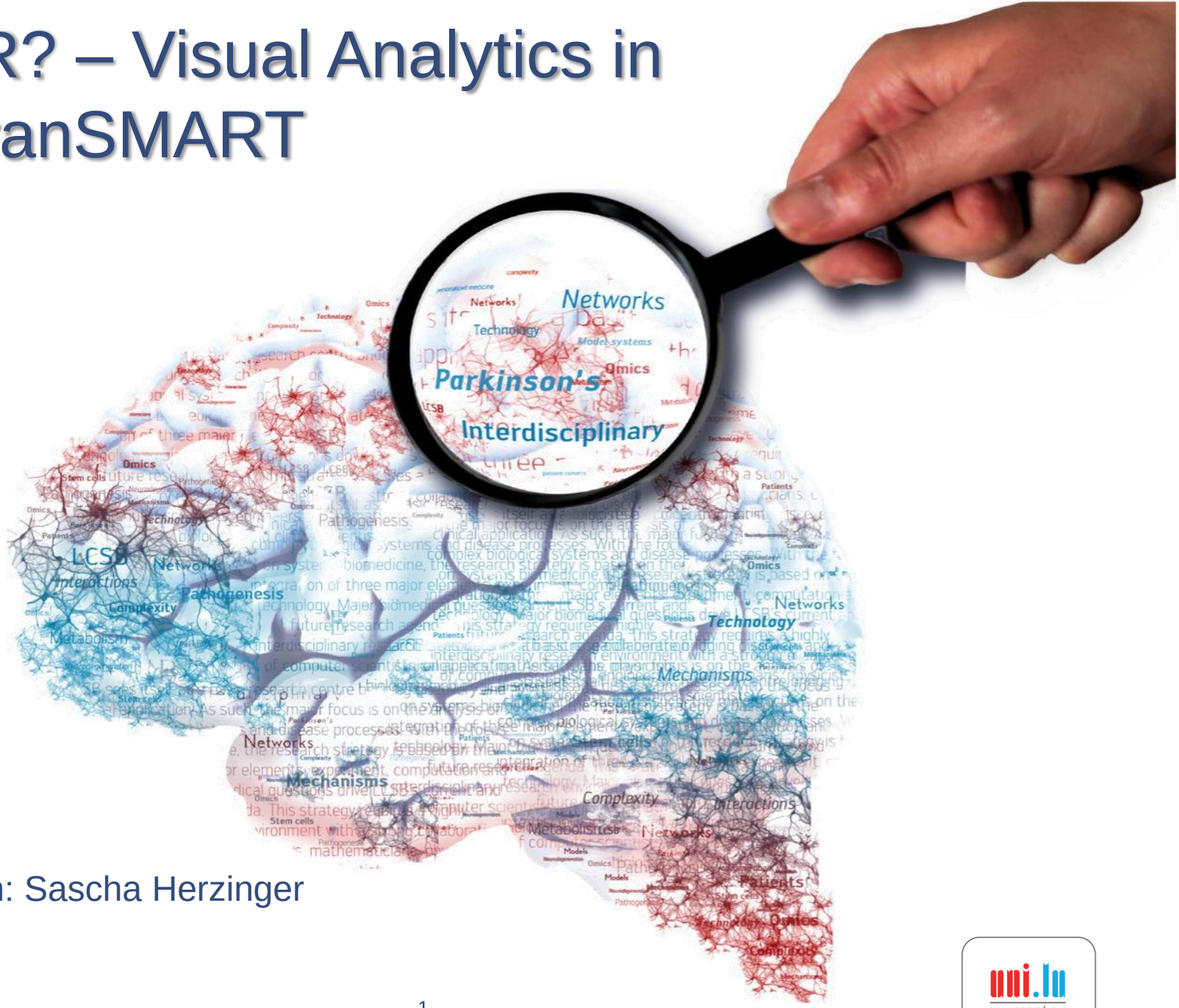
SmartR? – Visual Analytics in i2b2/TranSMART



Presentation: Sascha Herzinger

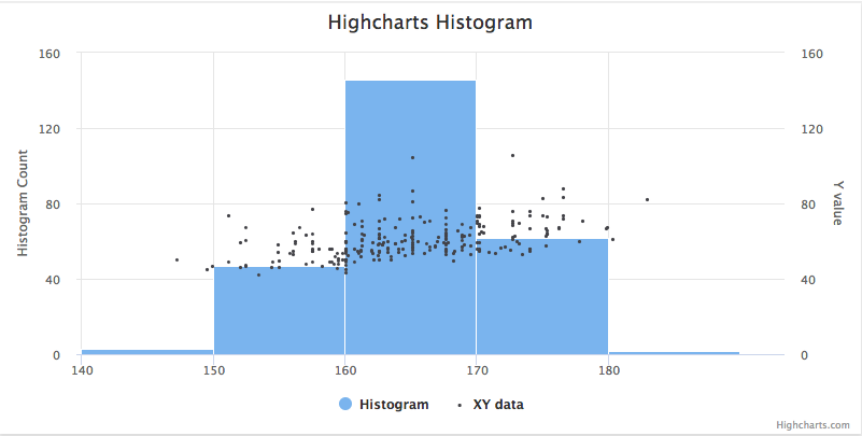
Goal

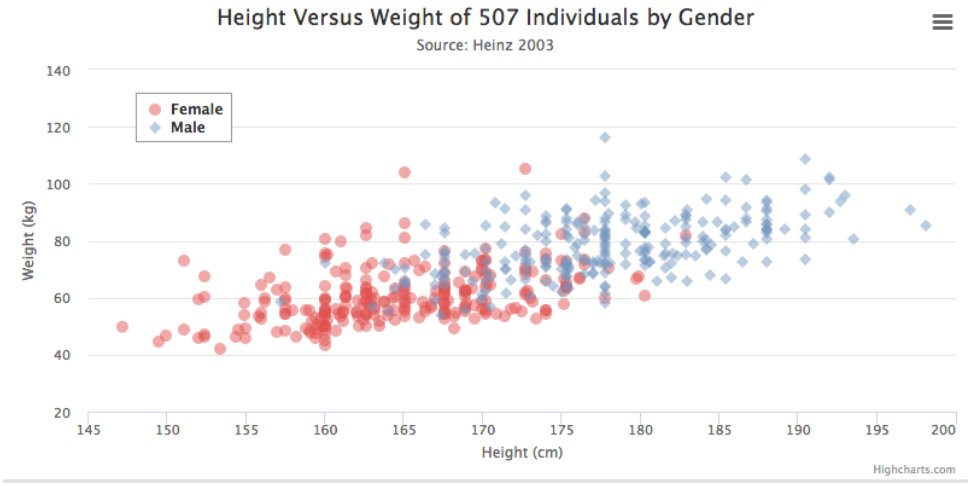
- Bring visual analytics to translational research platforms
- Permit explorative analysis that complements hypothesis driven research
- Explorative analysis is about finding the question first, because providing answers is expensive

