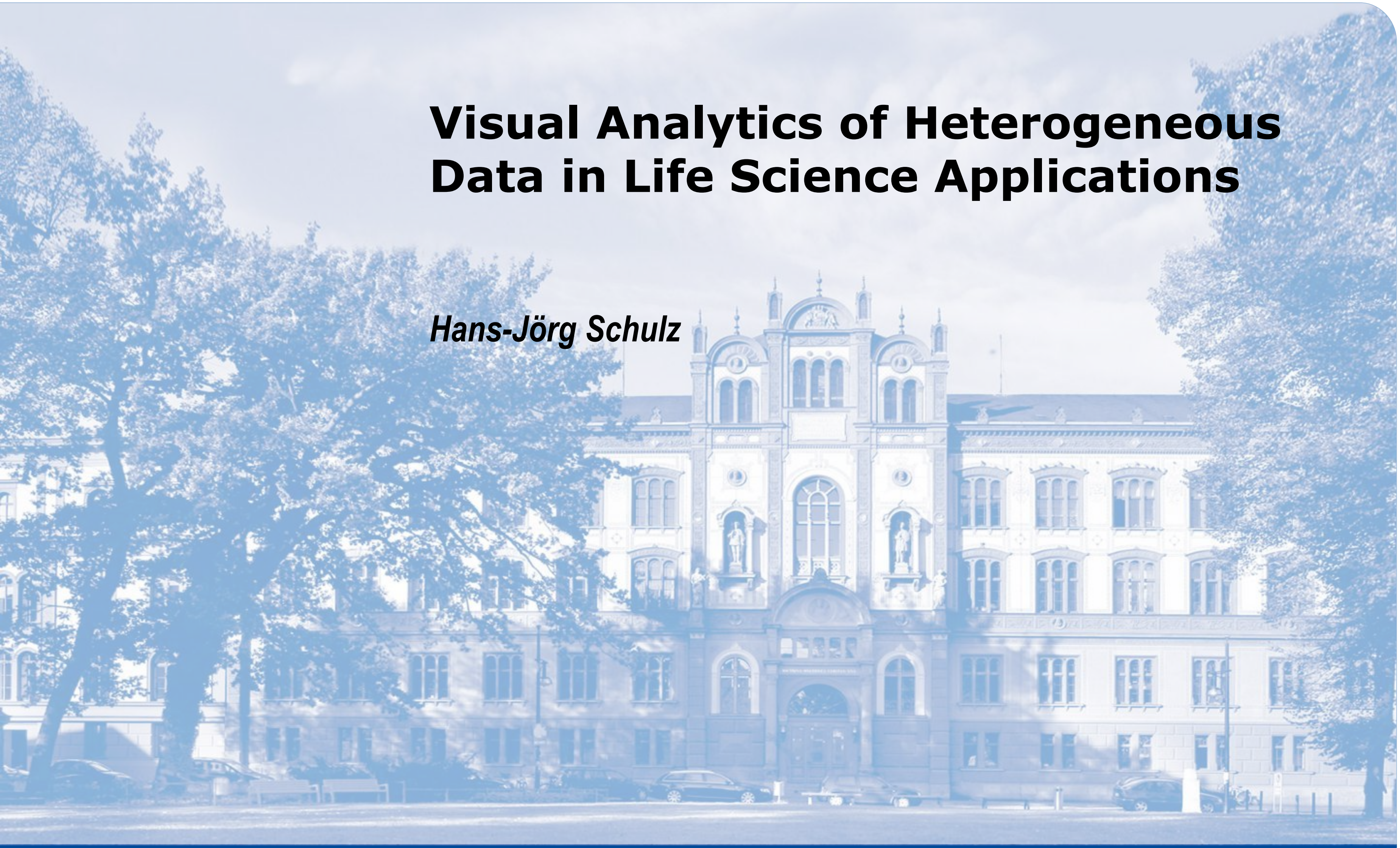
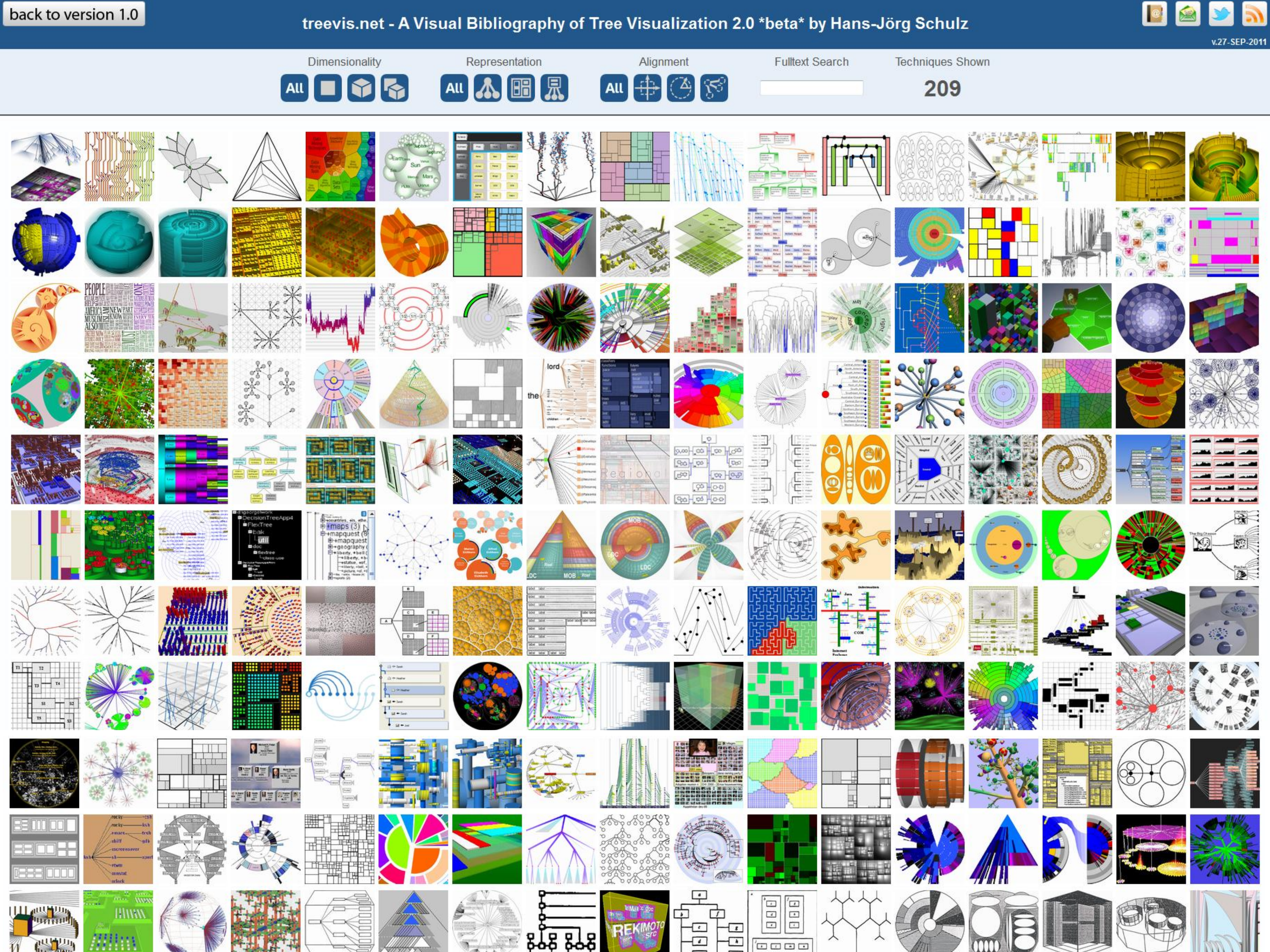
Universität Rostock

Traditio et Innovatio



Visual Analytics of Heterogeneous Data in Life Science Applications

Hans-Jörg Schulz



treevis.net - A Visual Bibliography of Tree Visualization 2.0 *beta* by Hans-Jörg Schulz

v.27-SEP-2011

Fulltext Search:

Techniques Shown: 209

Agenda

I. Motivation, Definitions

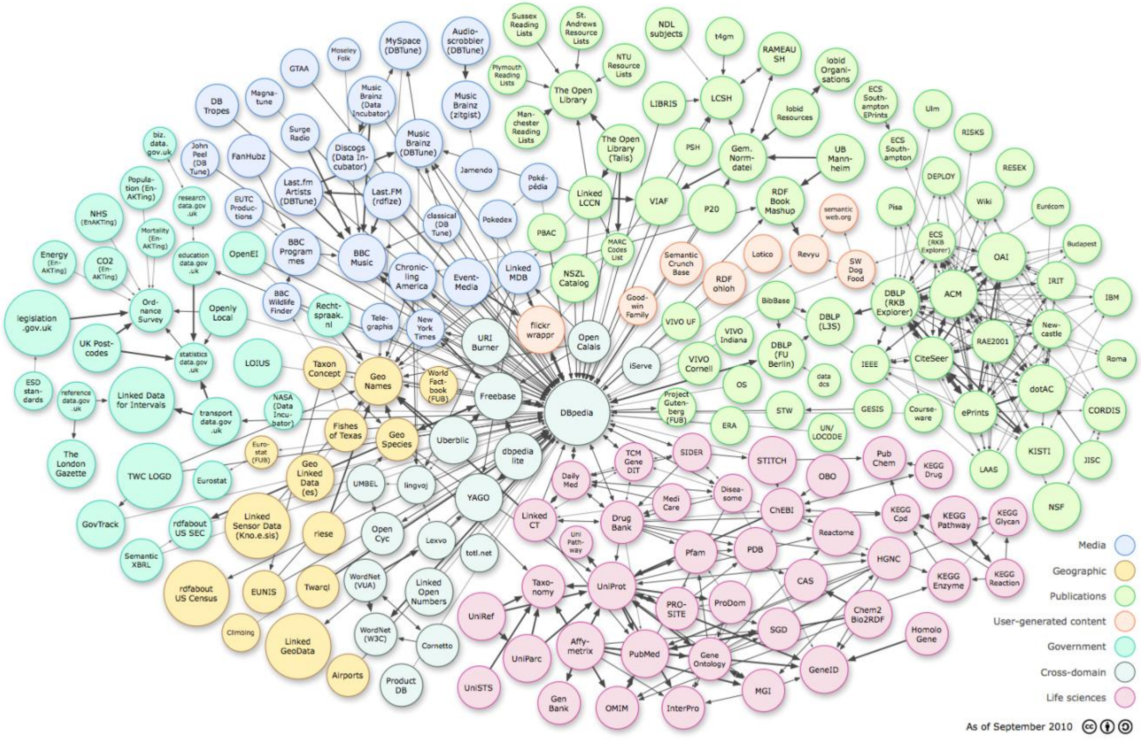
II. Visual Analytics of Inhomogeneous Data

III. Orientation and Navigation in Heterogeneous Data

IV. Conclusion, Food for Thought

I. Motivation, Definitions

Linked Data, Open Data



Source: Richard Cyganiak, Anja Jentzsch

I. Motivation, Definitions

Challenges of Multiple Data Sources

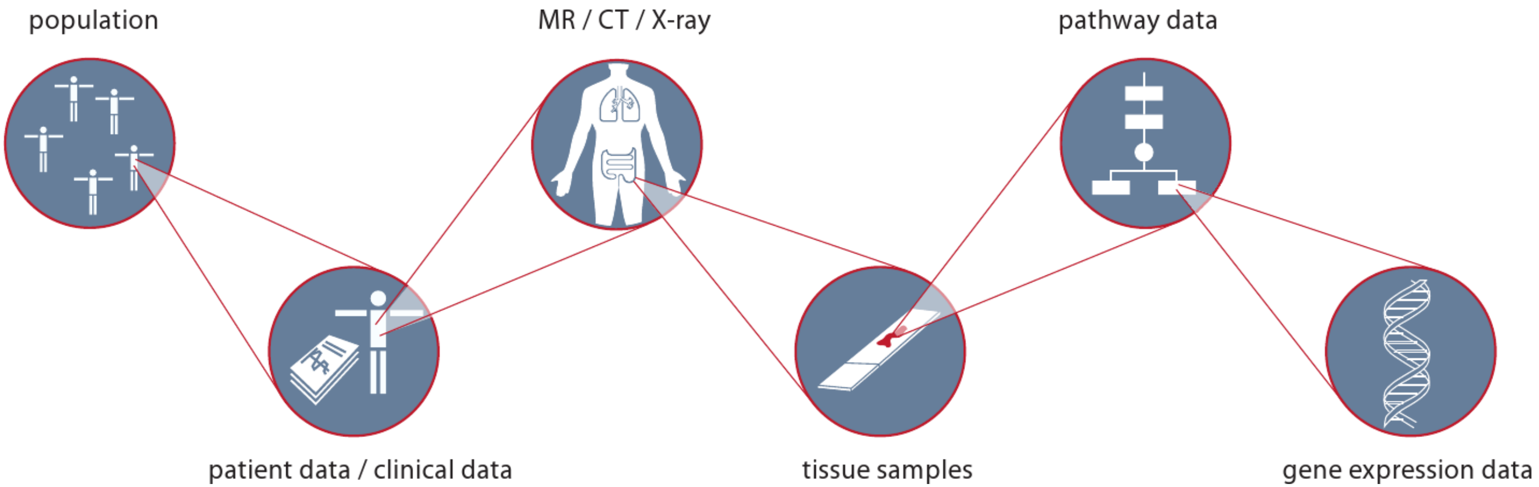
- Multiple **origins** (different measurement precisions, different languages,…)
- Multiple **formats** (for querying – SQL, SPARQL,… and for query results – JSON, XML,…)
- Multiple **access restrictions** (and authorities to grant access are also distributed)
- Multiple **data types** (images, documents, numerical values, graphs/structures,…)
- Multiple **data scales** (different value ranges)

- Multiple **analysis methods** (R or Weka for numerical data, Lucene and LingPipe for text,…)
- Multiple **visualization techniques** (image viewers, text visualization, charts+plots,…)

How to do Visual Analytics in such a scenario?

