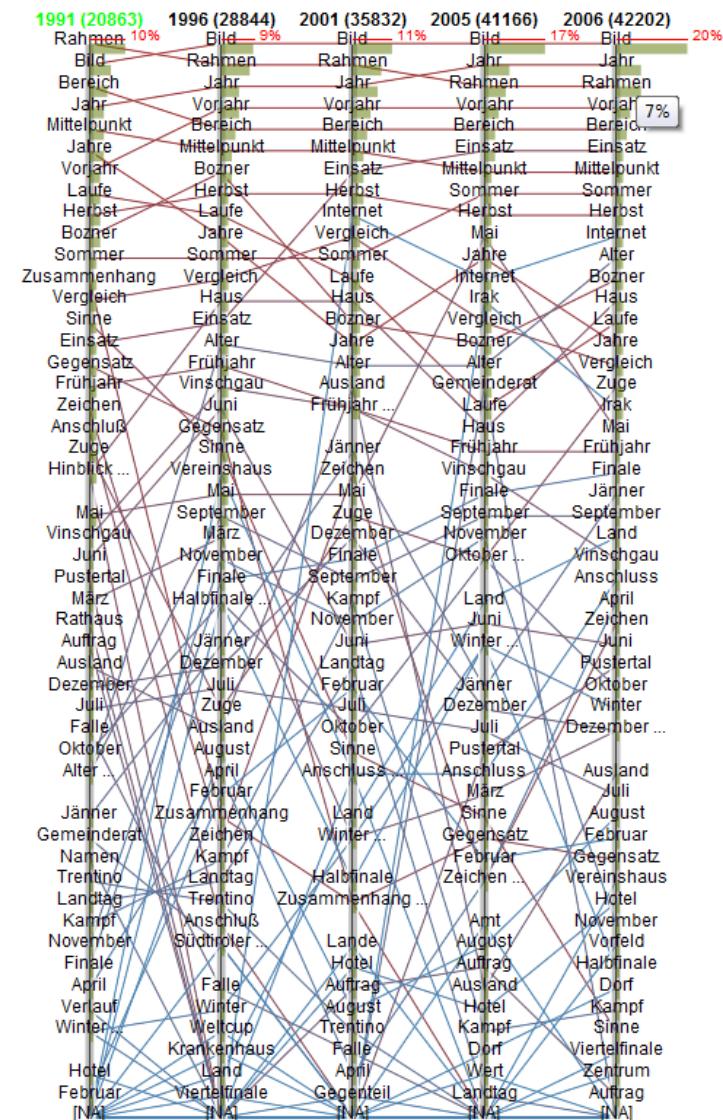
# Structured Parallel Coordinates

Jahr 4 3 3 2 2

Information visualization has been around as a field for ~15 years, so there are lots of visualization types to use as a basis for LInfoVis.

Here is one based on Parallel Coordinates:

# Rankings of nouns following „im“ in 5 years of the *Dolomiten* newspaper



# InfoVis and WebLicht, generally

- Functional aspects
  - Visualization for analysis
  - Visualization for data creation/manipulation (tools)
    - E.g. annotation, evaluation, selection
- Structural aspects
  - Visualization as “pure consumer” ( = dataflow “sink”)
    - The visualization does not contribute results back to WebLicht
  - Visualization as contributor
    - The visualization does contribute results back to WebLicht
    - It is harder to come up with a scenario for this case

## LInfoVis and WebLicht, research issues

1. What are relevant kinds of visualizations and applications for LInfoVis (in the context of WebLicht)
  - How do we address the special challenges of LInfoVis
  - Who are the target users and what do they need from LInfoVis?
    - Linguists/specialists - in what domains?
    - Non-language researchers
    - General public
    - What are the tasks they are trying to accomplish?
2. What is the right level of generality for encapsulating visualizations?

## LInfoVis and WebLicht, research issues (continued)

3. Can we classify the appropriateness of visualizations enough to create a “Vis Assistant”? (cf. “Show Me” in Tableau)
4. How do we deal with the massive amounts of data?
  - This is a current topic in Infovis more generally
5. How do we do evaluation? (Also a more general Infovis question)
  - No consensus, especially for exploratory analysis visualizations

## LInfoVis and WebLicht, technical issues (for breakout)

1. What are the input/output specifications/formats needed?
  - Tables are standard (and easy)
  - Some data type inference is possible (cf. Tableau)
    - Probably limited
    - What are the data types?
  - Encoding of the types of relations among information?
  - What about including task-appropriateness?
3. Client side vs. server side visualization construction
  - Most visualization these days is done on the client
  - But there has been some (mostly older) research on server side visualization
  - Depends on amount of data, type of visualization, practicalities
  - Could use help from computer scientists, especially in Grid
    - (There has been some work on visualization in Grid computing)

## LInfoVis and WebLicht, technical issues (continued)

### 3. Integrateability

- Cf. Tom's talk later about loose coupling vs. high integration
- Standalone (cf. Tableau, Excel)
- Flexible components (cf. Google VisAPI)
- Integrated, application specific (cf. *Dolomiten* time series)
  - Cf. Visual Analytics

# Conclusions

- Lots to be done
  - Solving basic challenges of LInfoVis
  - Placing LInfoVis applications in bigger context (e.g. WebLicht)
    - Serving the user
    - Technical challenges
- Promising area, both for research and feasibility
  - Innovative approach to handling/dealing with language data
  - Integration of technical aspects on different levels
  - There is a solid base of LInfoVis visualizations
    - EURAC
    - Other annotation tools
    - General visualizations that can be adapted easily

Thank you for your attention

Chris Culy & Verena Lyding  
christopher.culy@eurac.edu  
verena.lyding@eurac.edu

<http://www.eurac.edu/linfovis>

# Title

Text