What are ”visual analytics”?

- The name suggests a visual and an analytical component

- Is this visual analytics?



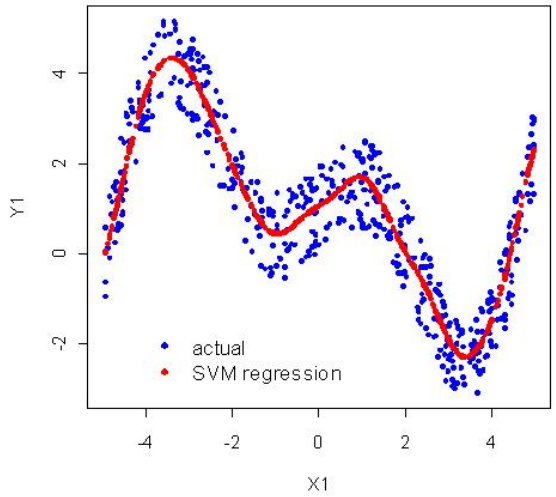


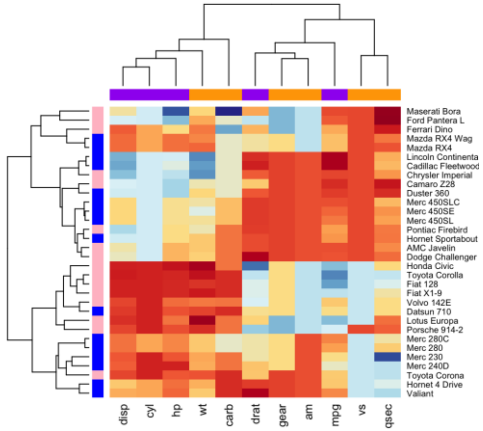
- No. There is a visual component, but the analytical component is negligible.

What are ”visual analytics”?

- How about R?

- Are these visual analytics?



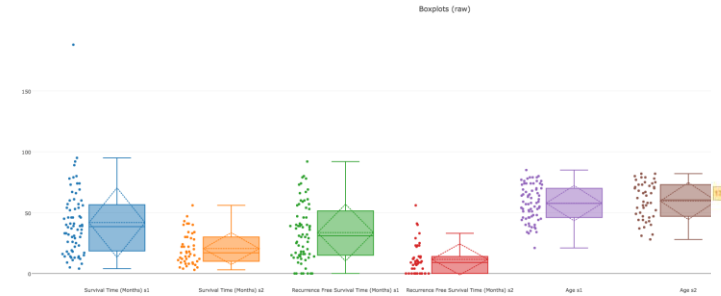


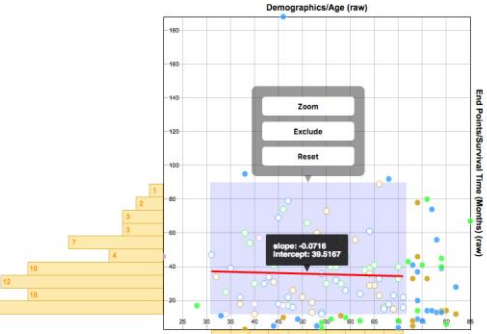
- Absolutely! There is a visual component and an analytical component (SVM regression / clustering)

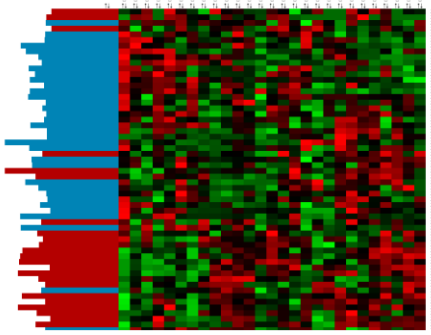
Ok, so why not just using R?

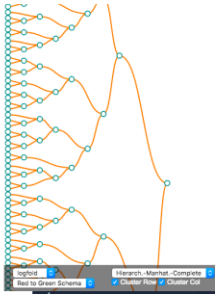
- Strong analytical component, but the visual capabilities are rather lacking.
- Possible Solution: Using R for analytics and the web browser for the visualisation.
- *SmartR* does just that!

What is SmartR?

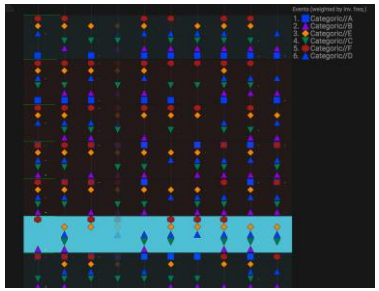




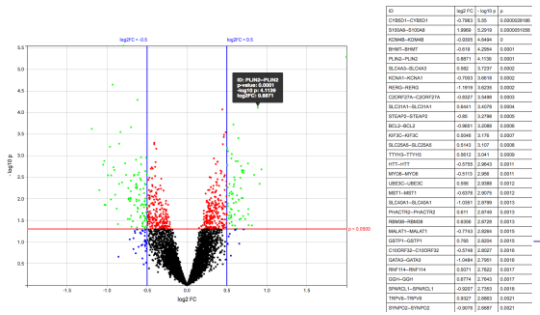


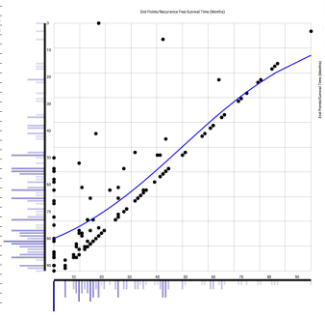


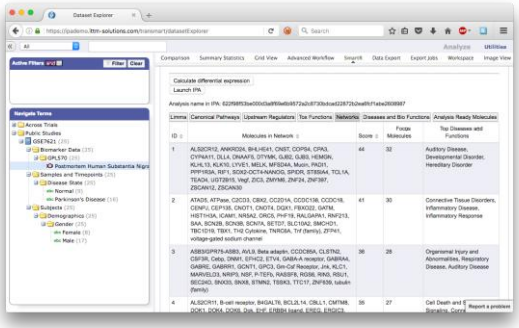
This.