I. Motivation, Definitions

Heterogeneous Data in Biomedical Applications



+ pharmaceutical data bases

+ PubMed publications data base

+ disease data bases (ICD-10, DSM-IV,…)

+ gene and protein data bases (NCBI)

I. Motivation, Definitions

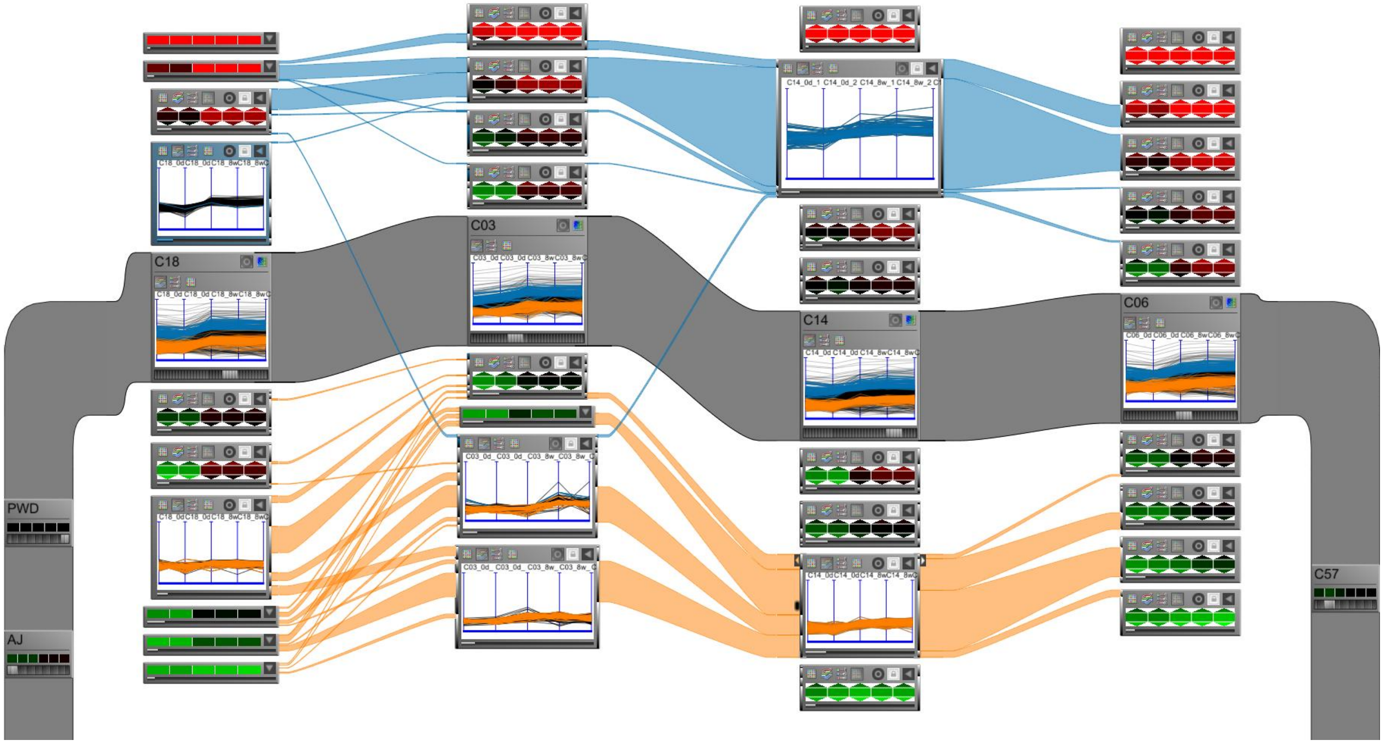
Heterogeneous Data

Data drawn from multiple separate data sets/data bases.

Inhomogeneous Data

Data from a single data set/data base which is non-uniformly distributed, contains values from different scales, or missing values.

II. Visual Analytics of Inhomogeneous Data



Alexander Lex, **Hans-Jörg Schulz**, Marc Streit, Christian Partl, and Dieter Schmalstieg: *VisBricks: Multiform Visualization of Large, Inhomogeneous Data*, appeared at InfoVis'11

II. Visual Analytics of Inhomogeneous Data

Premise

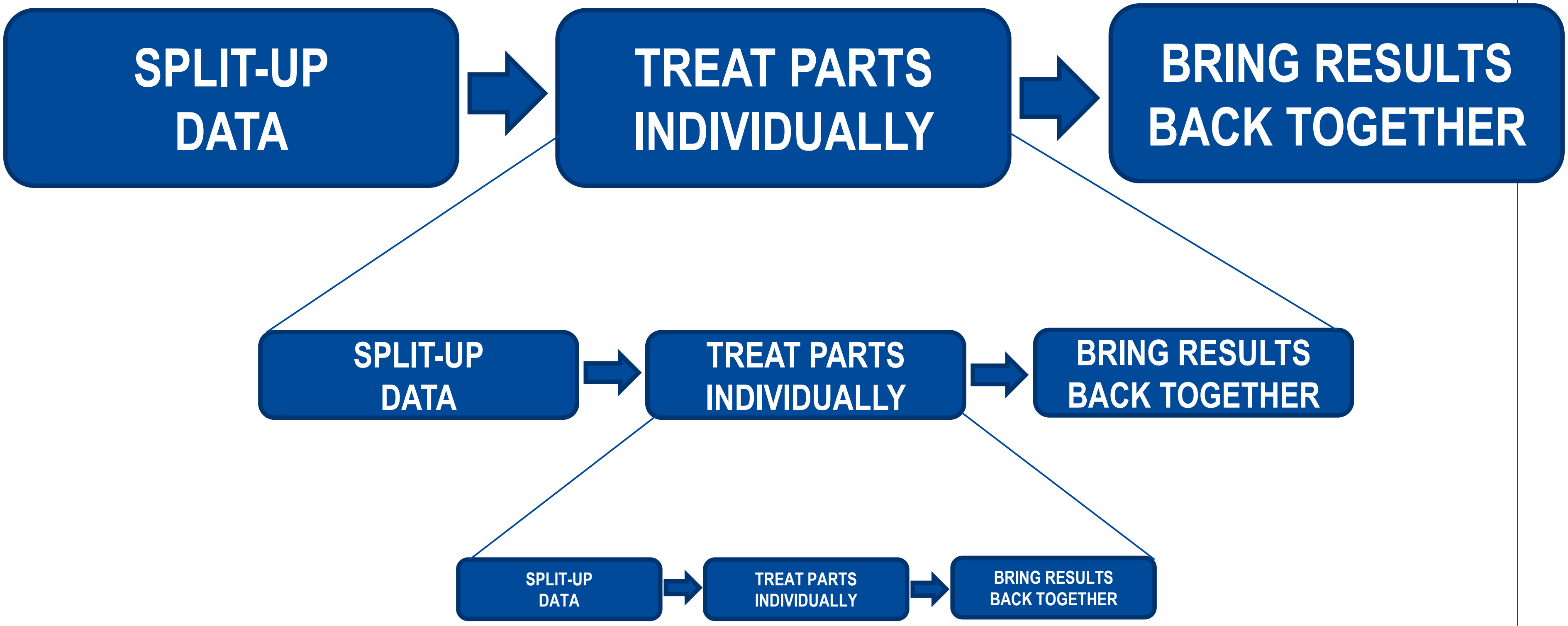
Different (homogeneous) subsets of an inhomogeneous data set

- exhibit different data characteristics
- which must be analyzed differently
- and shown differently

within the context of the whole data set.

II. Visual Analytics of Inhomogeneous Data

Overall Approach: Divide & Conquer



SPLIT-UP DATA

TREAT PARTS INDIVIDUALLY

BRING RESULTS BACK TOGETHER

SPLIT-UP DATA

TREAT PARTS INDIVIDUALLY

BRING RESULTS BACK TOGETHER

SPLIT-UP DATA

TREAT PARTS INDIVIDUALLY

BRING RESULTS BACK TOGETHER

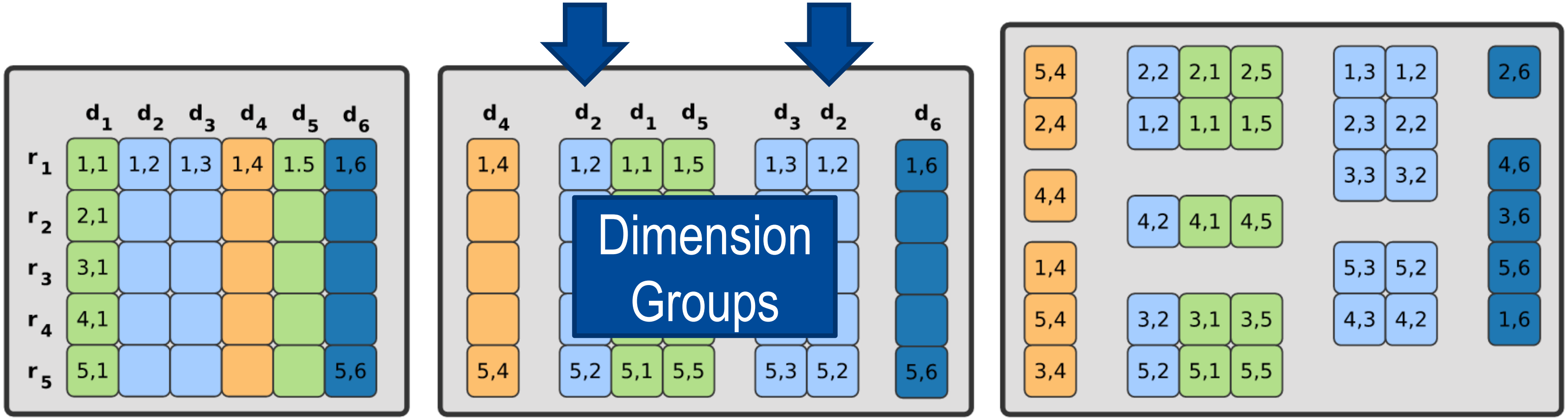
II. Visual Analytics of Inhomogeneous Data

The Divide Step: Possible Inhomogeneities in Tabular Data

	Dimensions	Records
Semantics	Columns First Name + Last Name *vs.* Age + ZIP Code	Symptoms Cough + Fever *vs.* Headache + Dangling Ankle
Characteristics	2 Columns of Scale [10^5...10^6] *vs.* Columns of Scale [0..1] + [10^5...10^6]	Undefined values *vs.* Defined values
Statistics	Correlated Columns *vs.* Uncorrelated Columns	Records from the same cluster *vs.* Records from different clusters

II. Visual Analytics of Inhomogeneous Data

The Divide Step: 2-Step-Subdivision of Inhomogeneous Data



Raw Input Data

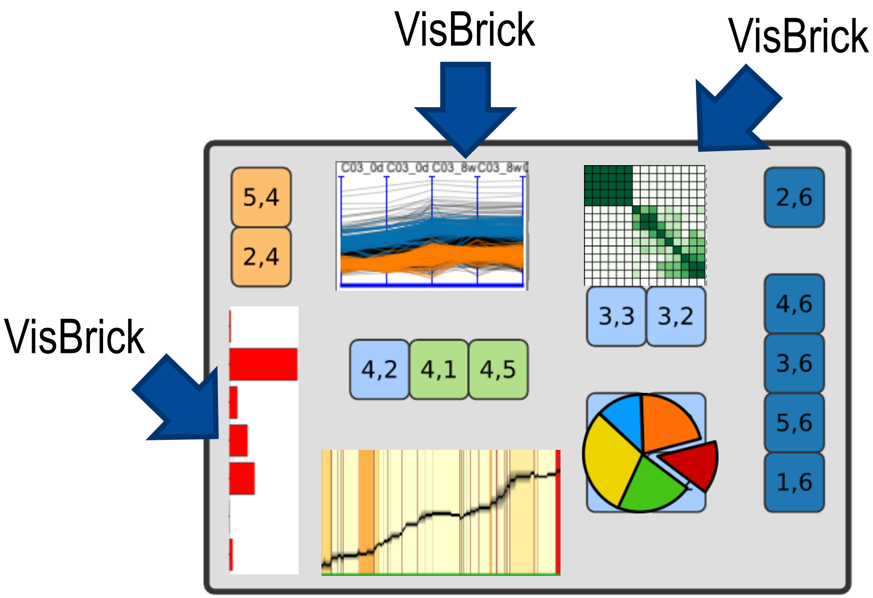
Step 1: Divide Dimensions

Step 2: Divide Records

Note: Division does not need to be disjoint – a dimension can appear in multiple groups.

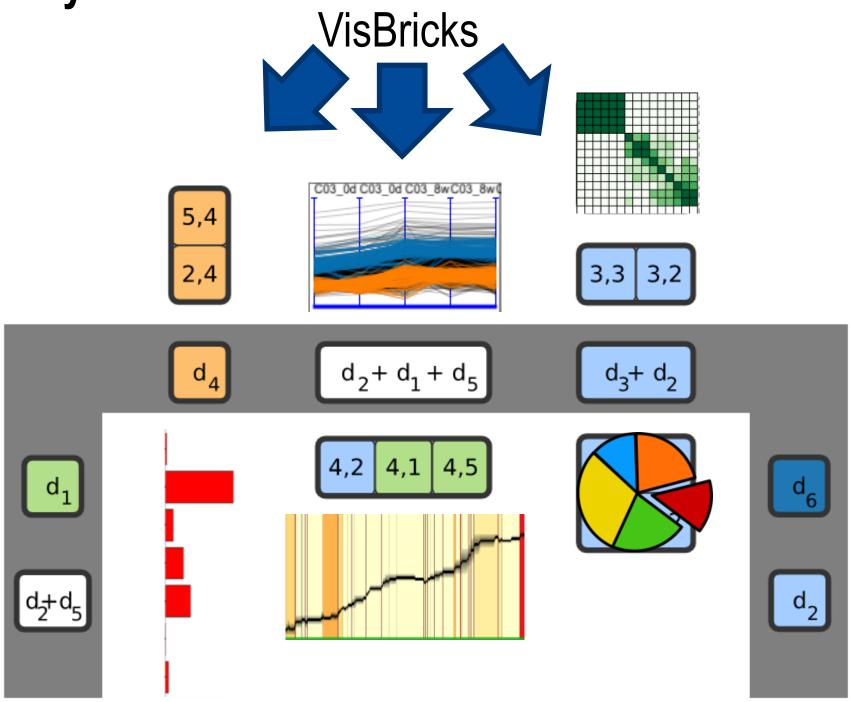
II. Visual Analytics of Inhomogeneous Data

Treat them differently:



