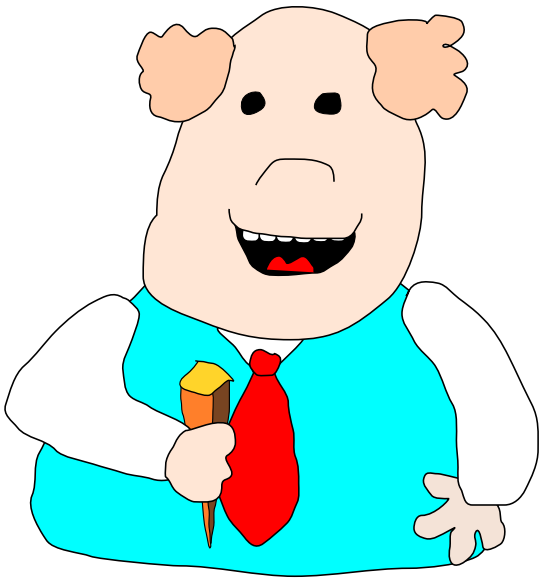
Software Process Recovery

Abram Hindle

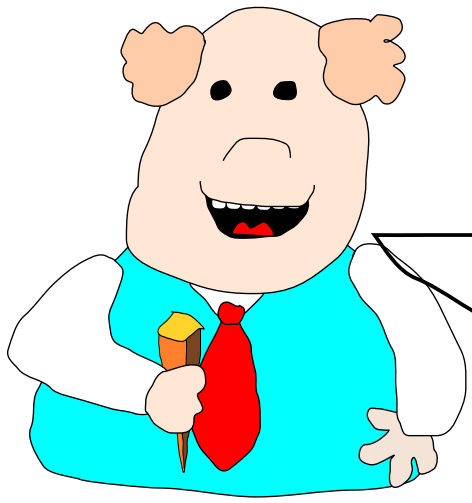
Software Architecture Group
David R. Cheriton School of Computer Science
University of Waterloo
Canada
http://swag.uwaterloo.ca/~ahindle/

ahindle@cs.uwaterloo.ca

what is going on in this project?