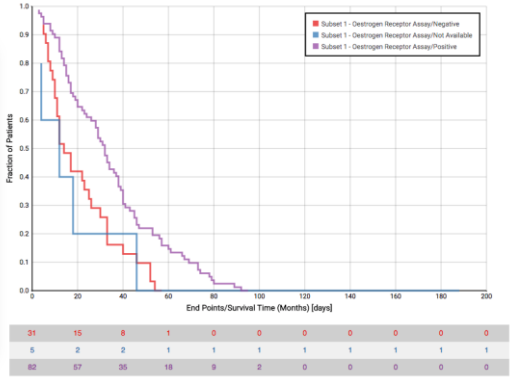
[1]

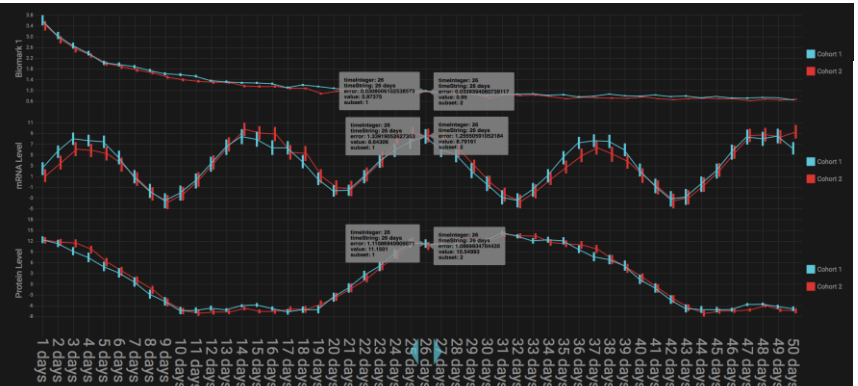


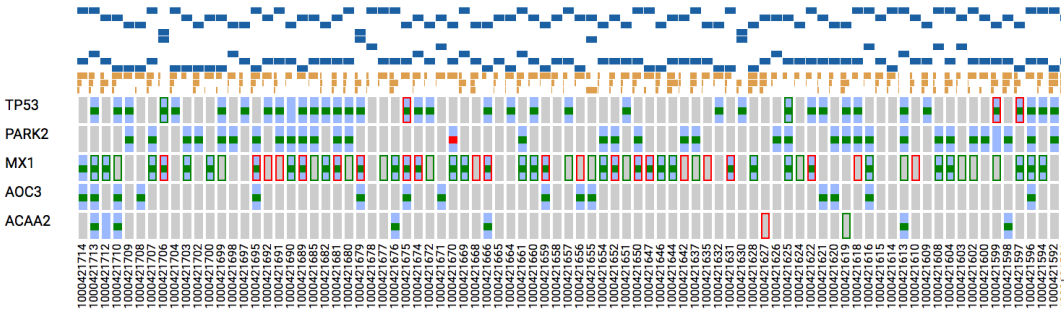




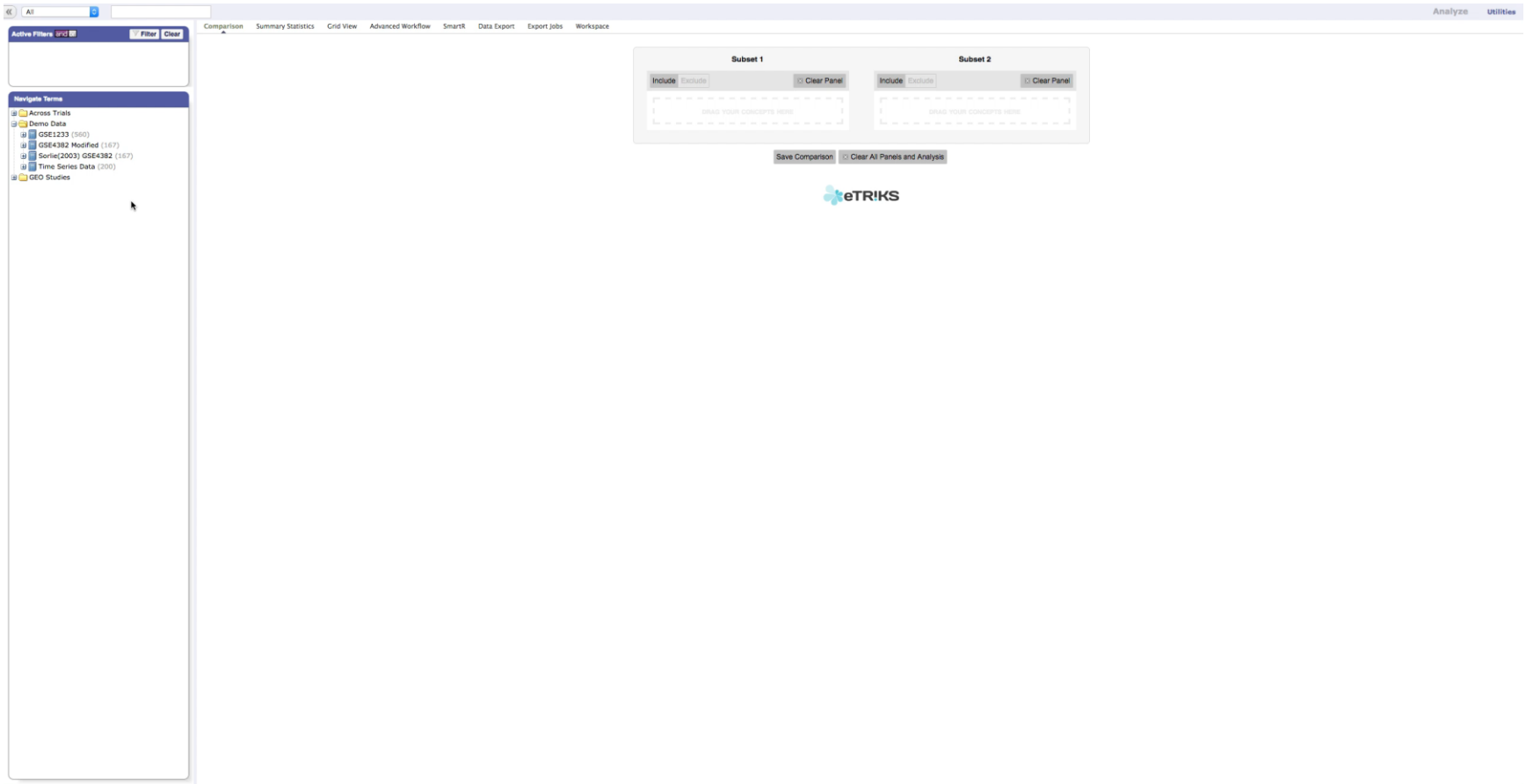
[2]