II. Visual Analytics of Inhomogeneous Data

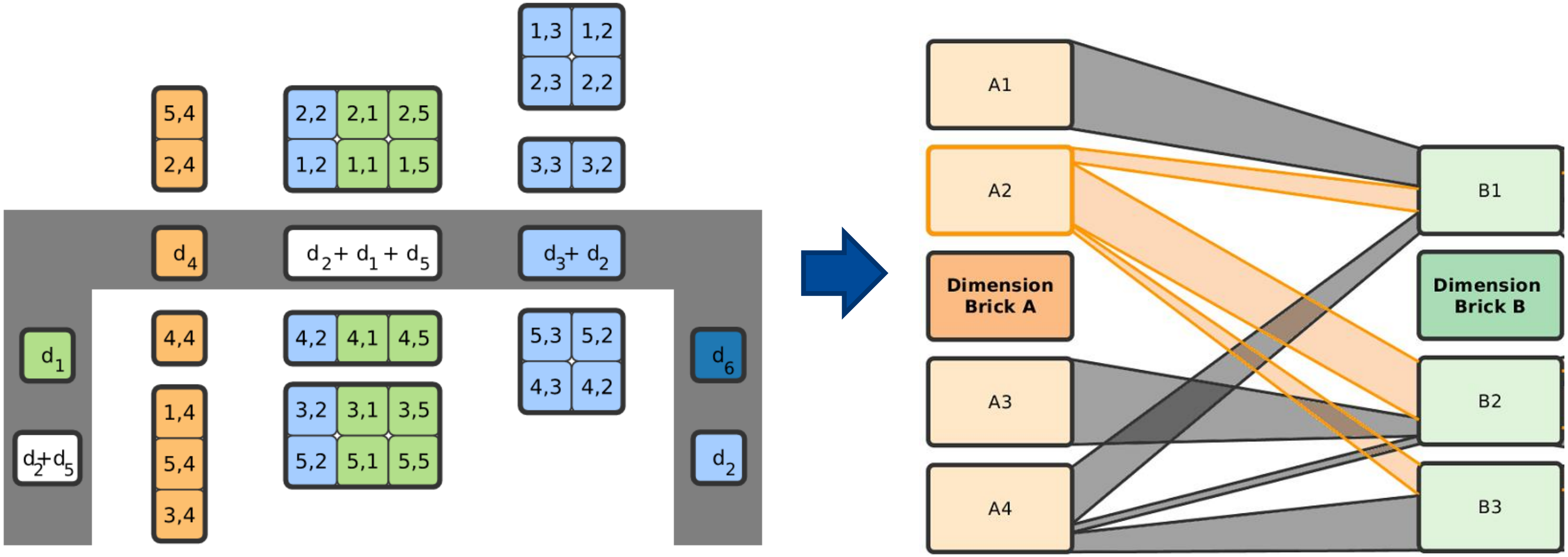
The Conquer Step: Layout



Multiform Visualization

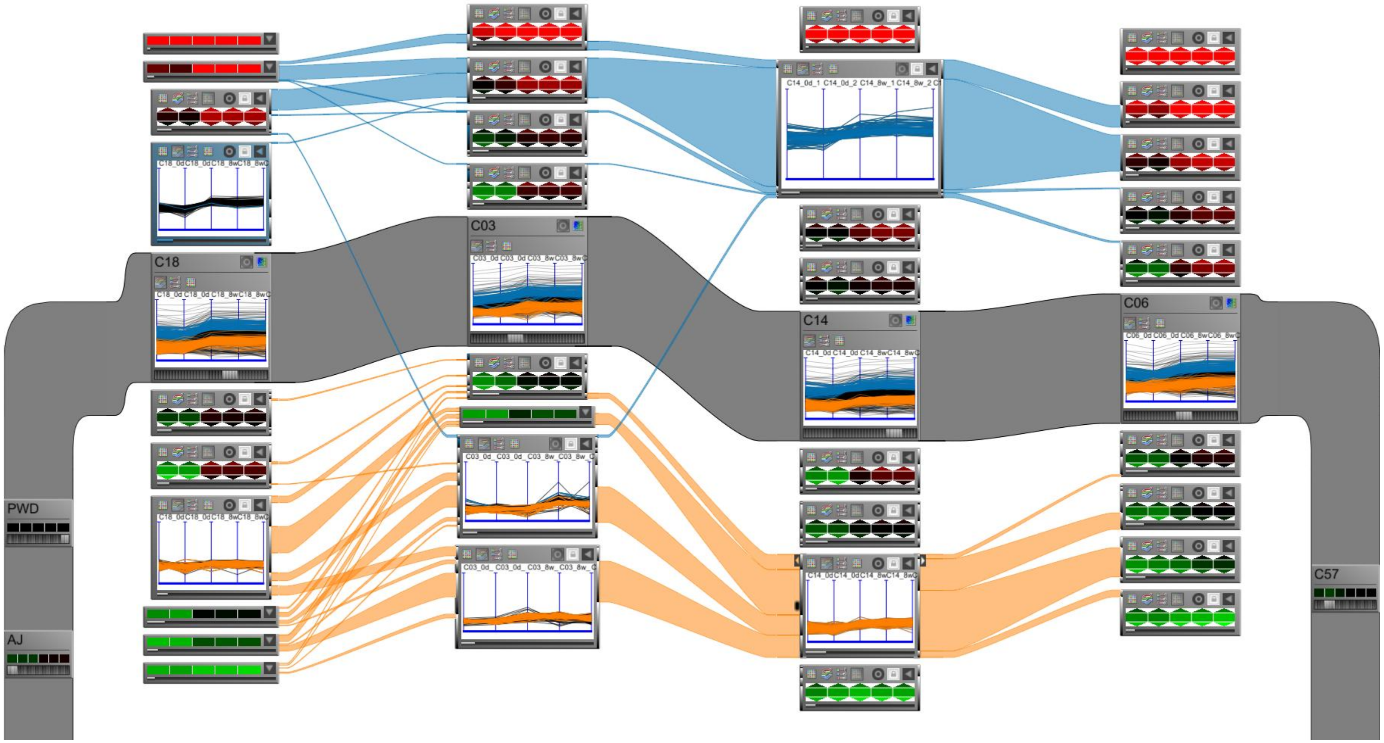
II. Visual Analytics of Inhomogeneous Data

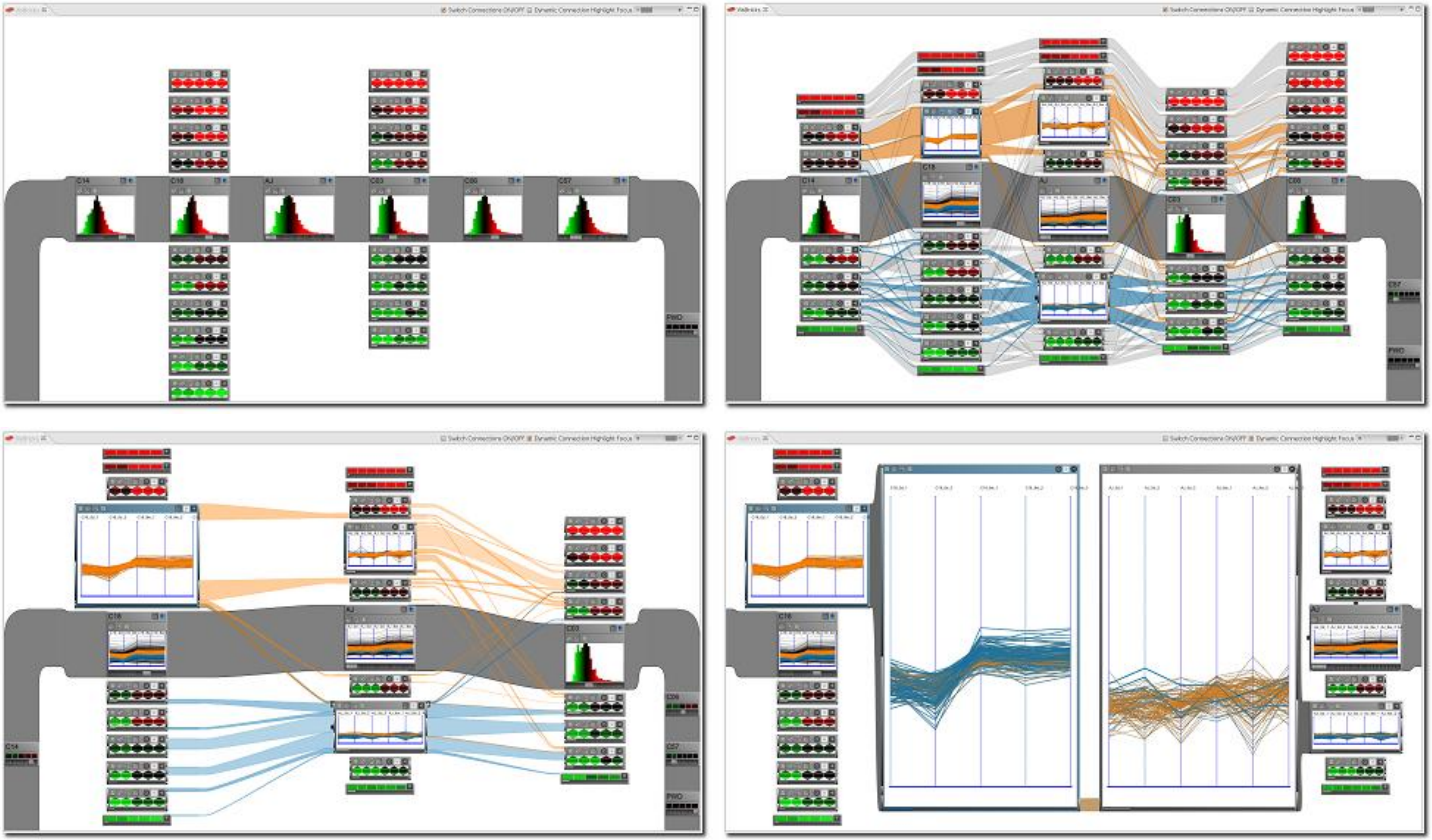
The Conquer Step: Linking



II. Visual Analytics of Inhomogeneous Data

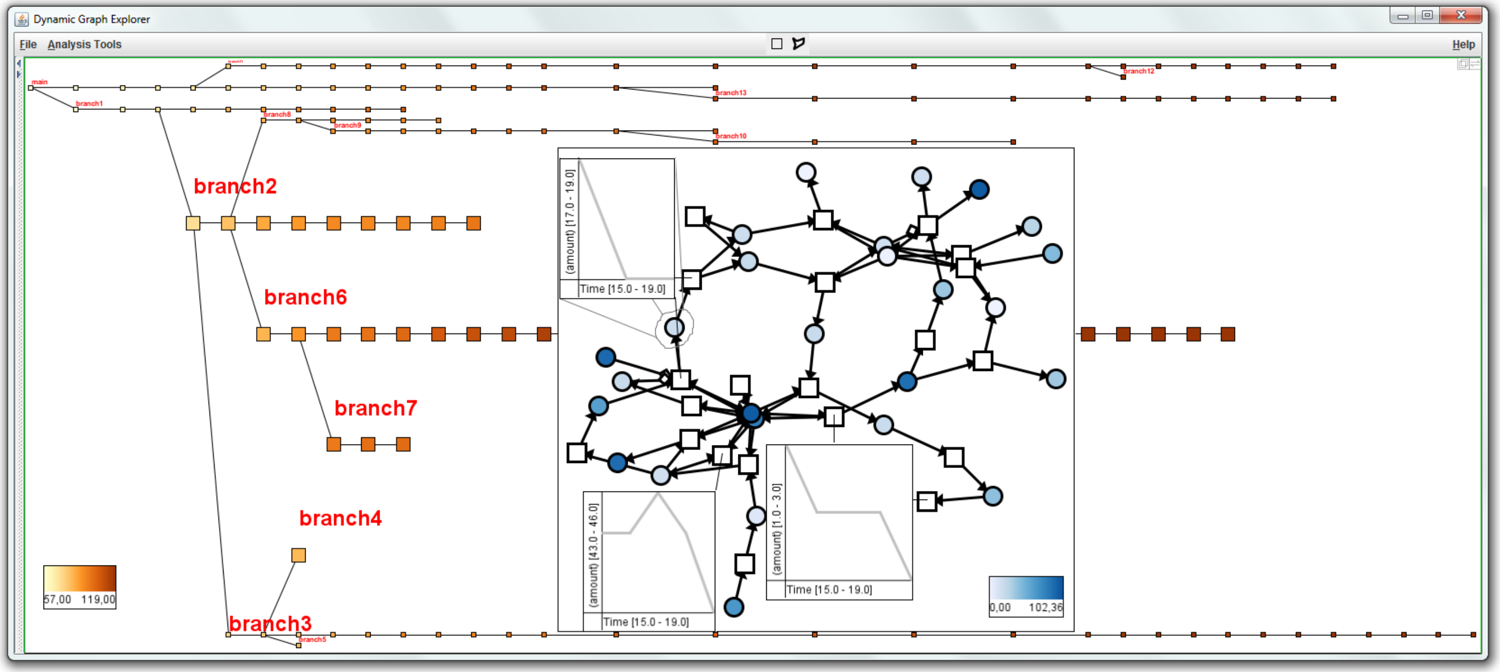
The Result




Switch Connections ON/OFF
Dynamic Connection Highlight Focus
C14
C18
AJ
C03
C00
C07
PWD

II. Visual Analytics of Inhomogeneous Data

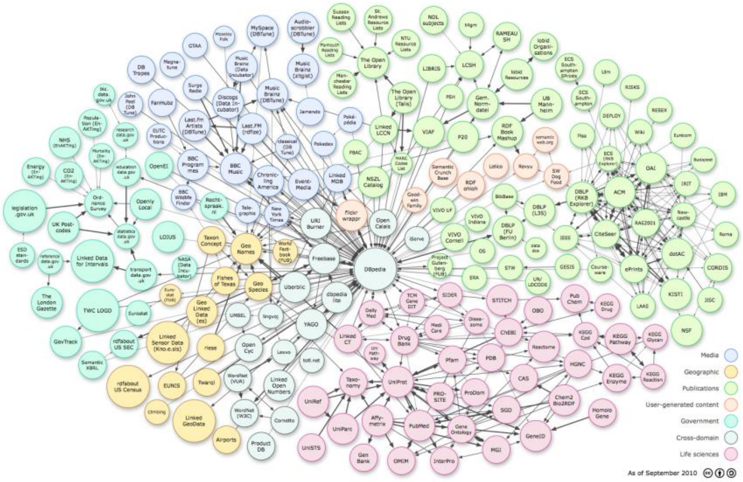
Current Research & Future Work: Extension to graph-structured data



Steffen Hadlak, **Hans-Jörg Schulz**, and Heidrun Schumann:
In Situ Exploration of Large Dynamic Networks, appeared at InfoVis'11

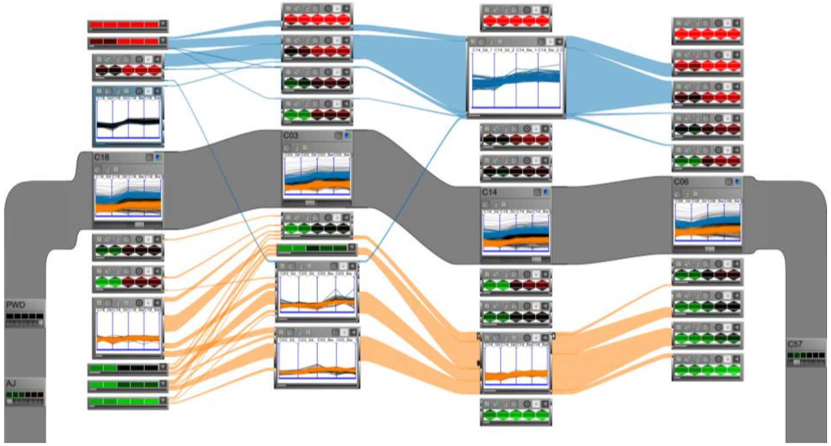
II. Visual Analytics of Inhomogeneous Data

Current Research & Future Work: Extension to heterogeneous data?