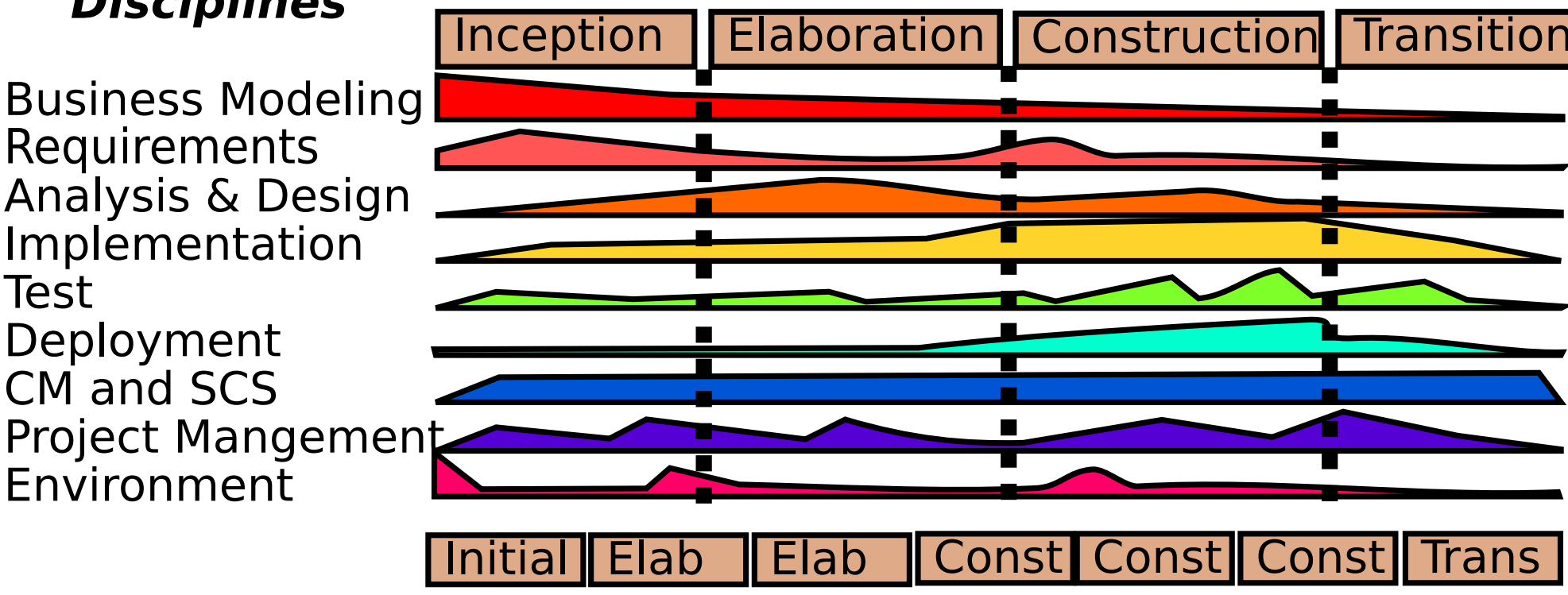
Software Project