SmartR Demo


All
Analyze
Utilities
Active Filters
Filter
Clear
Comparison
Summary Statistics
Grid View
Advanced Workflow
SmartR
Data Export
Export Jobs
Workspace
Navigate Terms
Across Trials
Demo Data
GSE1233 (560)
GSE4382 Modified (167)
Sorlie(2003) GSE4382 (167)
Time Series Data (200)
GEO Studies
Subset 1
Include
Exclude
Clear Panel
DRAG YOUR CONCEPTS HERE
Subset 2
Include
Exclude
Clear Panel
DRAG YOUR CONCEPTS HERE
Save Comparison
Clear All Panels and Analysis
eTRIKS

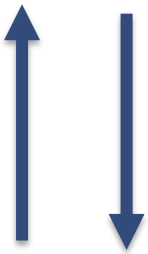
Ok, so why not just using R?

- Strong analytical component, but the visual capabilities are rather lacking.
- Possible Solution: Using R for analytics and the web browser for the visualisation.
- SmartR does just that!
- Pro: Highly dynamic and modern visualisations with strong analytical capabilities.
- Contra: This approach requires a deep integration with the web service → platform-bound

Why are visual analytics platform-bound?

This is what it looks like:

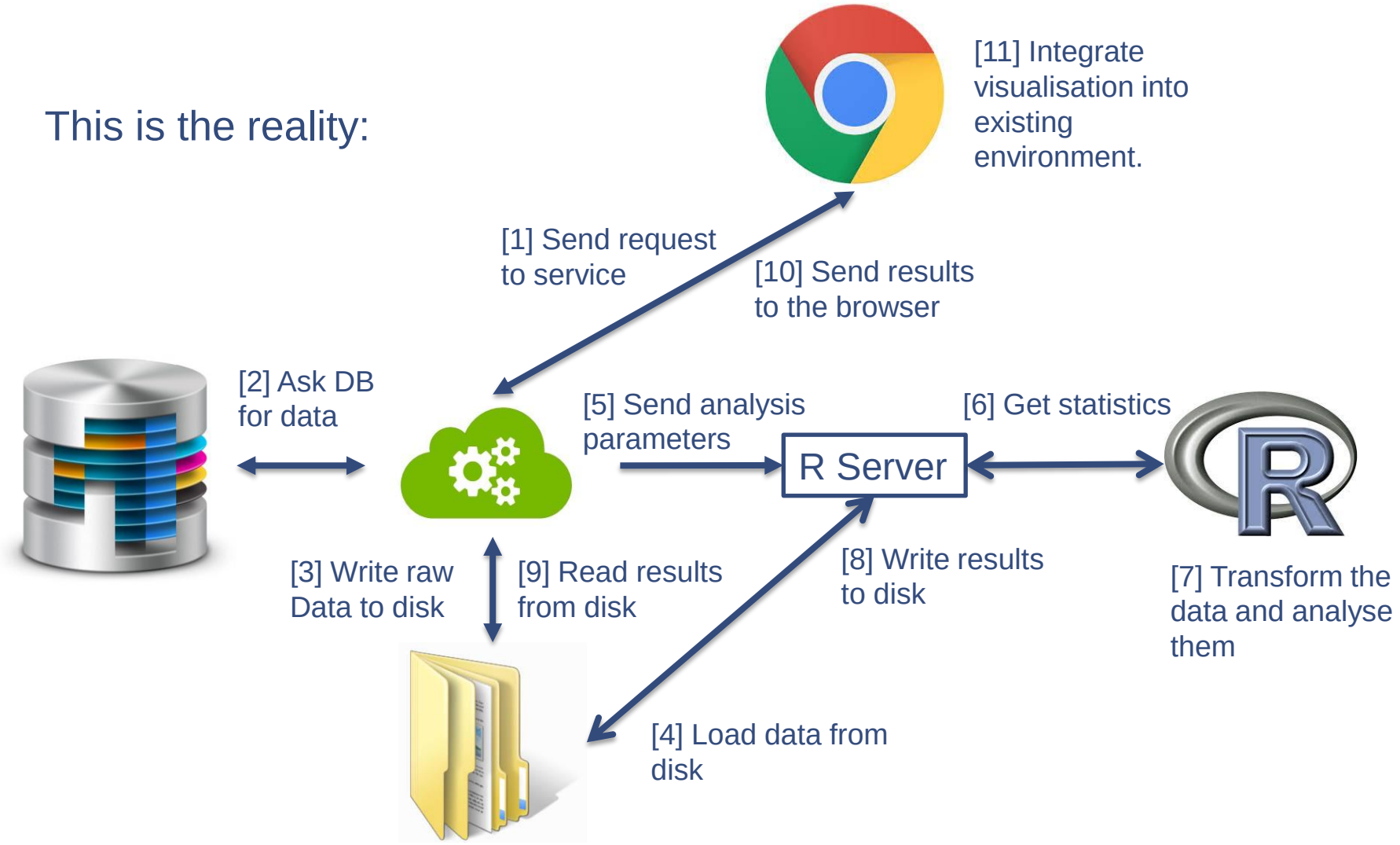
Visualize this for me!