Source: Richard Cyganiak, Anja Jentzsch

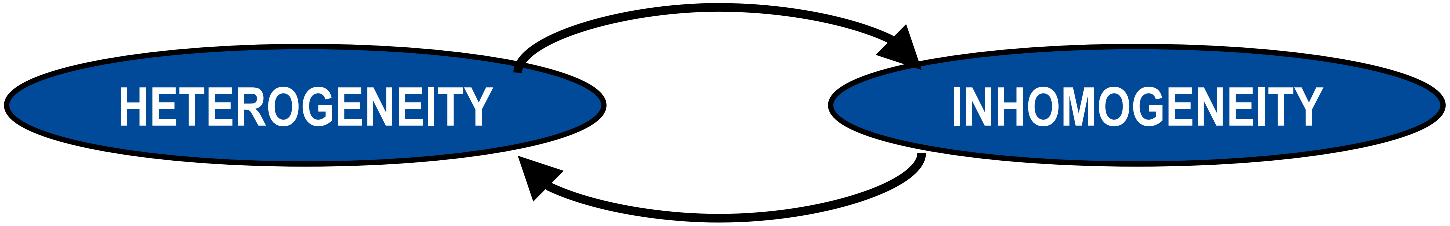
+



II. Visual Analytics of Inhomogeneous Data

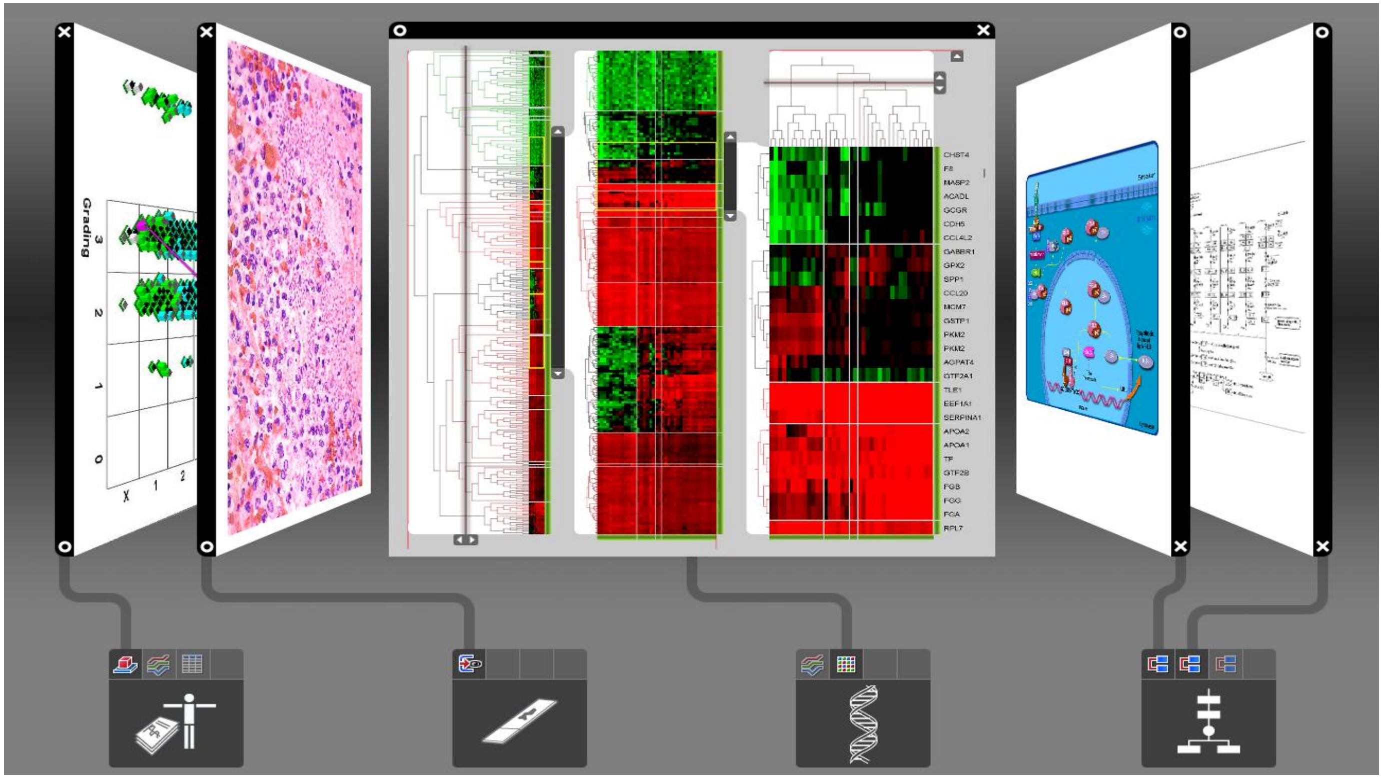
Blatantly overreaching conjecture:
The Heterogeneity-Inhomogeneity-Duality (*working title*)

Data Federation/Warehousing/Fusion
(*conquer*)



Data Clustering/Partitioning
(*divide*)

III. Orientation and Navigation in Heterogeneous Data

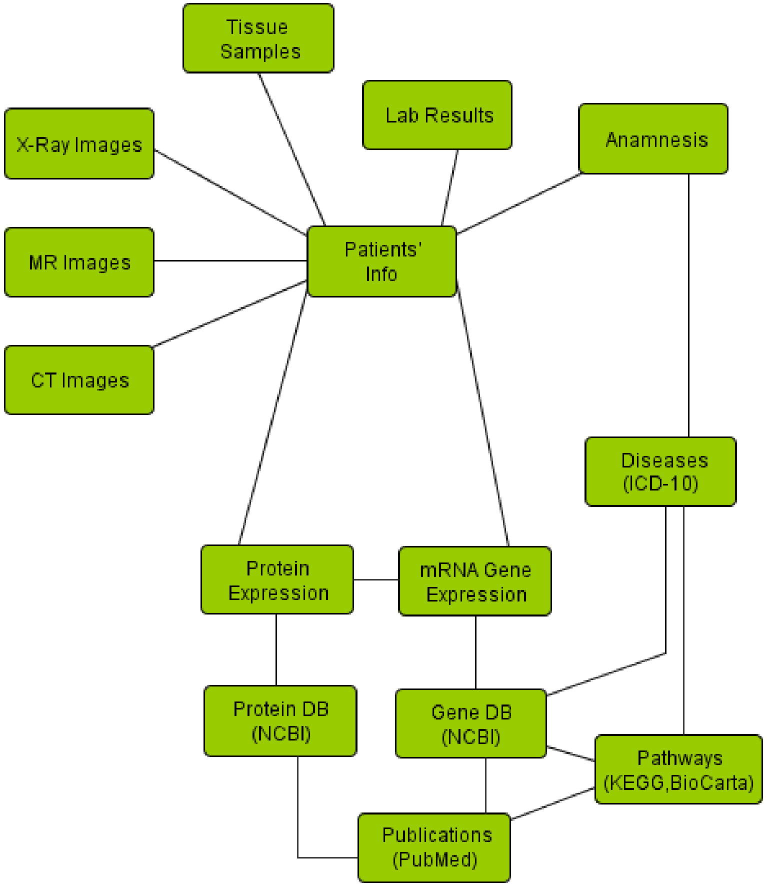


Marc Streit, **Hans-Jörg Schulz**, Alexander Lex, Dieter Schmalstieg, and Heidrun Schumann: *Model-Driven Design for the Visual Analysis of Heterogeneous Data*, to appear in IEEE TVCG

III. Orientation and Navigation in Heterogeneous Data

Data Heterogeneity

- multiple data sources
- which are linked via IDs, etc.

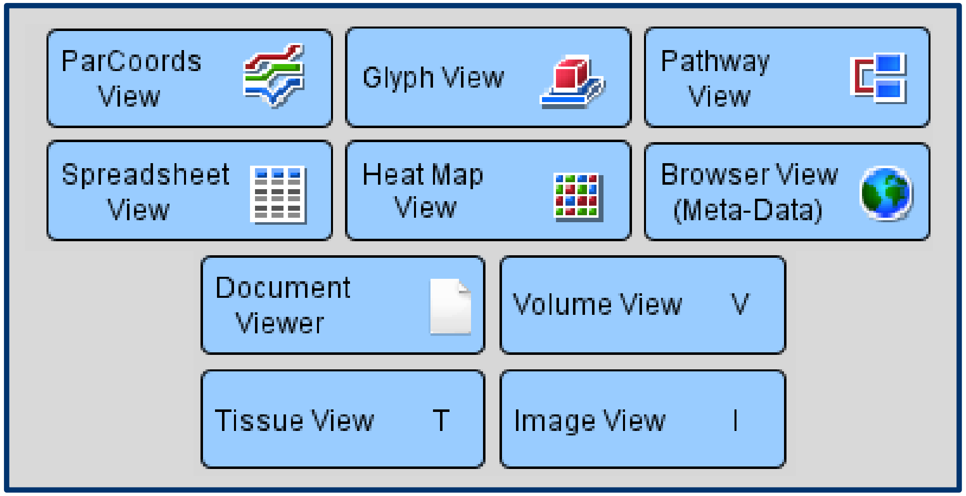


III. Orientation and Navigation in Heterogeneous Data

Data Heterogeneity

- multiple data sources
- which are linked via IDs, etc.

→ **Visual Heterogeneity**



ParCoords View

Glyph View

Pathway View

Spreadsheet View

Heat Map View

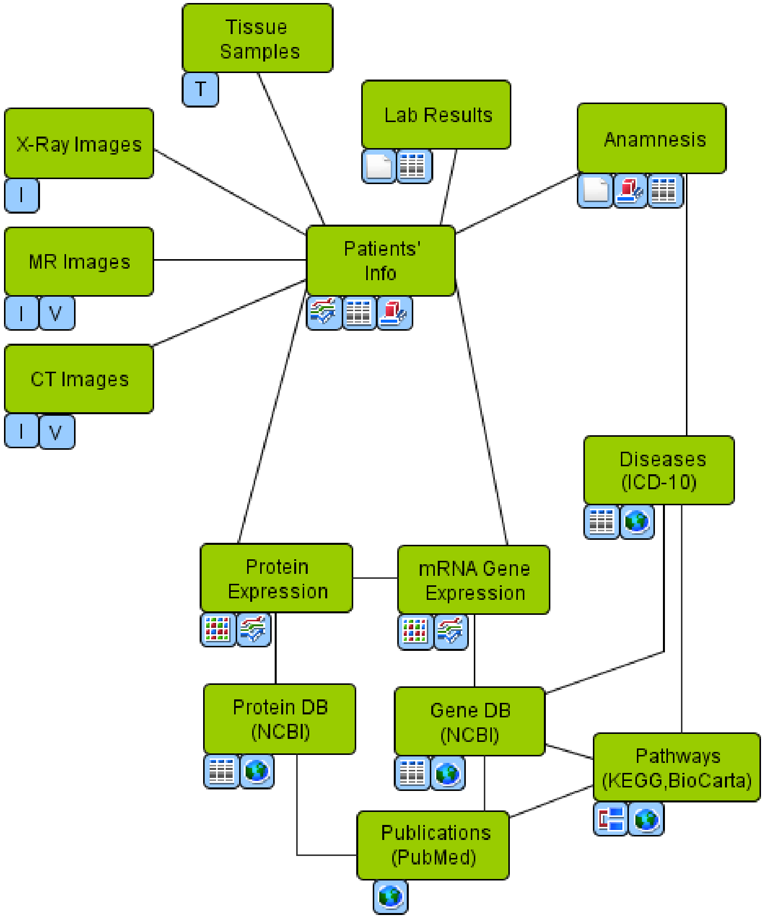
Browser View (Meta-Data)

Document Viewer

Volume View V

Tissue View T

Image View I



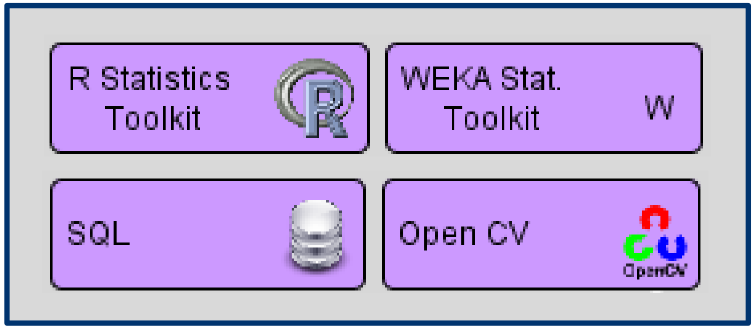
III. Orientation and Navigation in Heterogeneous Data

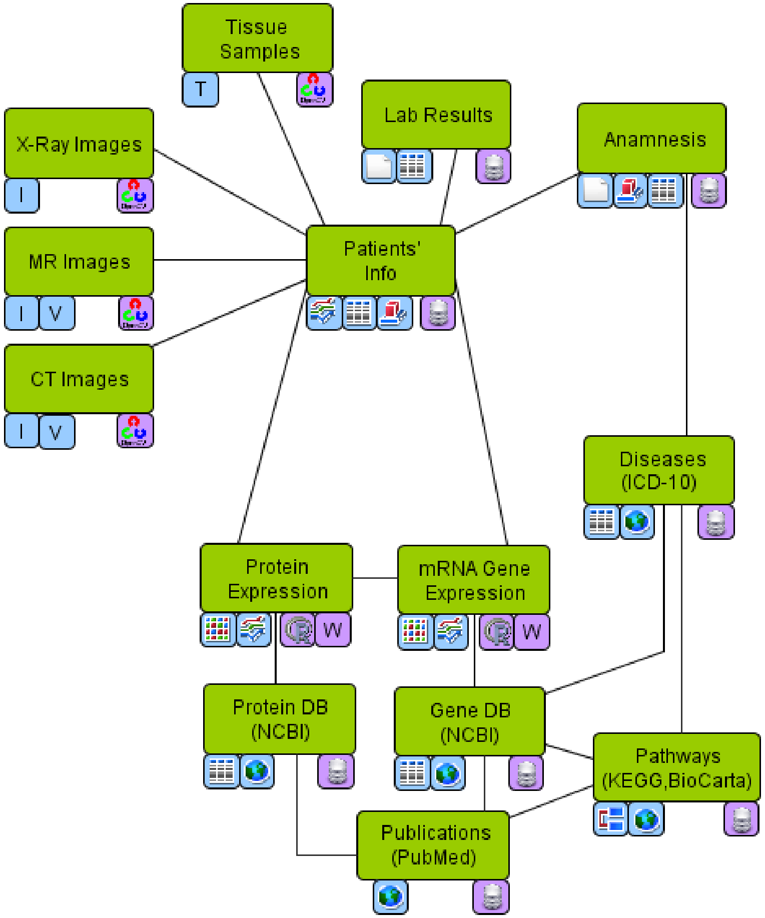
Data Heterogeneity

- multiple data sources
- which are linked via IDs, etc.