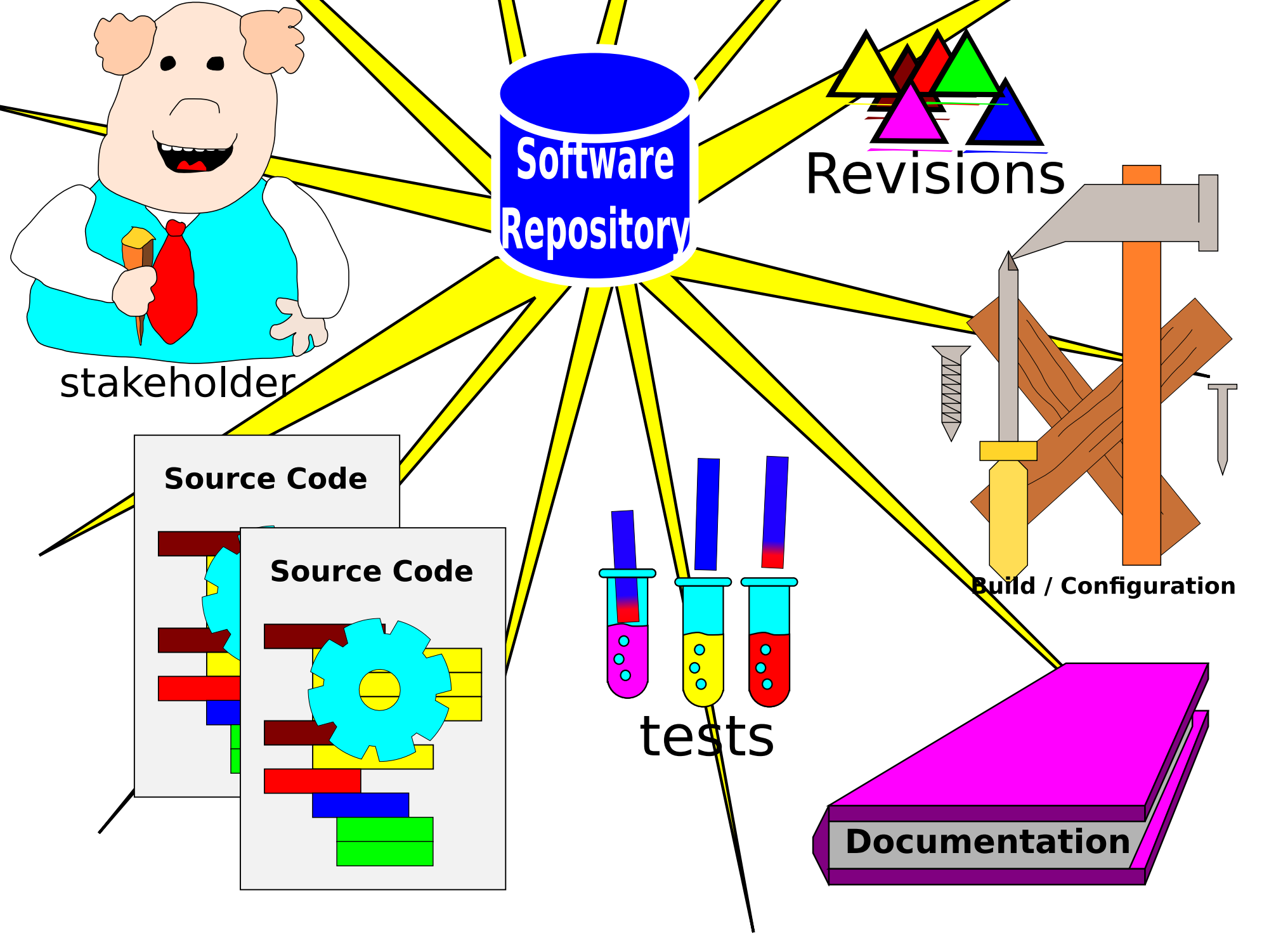


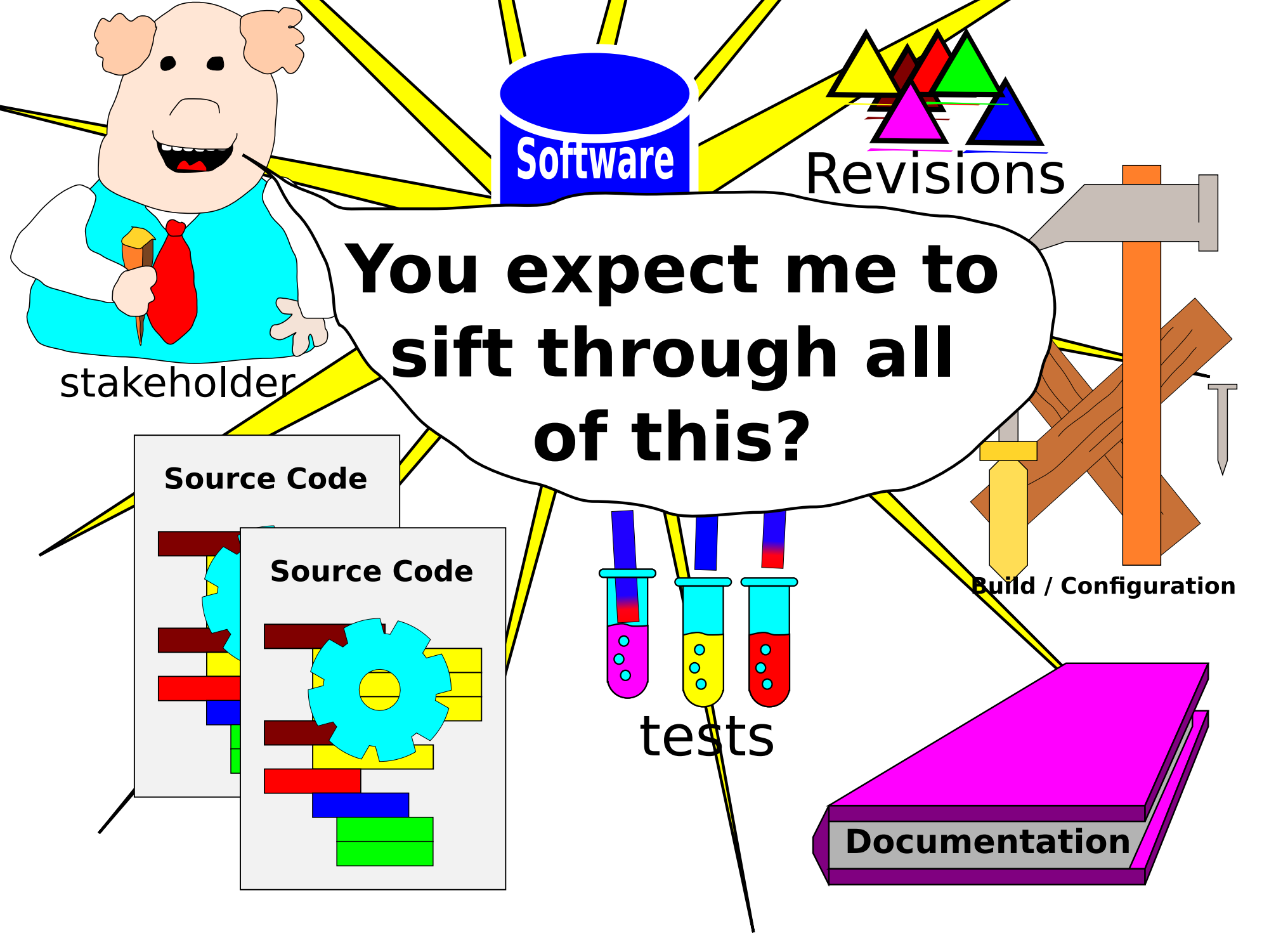
An overview of the project's processes and development would be nice!