Analyse this for me!

Why are visual analytics platform-bound?

This is the reality:



Why are visual analytics platform-bound?

- Things that have not been mentioned:
  - Session management
  - Clean-up services
  - Error handling
  - Job handling
  - Caching

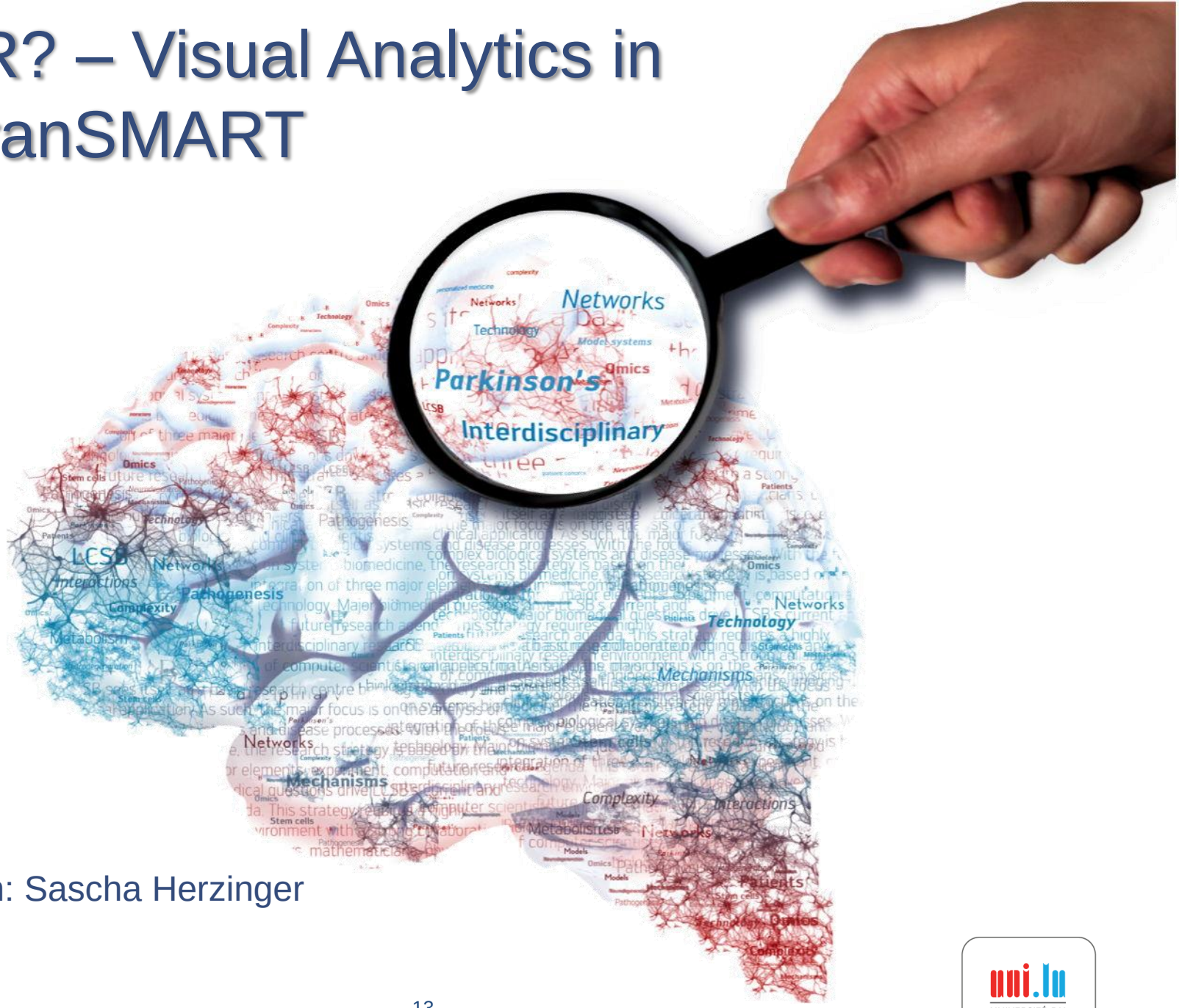
- Main reasons for platform-bondage:
  - Different programming language (Groovy, Python, Scala, …)
  - Different web framework (Grails, Django, Play, …)
  - Different internal class structure and architecture

The future of SmartR

- Must resolve platform-binding
- Must support multiple upcoming UIs
- Needs a more scalable solution than R / Rserve
- Needs constant maintenance, due to changing interfaces