→ **Visual Heterogeneity**

→ **Analytical Heterogeneity**





III. Orientation and Navigation in Heterogeneous Data

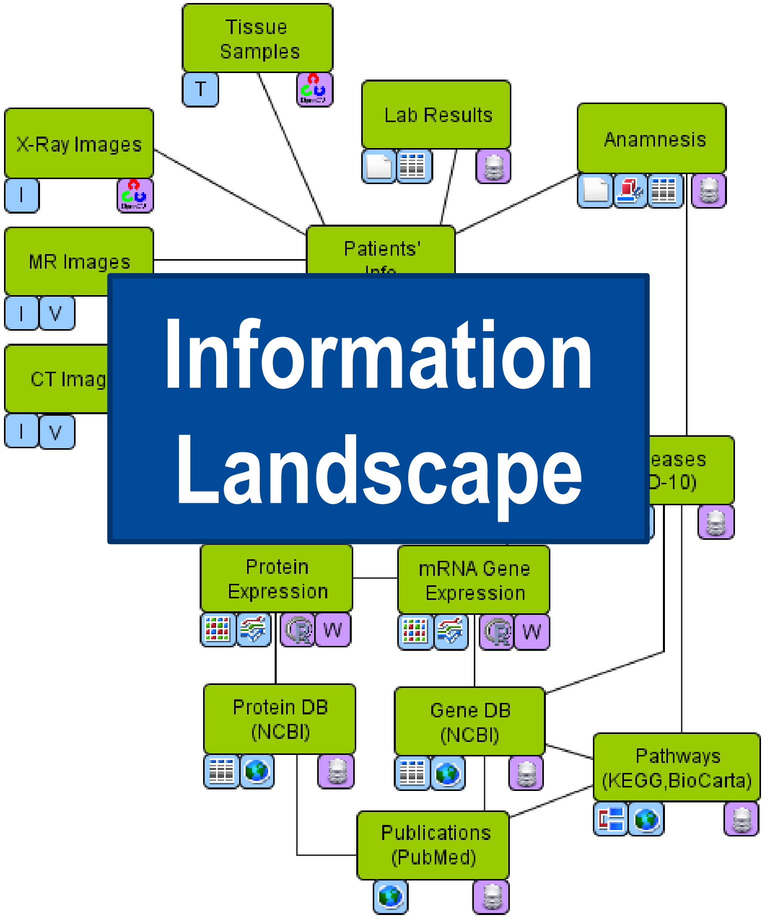
Orientation:

Where am I and **where** can I go from here?

Navigation:

<u>Given a goal</u>, **which** visual and/or analytical interface to use on **which** data set with **which** objective and in **which** order to reach this goal?

A typical goal is, for example:
treatment planning for cancer patients

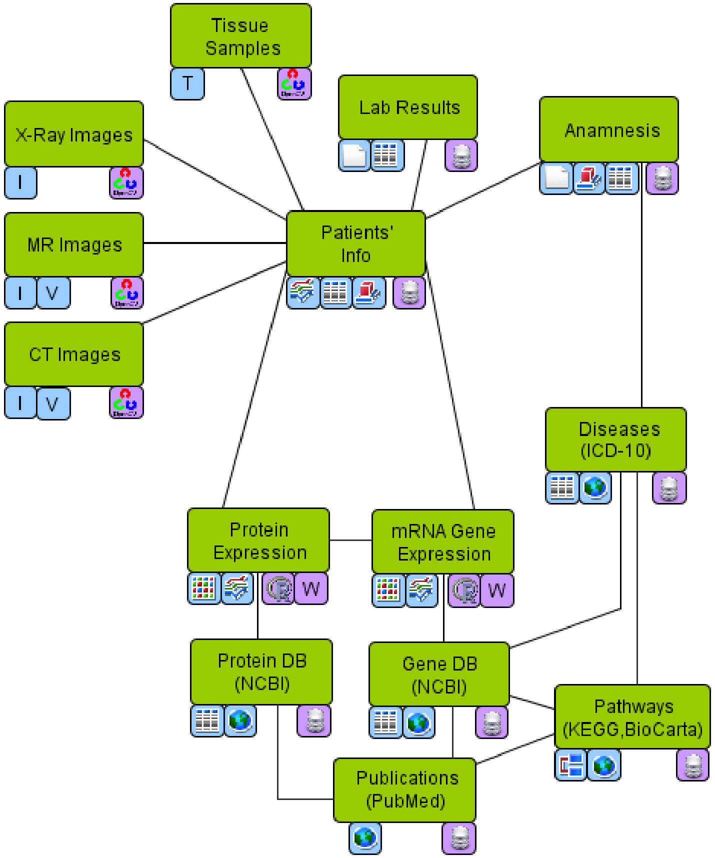


Information Landscape

III. Orientation and Navigation in Heterogeneous Data

Spell out the situation

Collect standard tasks



- View tissue
- Inspect lab results
- View anamnesis
- Record treatment decision
- Determine patients

- Browse patients
- Select patients
- Discard patients
- Inspect images
- Compare images
- Segment tumor
- Search disease
- View disease description
- Cluster expression data
- Inspect expression data
- Filter expression data
- Find gene
- View gene information
- Explore related pathways

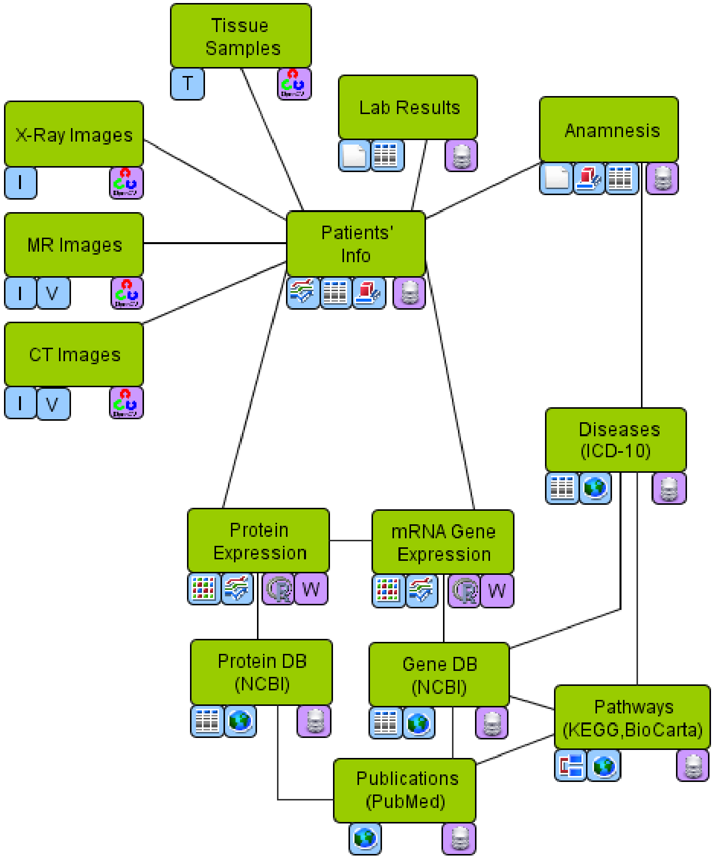
- Find pathway
- Investigate related articles

III. Orientation and Navigation in Heterogeneous Data

Spell out the situation

Collect standard tasks

Strip away their domain specificity…



View tissue	Inspect
Inspect lab results	Inspect
View anamnesis	Inspect
Record treatment decision	Record
Determine patients	Filter
	P. Clustering
Browse patients	Inspect
Select patients	Filter
Discard patients	Filter
Inspect images	Inspect
Compare images	Similarity A.
Segment tumor	Segment
Search disease	Query
View disease description	Inspect
Cluster expression data	H. Clustering
Inspect expression data	Inspect
Filter expression data	Filter
Find gene	Query
View gene information	Inspect
Explore related pathways	Filter
	Inspect
Find pathway	Query
Investigate related articles	Filter
	Inspect

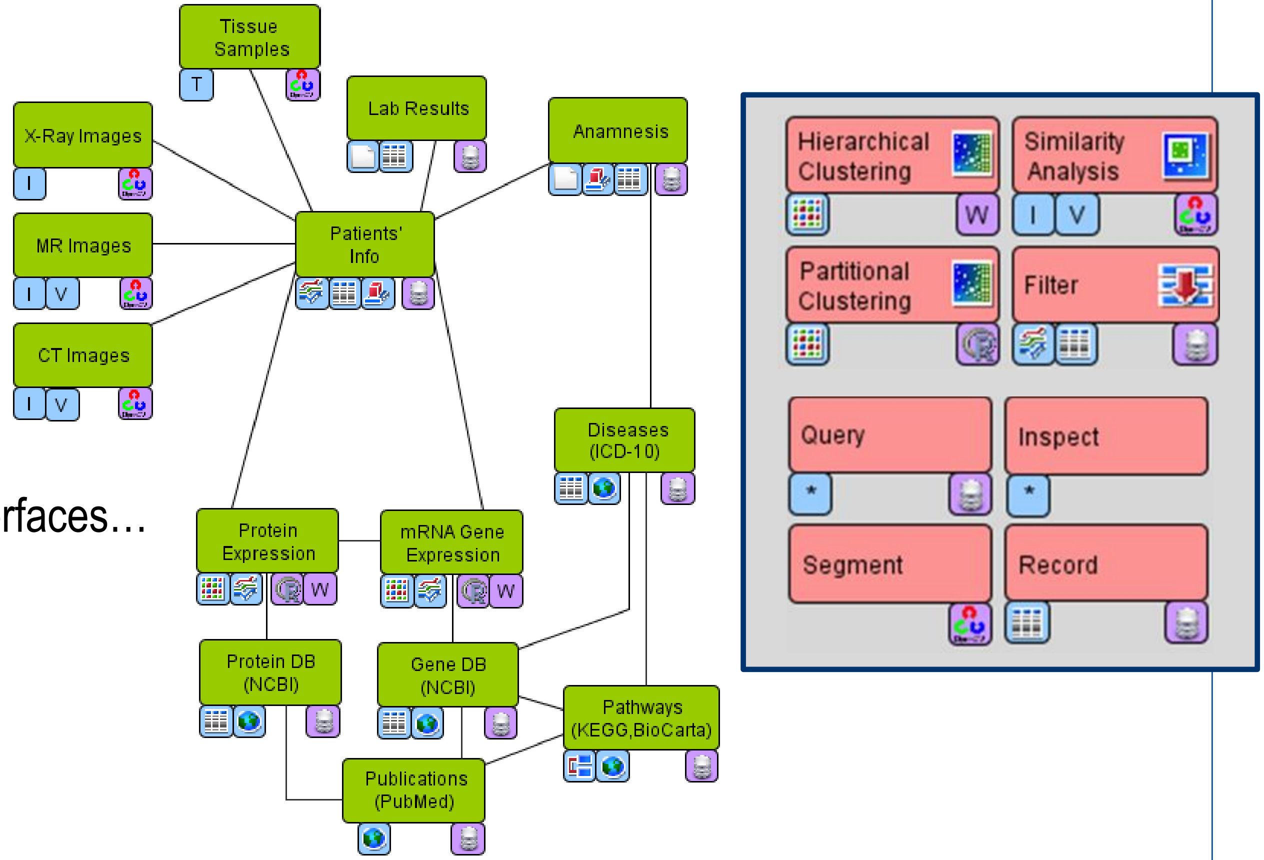
III. Orientation and Navigation in Heterogeneous Data