Phases

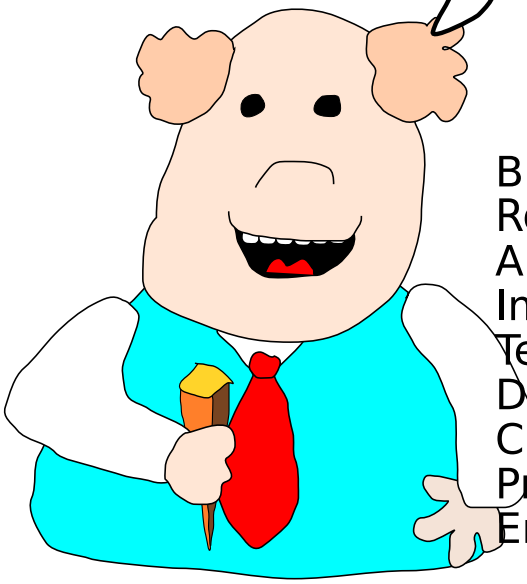
Disciplines

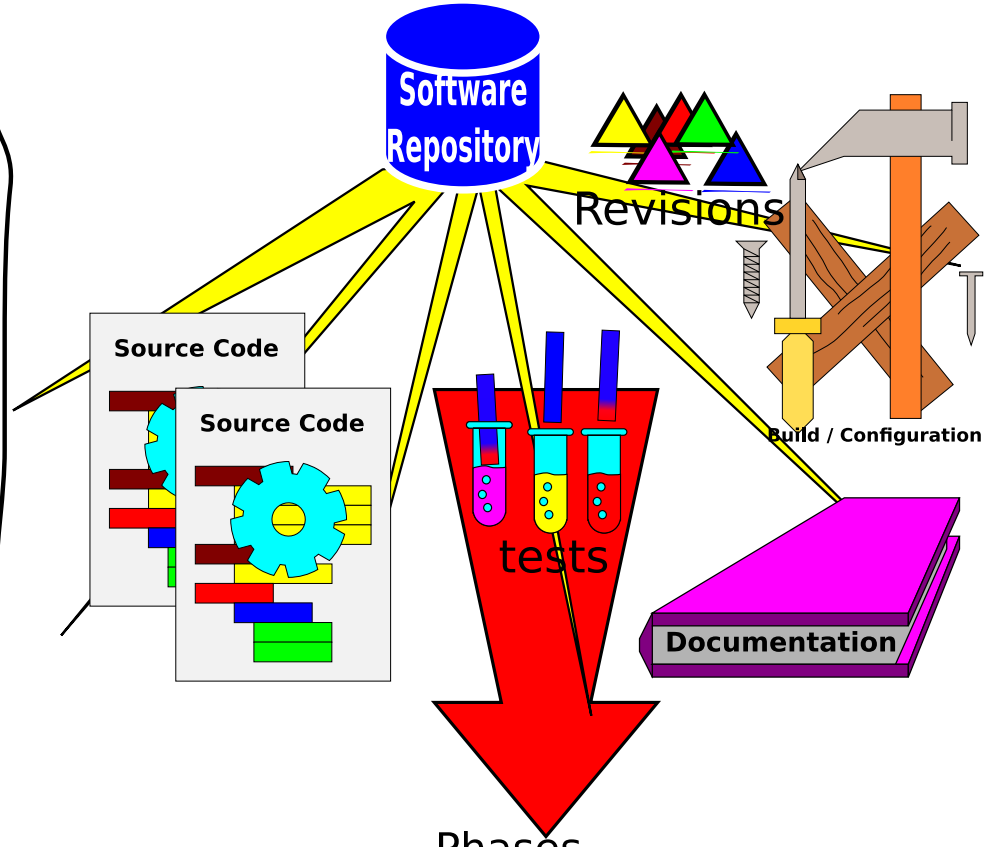



Software Repository
Revisions
stakeholder
Source Code
Source Code
tests
Build / Configuration
Documentation


Software
Revisions
You expect me to sift through all of this?
stakeholder
Source Code
Source Code
tests
Build / Configuration
Documentation

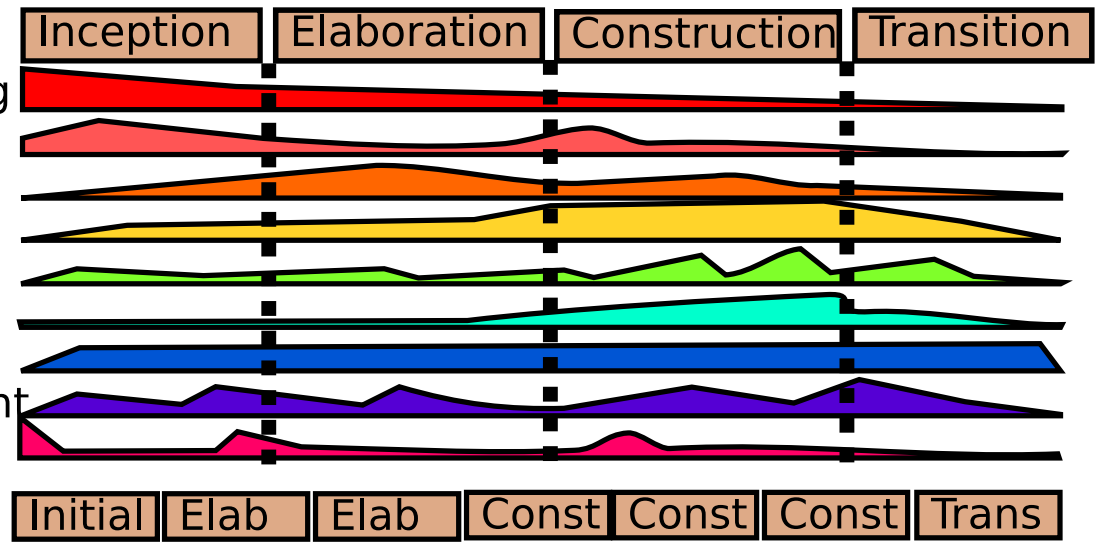
Can't we just summarize what is going on within this project?