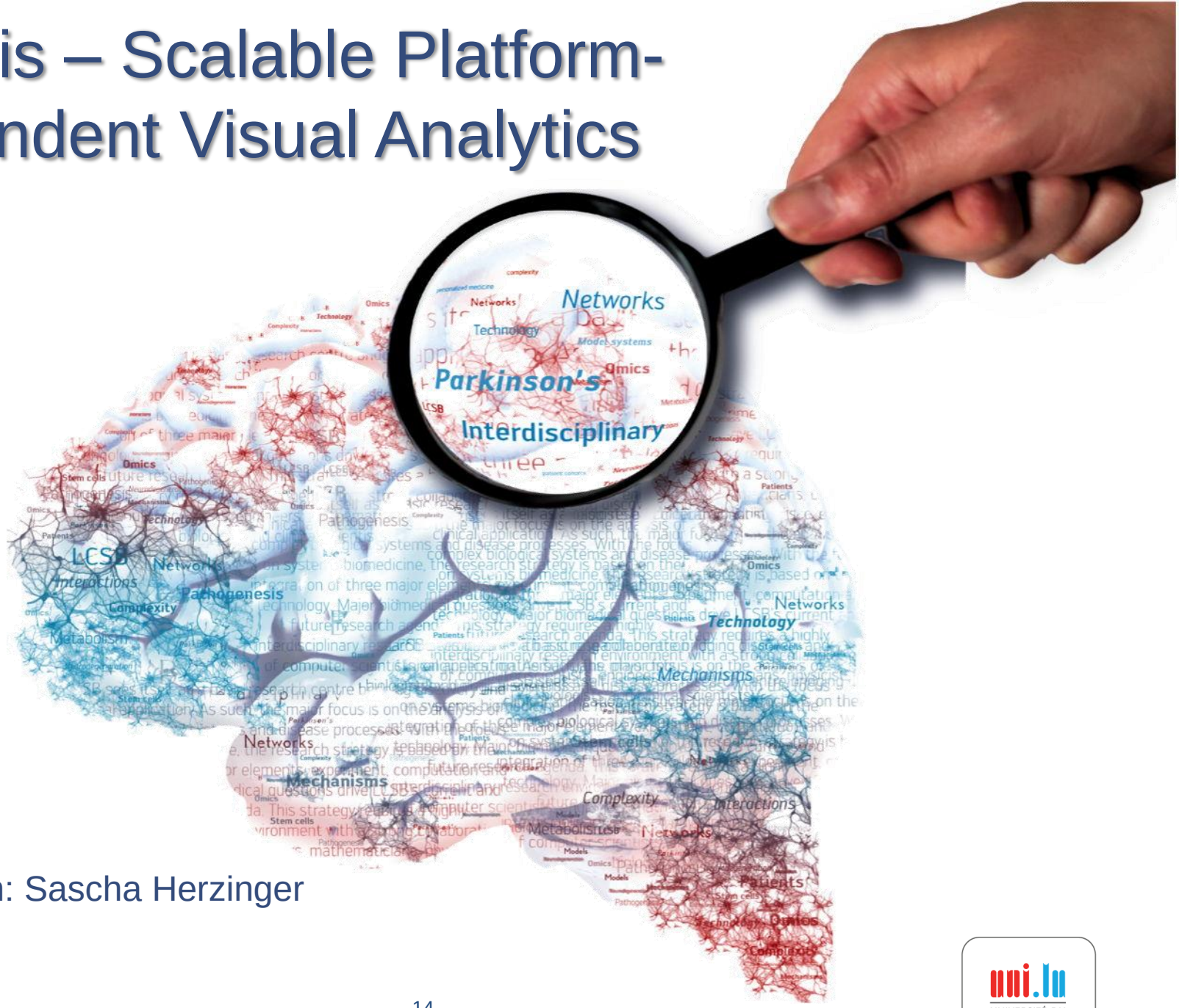
- Conclusion: **Dead End**

SmartR? – Visual Analytics in i2b2/TranSMART



Presentation: Sascha Herzinger

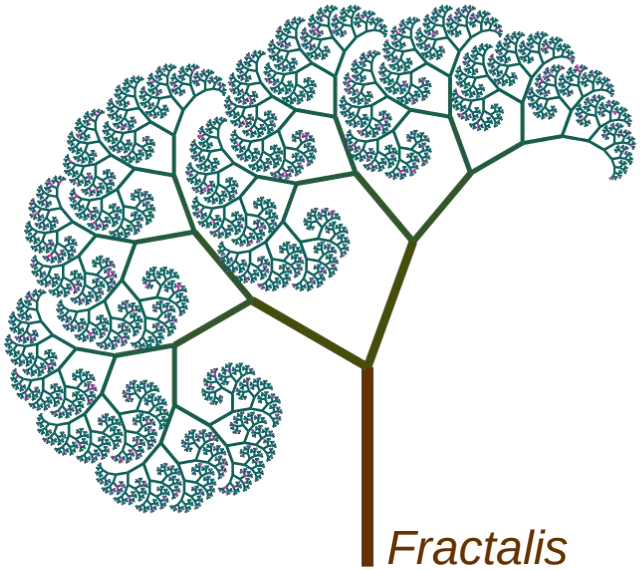
Fractalis – Scalable Platform-independent Visual Analytics



Presentation: Sascha Herzinger

A New Start

Name inspiration: https://en.wikipedia.org/wiki/Mandelbrot_set



Requirements:

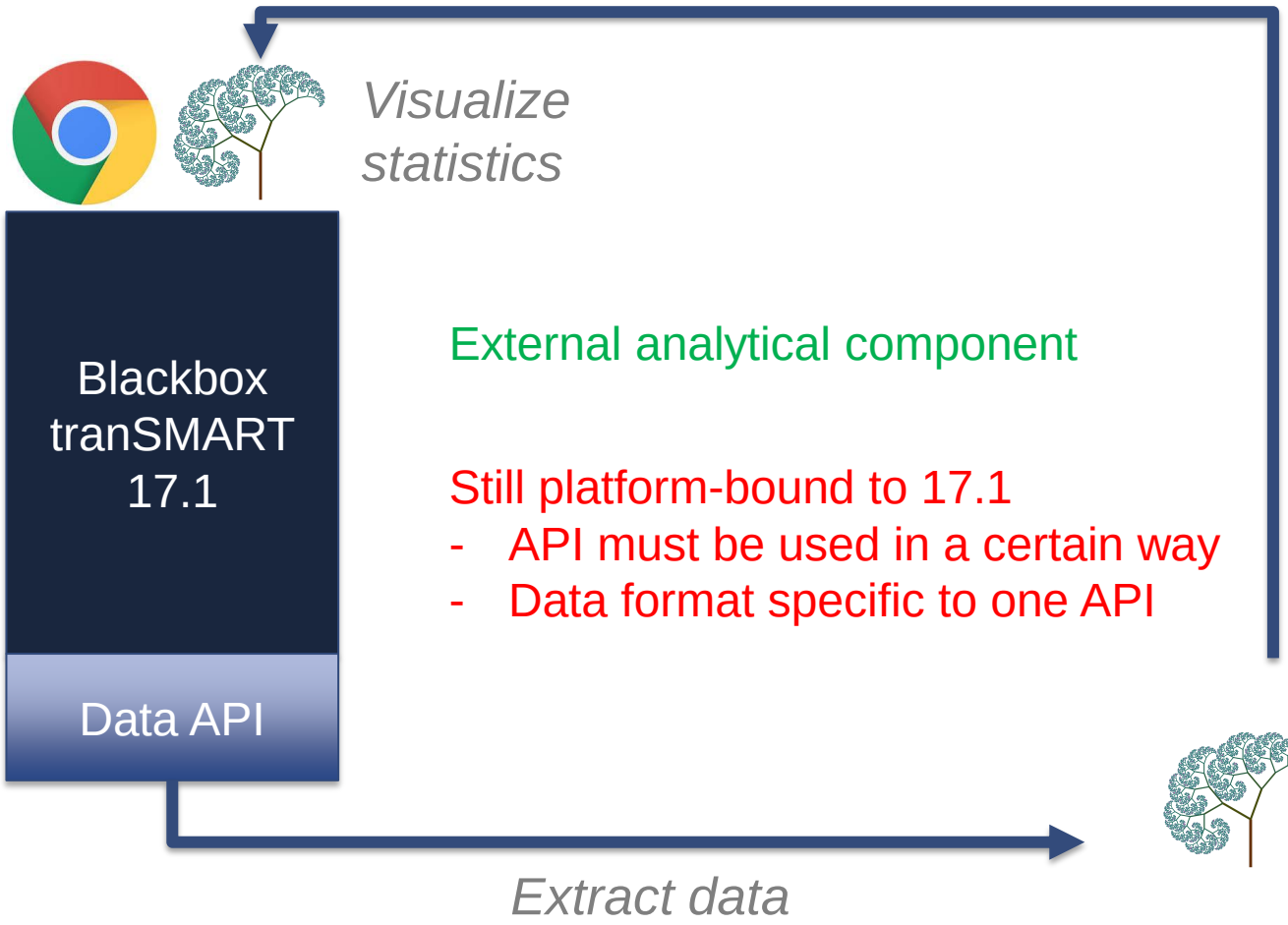
- Scalable
- Extendable
- Platform independent
- Robust to changes
- Documented
- Well tested
- Simple to integrate

How to fulfil these new requirements?

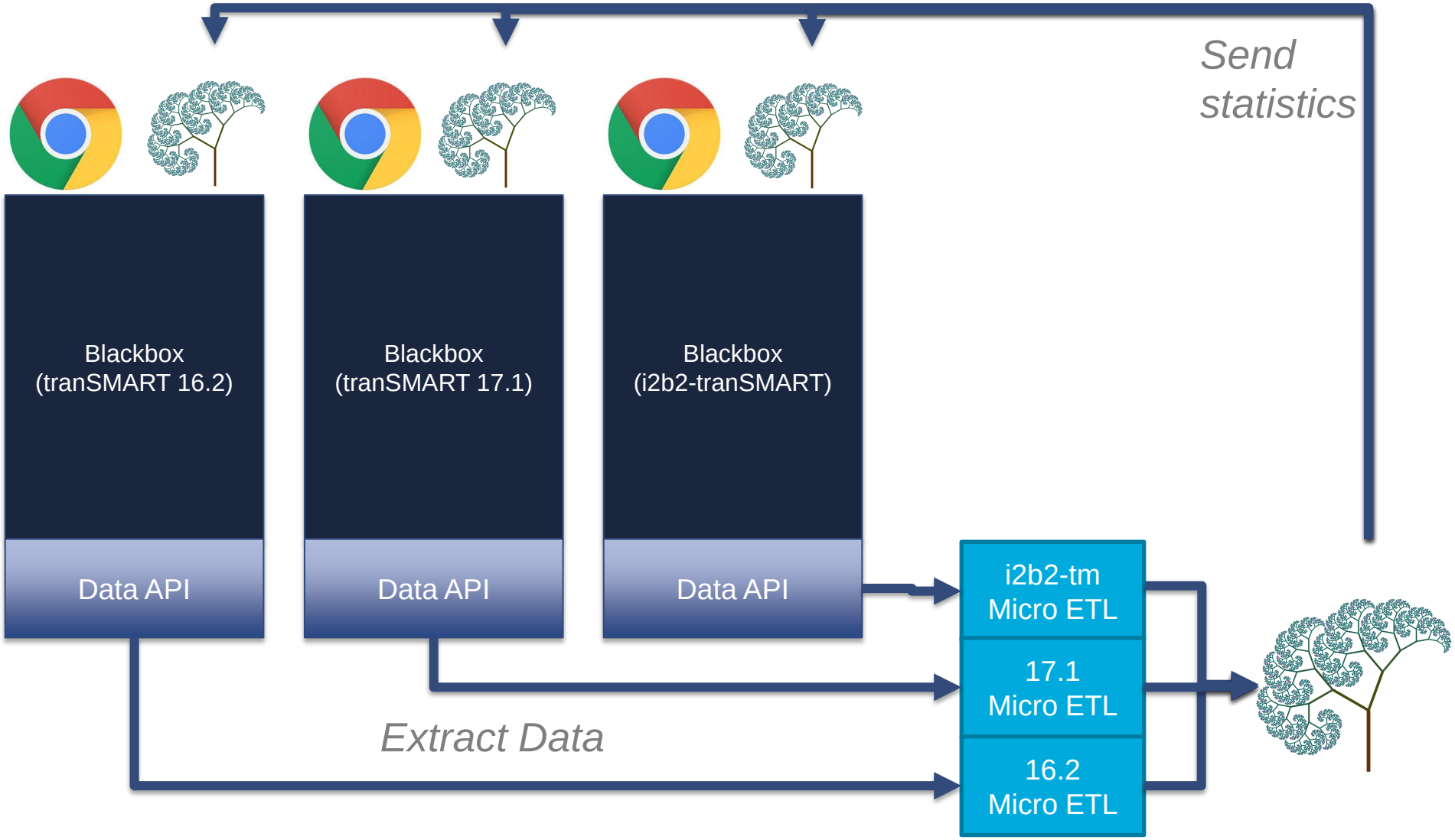
Blackbox

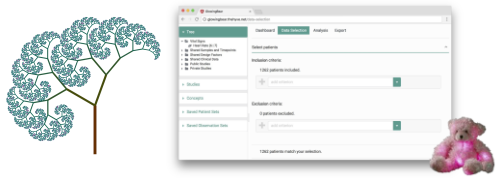
This means that we only care about what goes into this box and what comes out of it!