Spell out the situation

Collect standard tasks

Strip away their
domain specificity…

Link them to appropriate
visual and analytical interfaces…



III. Orientation and Navigation in Heterogeneous Data

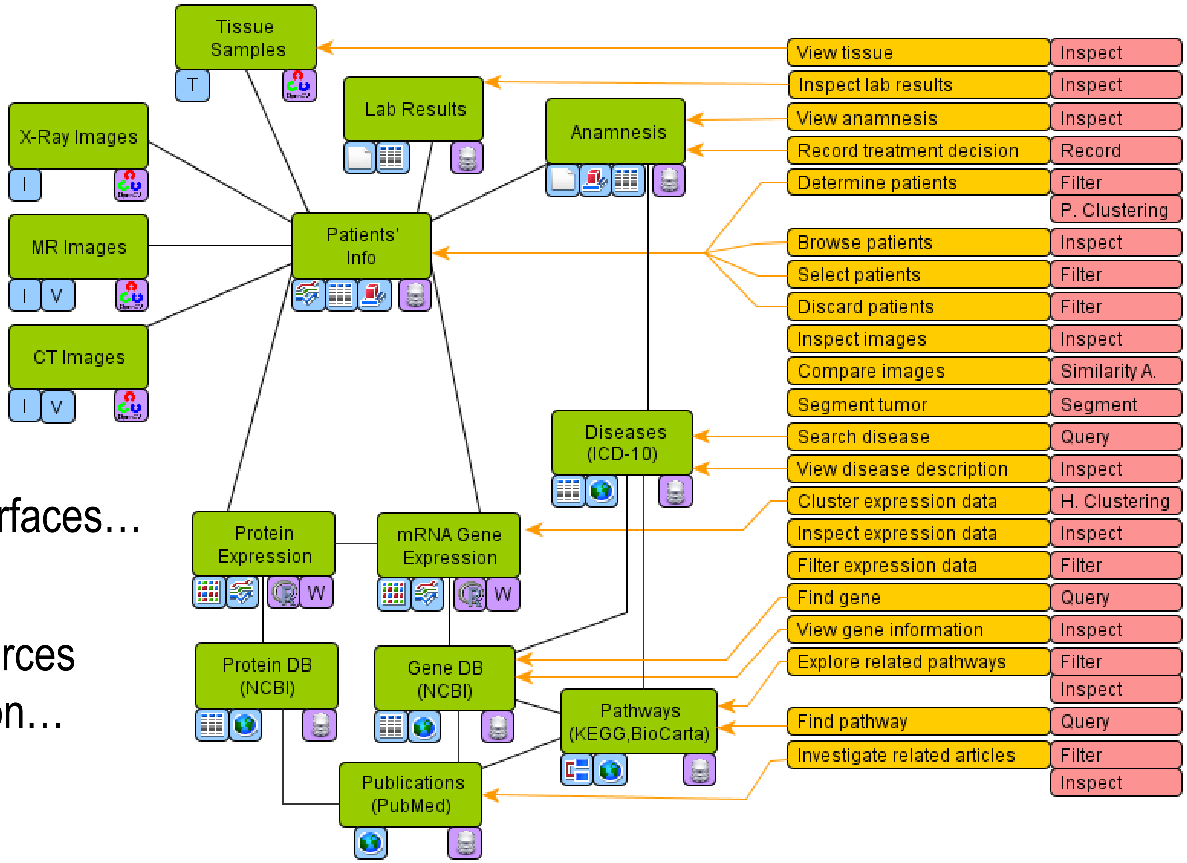
Spell out the situation

Collect standard tasks

Strip away their domain specificity…

Link them to appropriate visual and analytical interfaces…

Link them to the data sources they can be carried out on…



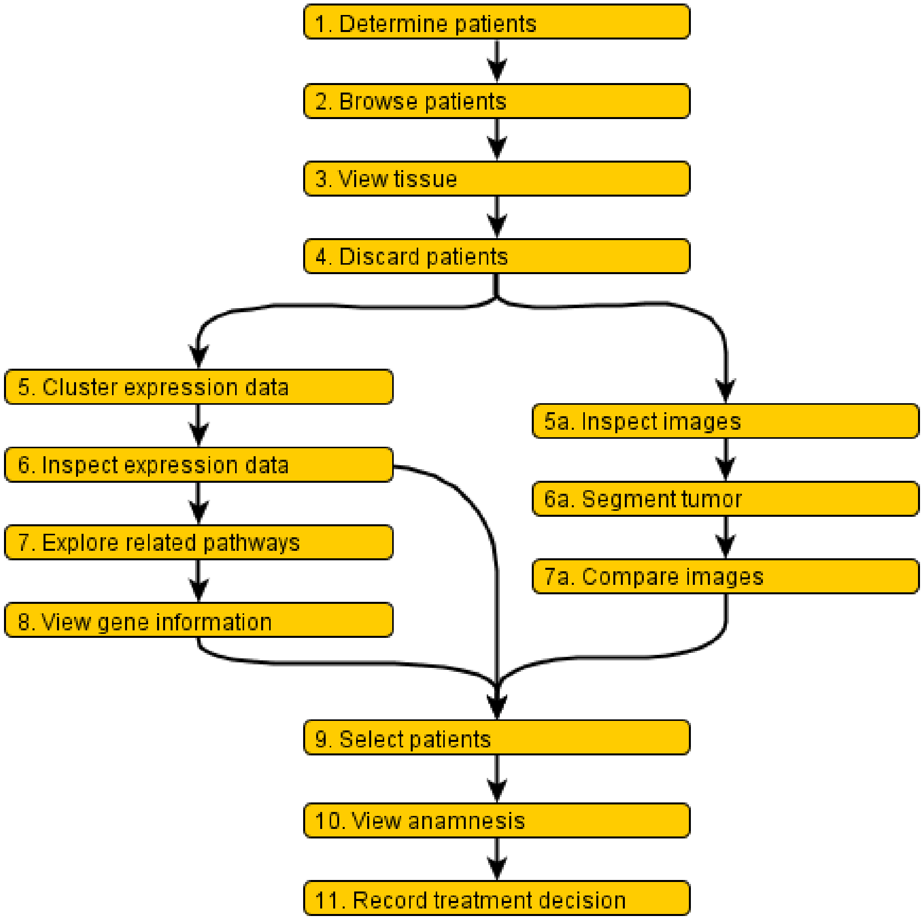
View tissue	Inspect
Inspect lab results	Inspect
View anamnesis	Inspect
Record treatment decision	Record
Determine patients	Filter
	P. Clustering
Browse patients	Inspect
Select patients	Filter
Discard patients	Filter
Inspect images	Inspect
Compare images	Similarity A.
Segment tumor	Segment
Search disease	Query
View disease description	Inspect
Cluster expression data	H. Clustering
Inspect expression data	Inspect
Filter expression data	Filter
Find gene	Query
View gene information	Inspect
Explore related pathways	Filter
	Inspect
Find pathway	Query
Investigate related articles	Filter
	Inspect

III. Orientation and Navigation in Heterogeneous Data

Model the Work Flow

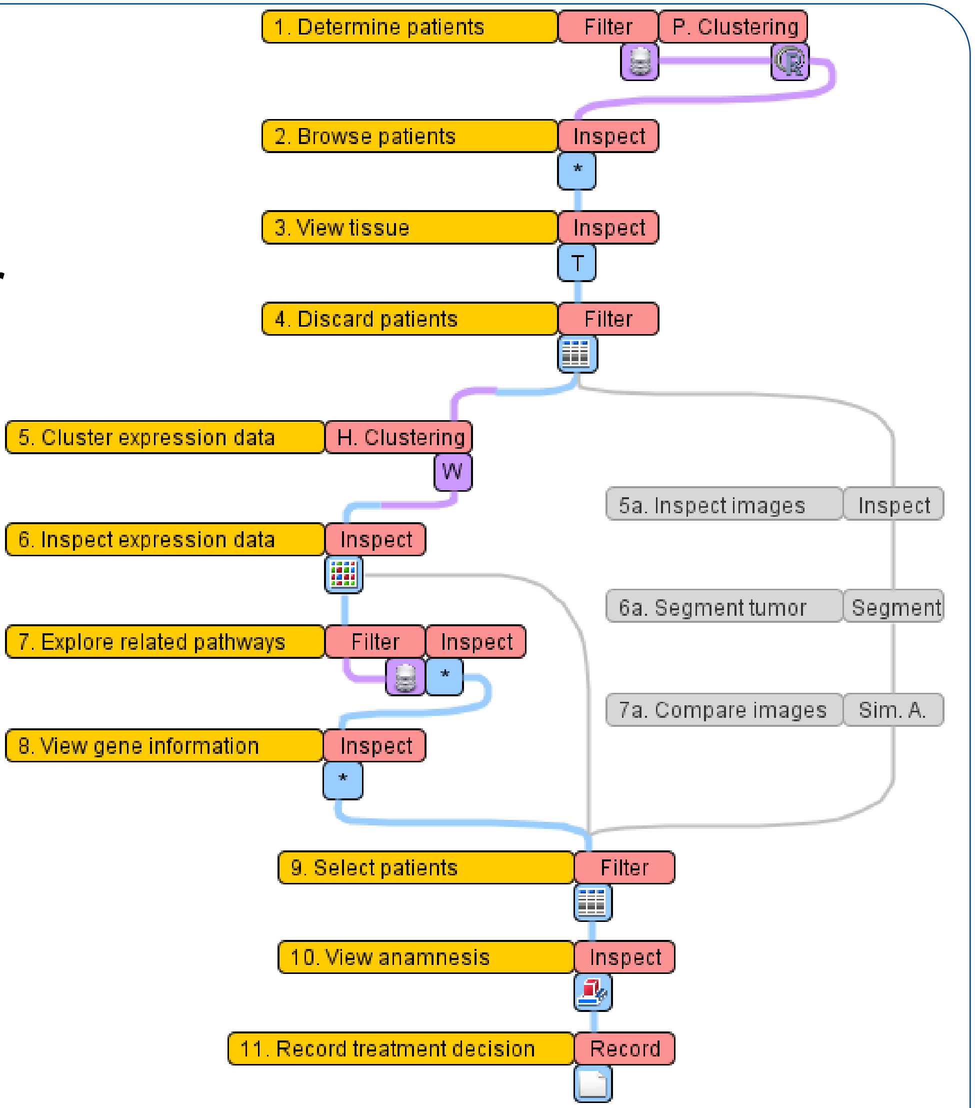
- Use the collected standard tasks as building blocks for the work flow model
- Model alternative paths by branching out the work flow
- Use pre- and post-conditions to define the objective of a task
- …

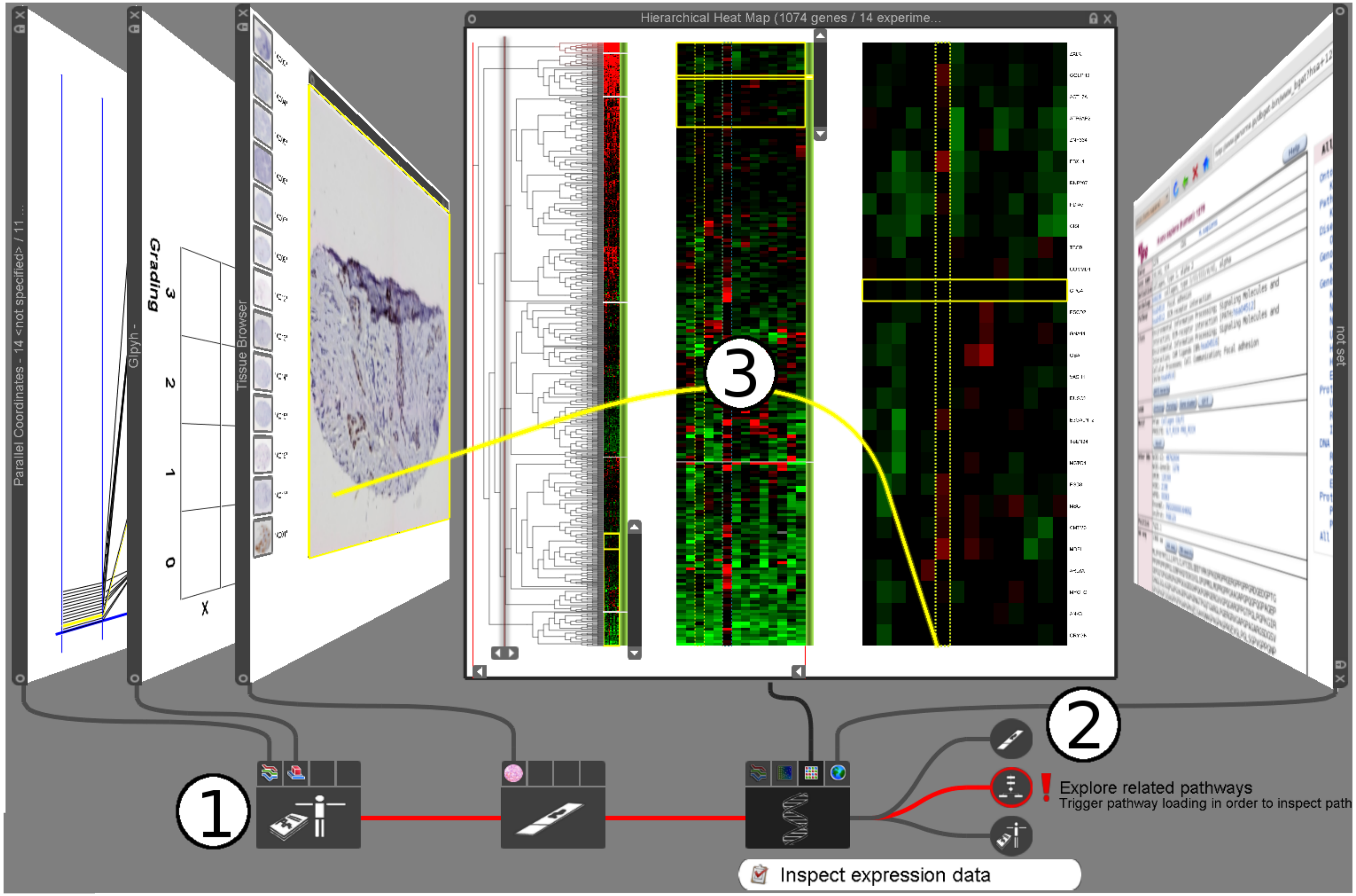
Note: The work flow is independent of the modeled setup!