How to fulfil these new requirements?



How to fulfil these new requirements?





User Interface

Server

POST /data
Platform: 17.1
Data Type: Integer
Field Name: Age
Study: DeNoPa

Robustness & platform independence

Flask
web development,
one drop at a time

ETL
Factory

16.2 int ETL

17.1 cat ETL

17.1 int ETL

i2b2-tm cat ETL

redis

Task ID: 1234567890
Data type: numeric
Location: /tmp/foo

Submit Task

Celery

Celery

Celery

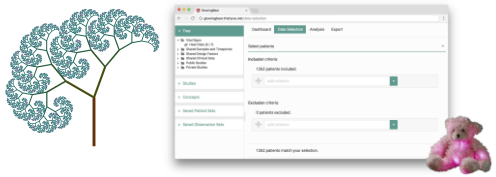
RabbitMQ

Store meta info

Shared
Cache

Store real data

Download Data



User Interface

Server

POST /analytics
Name: correlation
Params: {x: 1234567890, y: 0987654321}

Expandable

PCA

Clustering

Regression

Correlation

Flask
web development,
one drop at a time

Task
Factory

redis

Task ID: 555123
P-value: 0.05
Coefficient: 1

Submit Task

Celery

Celery

Celery

RabbitMQ

Task ID: 1234567890
Data type: numeric
Location: /tmp/foo

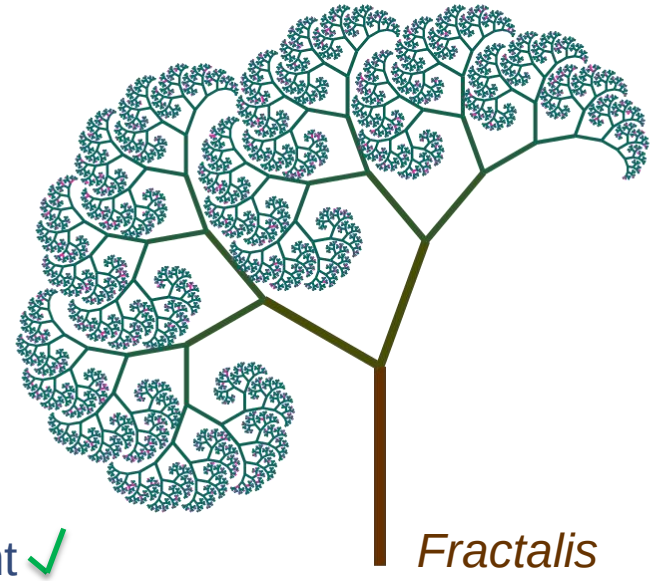
Shared
Cache

Actual Data

Scalable

A New Start

Name inspiration: https://en.wikipedia.org/wiki/Mandelbrot_set



Fractalis

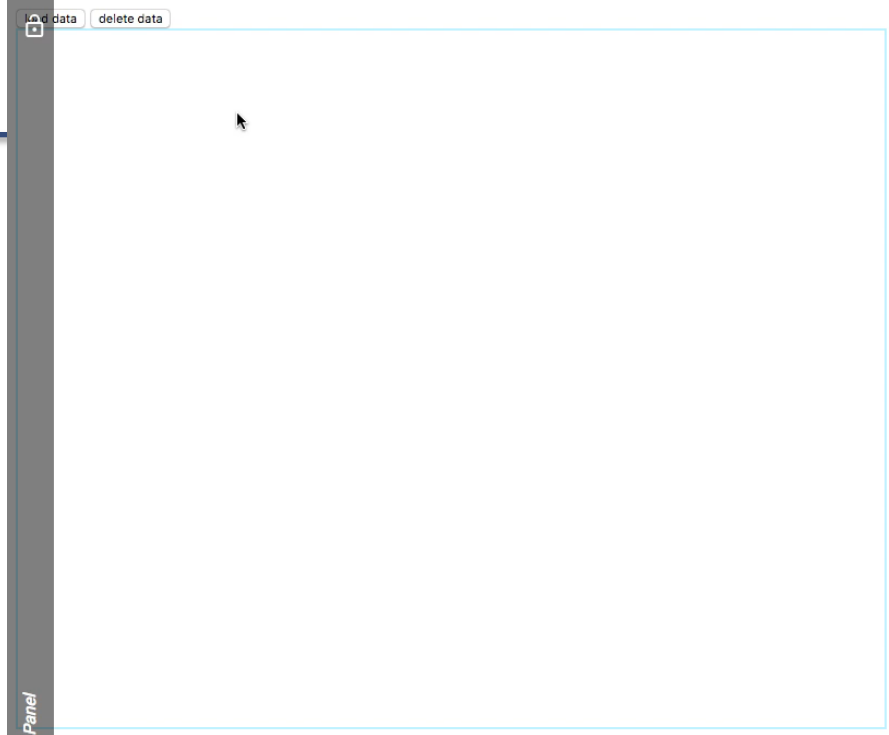
Requirements:

- Scalable ✓
- Extendable ✓
- Platform independent ✓
- Robust to changes ✓
- Documented ✓
- Well tested ✓
- Simple to integrate ✓


d data
delete data
Control Panel

load data delete data

Control Panel


load data
delete data
Control Panel

load data delete data

Control Panel

Special Thanks

- To all my colleagues at the LCSB who gave me inspiration and advice

- To all the people and organisations I worked together with during the last two years (eTRIKS, The Hyve, ITTM, and many more)

Some important logos