1. Determine patients
2. Browse patients
3. View tissue
4. Discard patients
5. Cluster expression data
6. Inspect expression data
7. Explore related pathways
8. View gene information
5a. Inspect images
6a. Segment tumor
7a. Compare images
9. Select patients
10. View anamnesis
11. Record treatment decision

Bring Work Flow and Setup together

- Determine exactly which visual or analytical interface to use for each operator of each step's task
- Missing data sources and missing operators lead to a **pruning** of the workflow
- If the pruning doesn't leave a path from start to finish, the goal cannot be achieved, but the smallest gap to close can be determined…




Parallel Coordinates - 14 <not specified> / 11 ...
Glyph -
Grading
3
2
1
0
X
Tissue Browser
Hierarchical Heat Map (1074 genes / 14 experime...
not set
1
2
3
Explore related pathways
Trigger pathway loading in order to inspect path
Inspect expression data

III. Orientation and Navigation in Heterogeneous Data

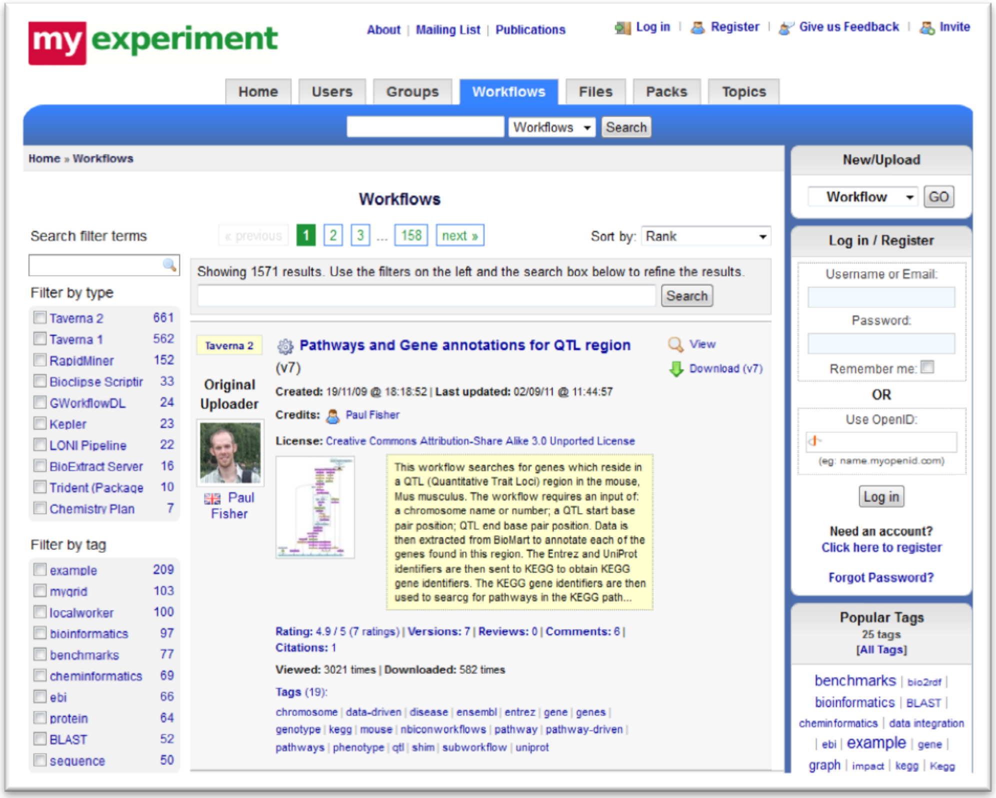
A few Words about the Cost of Modeling

- Mostly suitable for highly repetitive tasks in which each step is of outmost importance (airplane checks, biomedical procedures,…)

- Cost is not as high as imagined

1) hospitals often already have a data model for their IT infrastructure

2) workflows can be crowdsourced (unless they are proprietary)



Workflows

Search filter terms

« previous 1 2 3 ... 158 next »

Sort by: Rank

Showing 1571 results. Use the filters on the left and the search box below to refine the results.

Search

Filter by type

□ Taverna 2	661
□ Taverna 1	562
□ RapidMiner	152
□ Bioclipse Scriptir	33
□ GWorkflowDL	24
□ Kepler	23
□ LONI Pipeline	22
□ BioExtract Server	16
□ Trident (Package	10
□ Chemistry Plan	7

Filter by tag

□ example	209
□ mygrid	103
□ localworker	100
□ bioinformatics	97
□ benchmarks	77
□ cheminformatics	69
□ ebi	66
□ protein	64
□ BLAST	52
□ sequence	50

Taverna 2

Pathways and Gene annotations for QTL region (v7)

View

Download (v7)

Original Uploader

Paul Fisher

Created: 19/11/09 @ 18:18:52 | **Last updated:** 02/09/11 @ 11:44:57

Credits: Paul Fisher

License: Creative Commons Attribution-Share Alike 3.0 Unported License

This workflow searches for genes which reside in a QTL (Quantitative Trait Loci) region in the mouse, Mus musculus. The workflow requires an input of: a chromosome name or number; a QTL start base pair position; QTL end base pair position. Data is then extracted from BioMart to annotate each of the genes found in this region. The Entrez and UniProt identifiers are then sent to KEGG to obtain KEGG gene identifiers. The KEGG gene identifiers are then used to searcg for pathways in the KEGG path...

Rating: 4.9 / 5 (7 ratings) | **Versions:** 7 | **Reviews:** 0 | **Comments:** 6 | **Citations:** 1

Viewed: 3021 times | **Downloaded:** 582 times

Tags (19):

chromosome | data-driven | disease | ensembl | entrez | gene | genes | genotype | kegg | mouse | nbiconworkflows | pathway | pathway-driven | pathways | phenotype | qtl | shim | subworkflow | uniprot

New/Upload

Workflow GO

Log in / Register

Username or Email:

Password:

Remember me: □

OR

Use OpenID:

(eg: name.myopenid.com)

Log in

Need an account?
Click here to register

Forgot Password?

Popular Tags
25 tags
[All Tags]

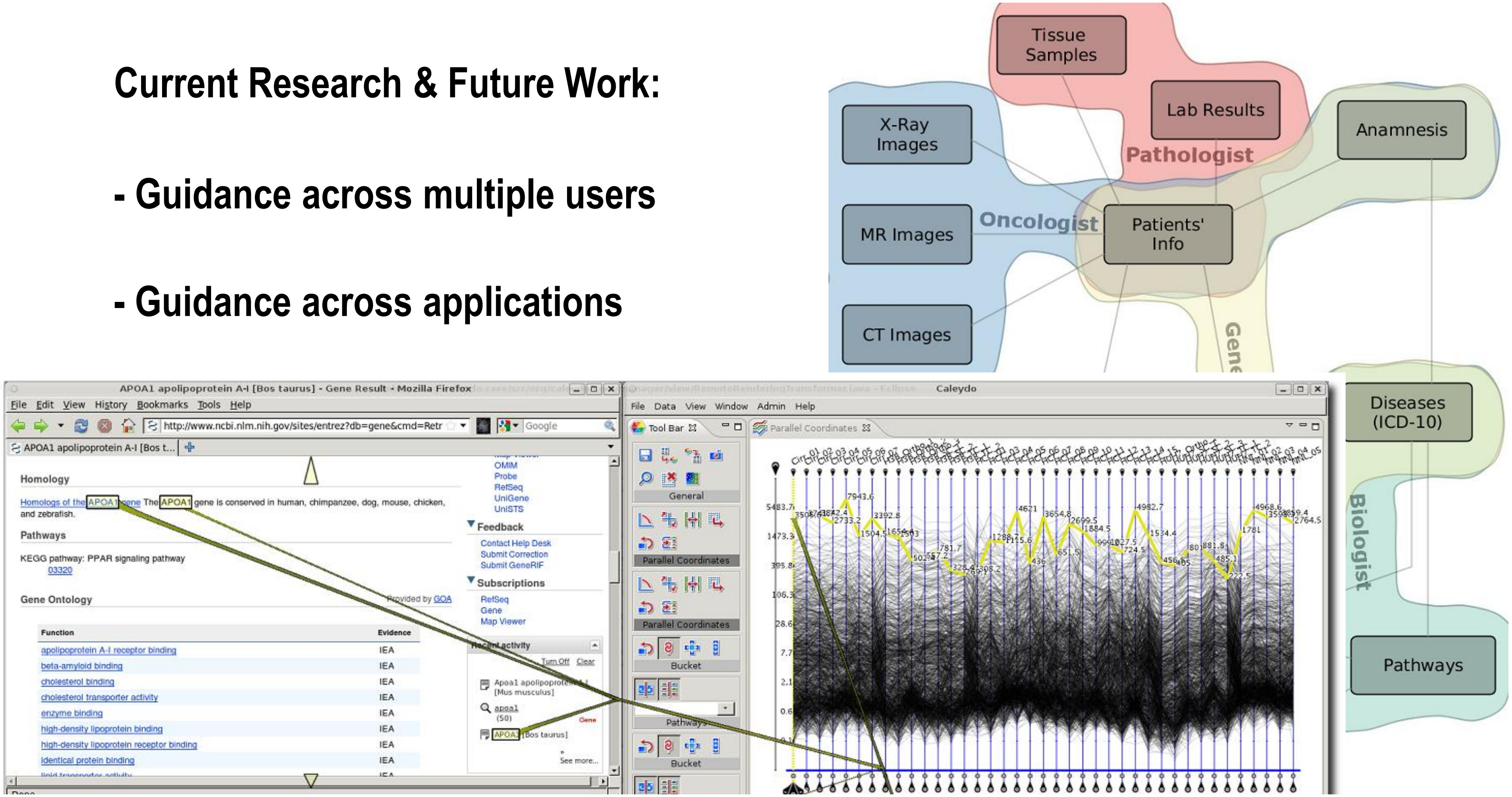
benchmarks | bio2rdf | bioinformatics | BLAST | cheminformatics | data integration | ebi | example | gene | graph | impact | kegg | Kegg

III. Orientation and Navigation in Heterogeneous Data

Current Research & Future Work:

- **Guidance across multiple users**

- **Guidance across applications**



APOA1 apolipoprotein A-I [Bos taurus] - Gene Result - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retr

APOA1 apolipoprotein A-I [Bos t...

Homology

Homologs of the APOA1 gene The APOA1 gene is conserved in human, chimpanzee, dog, mouse, chicken, and zebrafish.

Pathways

KEGG pathway: PPAR signaling pathway
03320

Gene Ontology

Provided by GOA

Function	Evidence
apolipoprotein A-I receptor binding	IEA
beta-amyloid binding	IEA
cholesterol binding	IEA
cholesterol transporter activity	IEA
enzyme binding	IEA
high-density lipoprotein binding	IEA
high-density lipoprotein receptor binding	IEA
identical protein binding	IEA
lipid transporter activity	IEA

OMIM
Probe
RefSeq
UniGene
UniSTS

Feedback

Contact Help Desk
Submit Correction
Submit GeneRIF

Subscriptions

RefSeq
Gene
Map Viewer

Recent activity

Turn Off Clear

Apoa1 apolipoprotein A-I [Mus musculus]

apoa1 (50) Gene

APOA1 [Bos taurus]

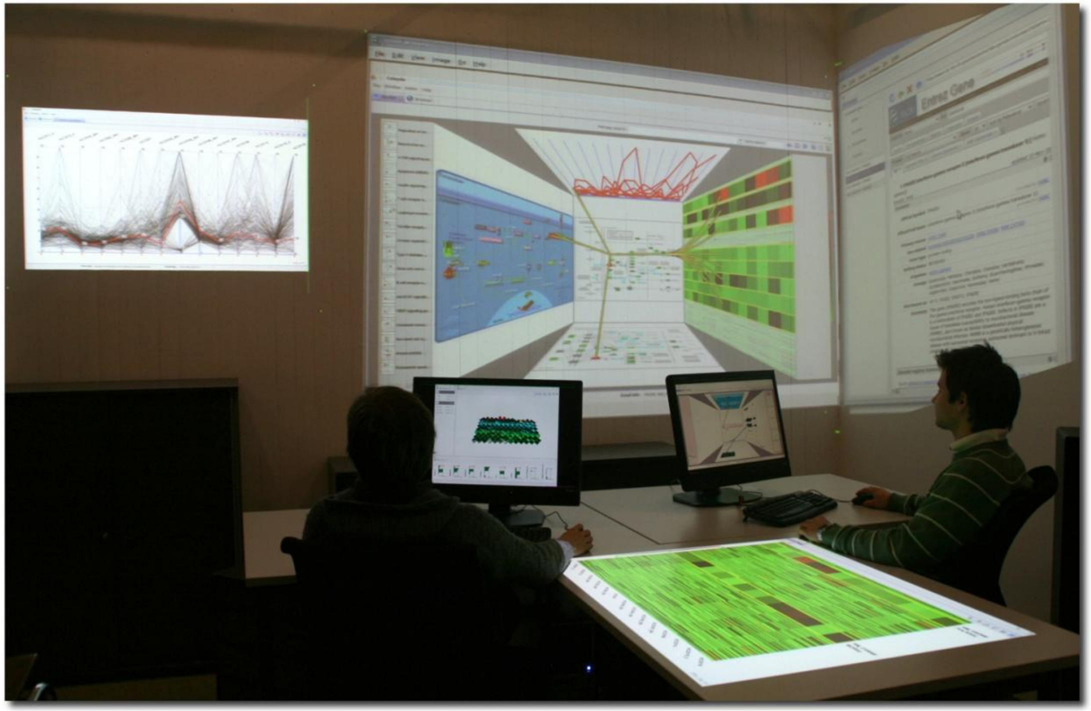
See more...

Done

Source: Marc Streit, PhD Thesis (Graz, 2011)

III. Orientation and Navigation in Heterogeneous Data

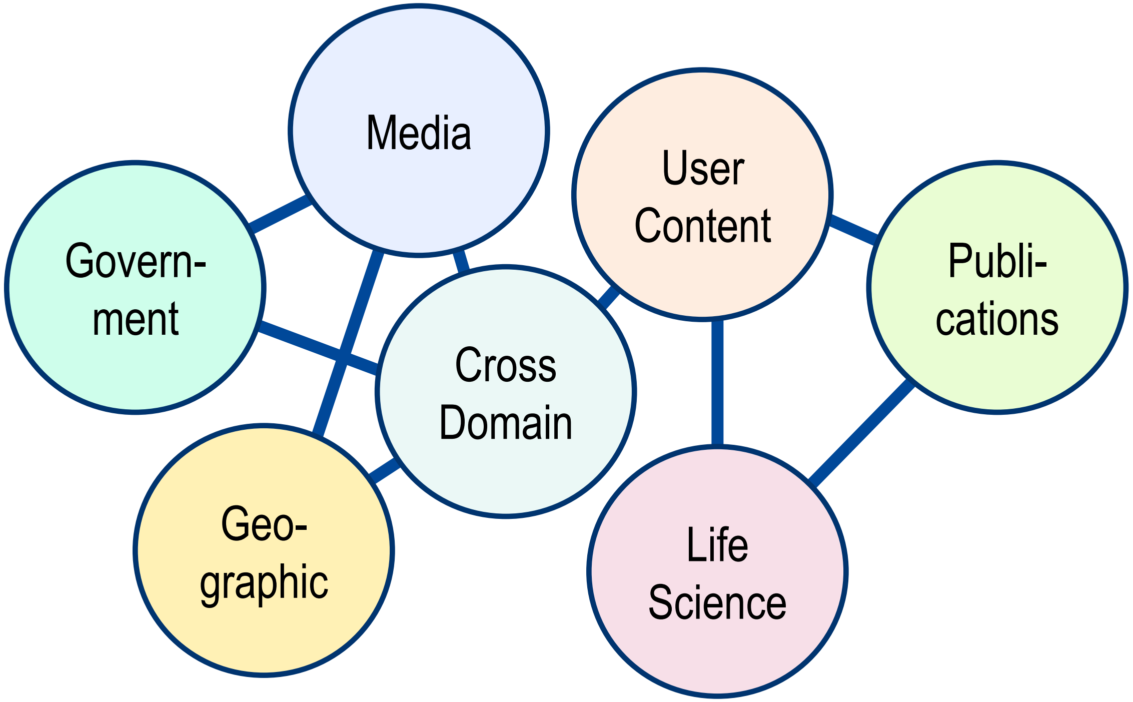
Current Research & Future Work: Guidance across multiple displays



Source: Marc Streit, PhD Thesis (Graz, 2011)

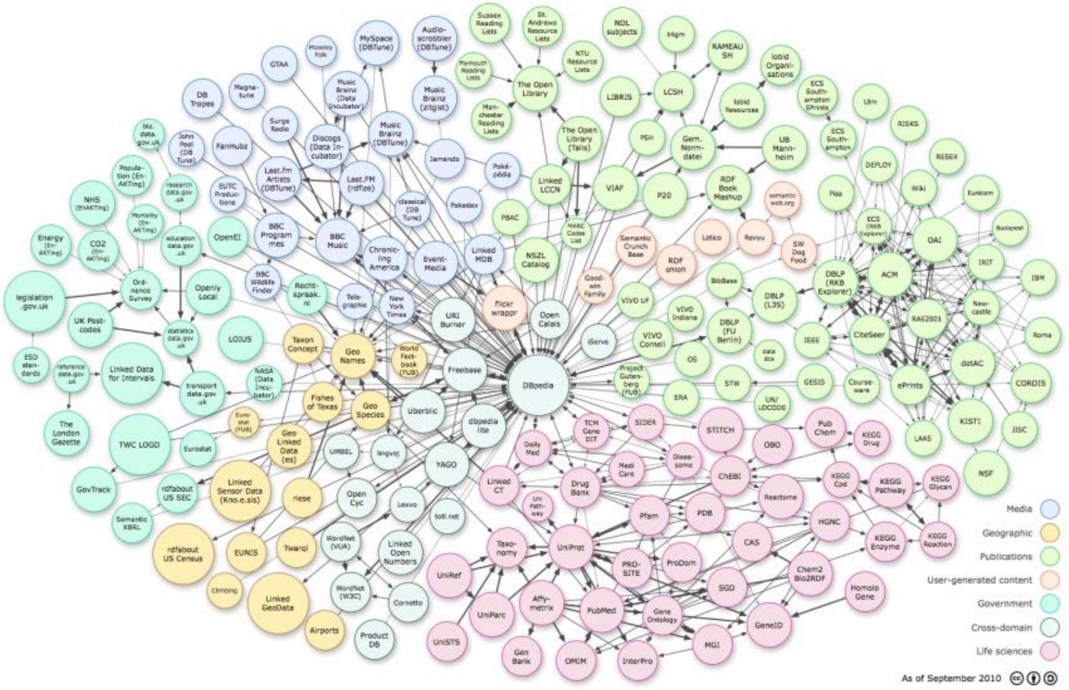
IV. Conclusion, Food for Thought

My vision for heterogeneous data: Google Maps for Information Landscapes
- to combine the meta view of the “data model” with multiform visualizations



IV. Conclusion, Food for Thought

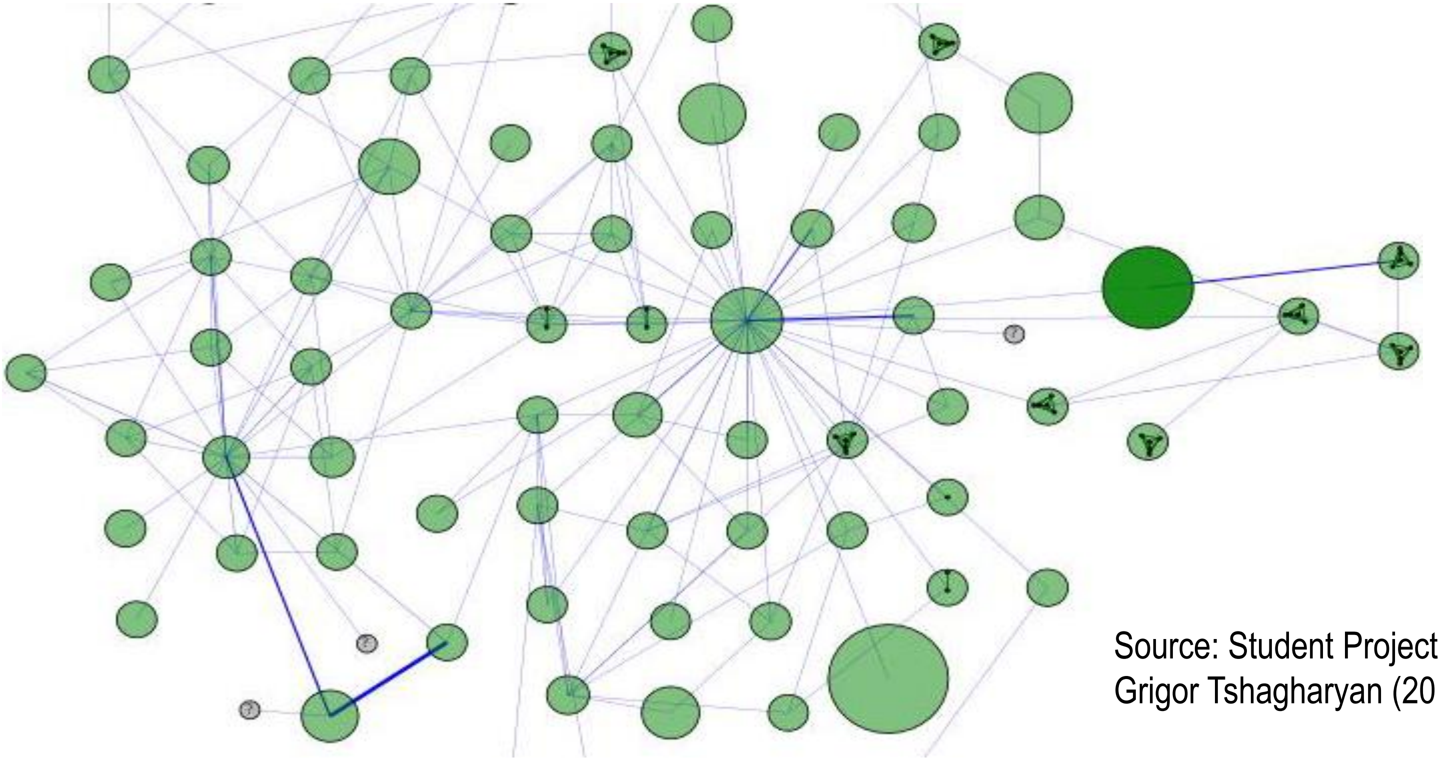
My vision for heterogeneous data: Google Maps for Information Landscapes
- to combine the meta view of the “data model” with multiform visualizations



Source: Richard Cyganiak, Anja Jentzsch

IV. Conclusion, Food for Thought

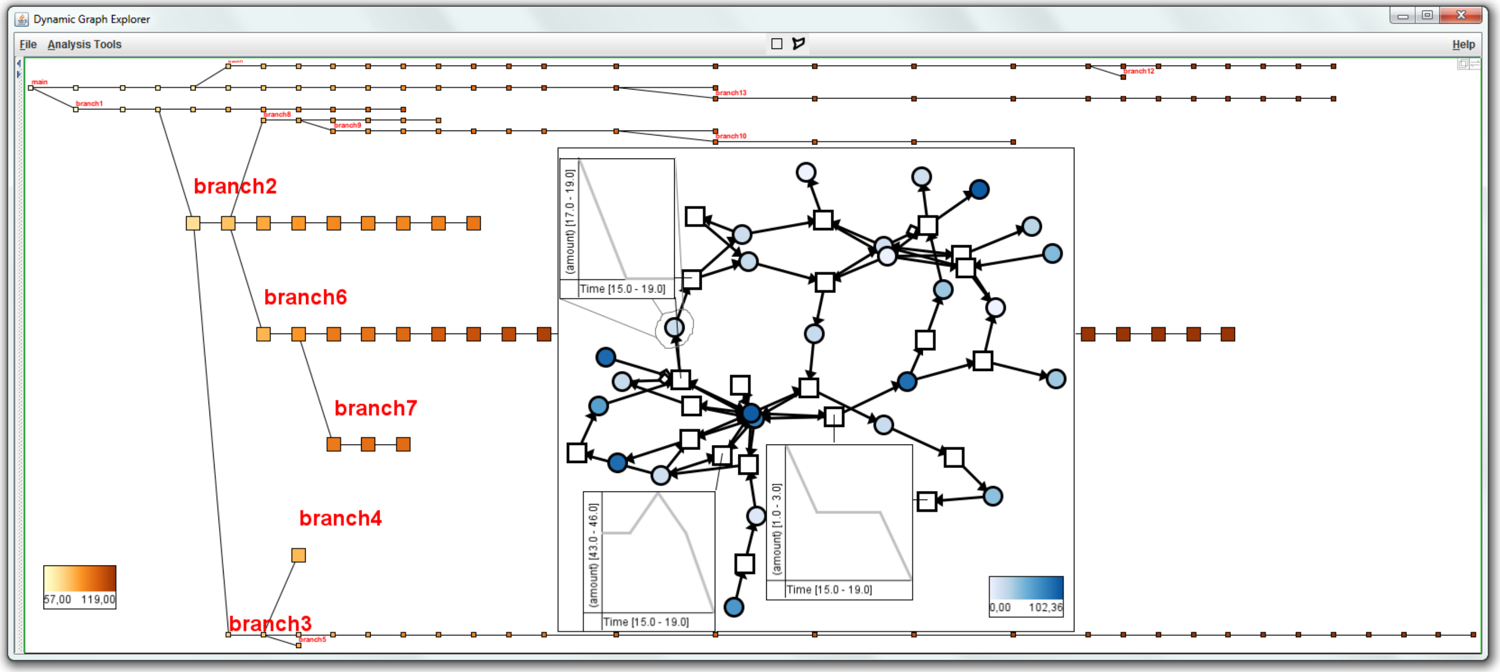
My vision for heterogeneous data: Google Maps for Information Landscapes
- to combine the meta view of the "data model" with multiform visualizations



Source: Student Project by
Grigor Tshagharyan (2011)

IV. Conclusion, Food for Thought

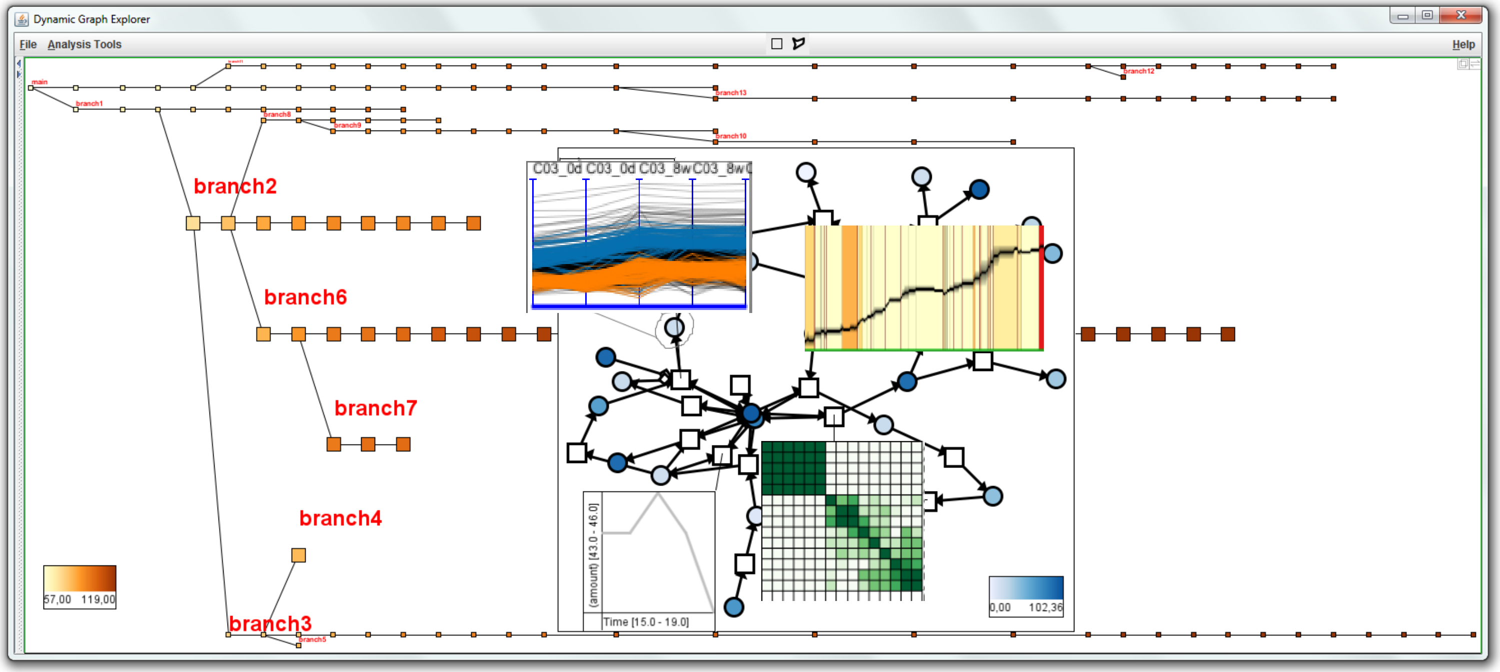
My vision for heterogeneous data: Google Maps for Information Landscapes
- to combine the meta view of the “data model” with multiform visualizations



IV. Conclusion, Food for Thought

My vision for heterogeneous data: Google Maps for Information Landscapes
- to combine the meta view of the "data model" with multiform visualizations

Publi-cations



Govern-ment

Acknowledgements

Deutsche Forschungsgemeinschaft DFG

TU Graz
Graz University of Technology

FFG

CALEYDO

Further Information

Dr.-Ing. Hans-Jörg Schulz

Web: http://hjschulz.net
eMail: contact@hjschulz.net