Disciplines

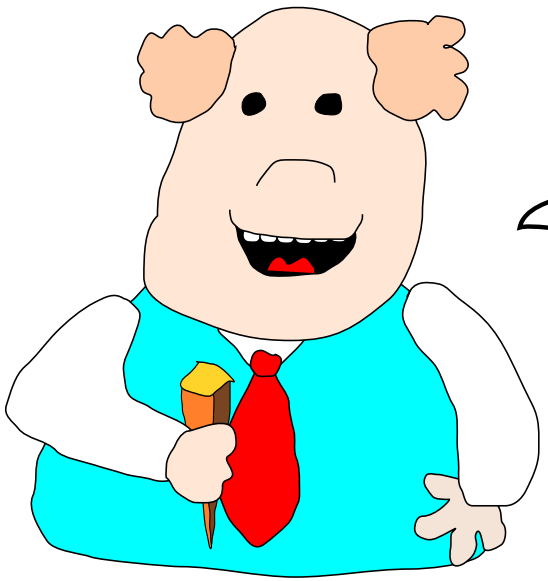
Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Mangement
Environment



Proposed Process

Recovered Process

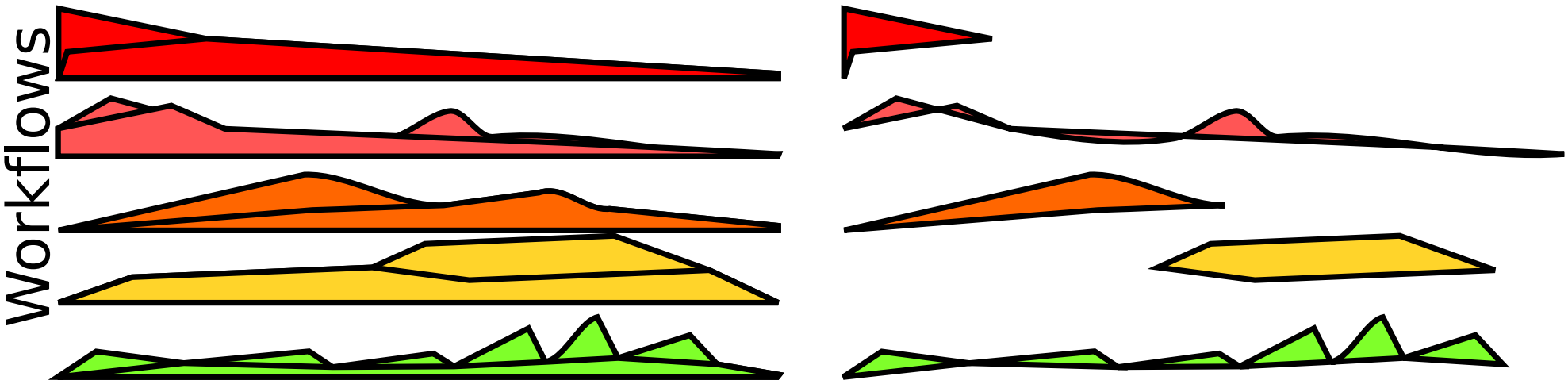
Workflows

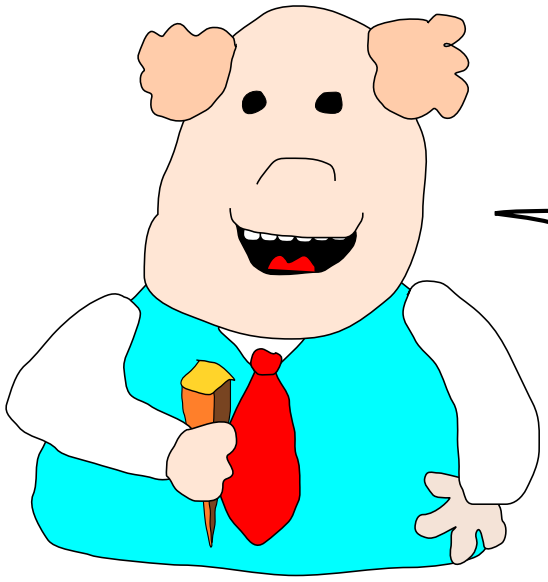


With an overview I can validate this process against what I expected!

Proposed and Recovered Process Overlayed

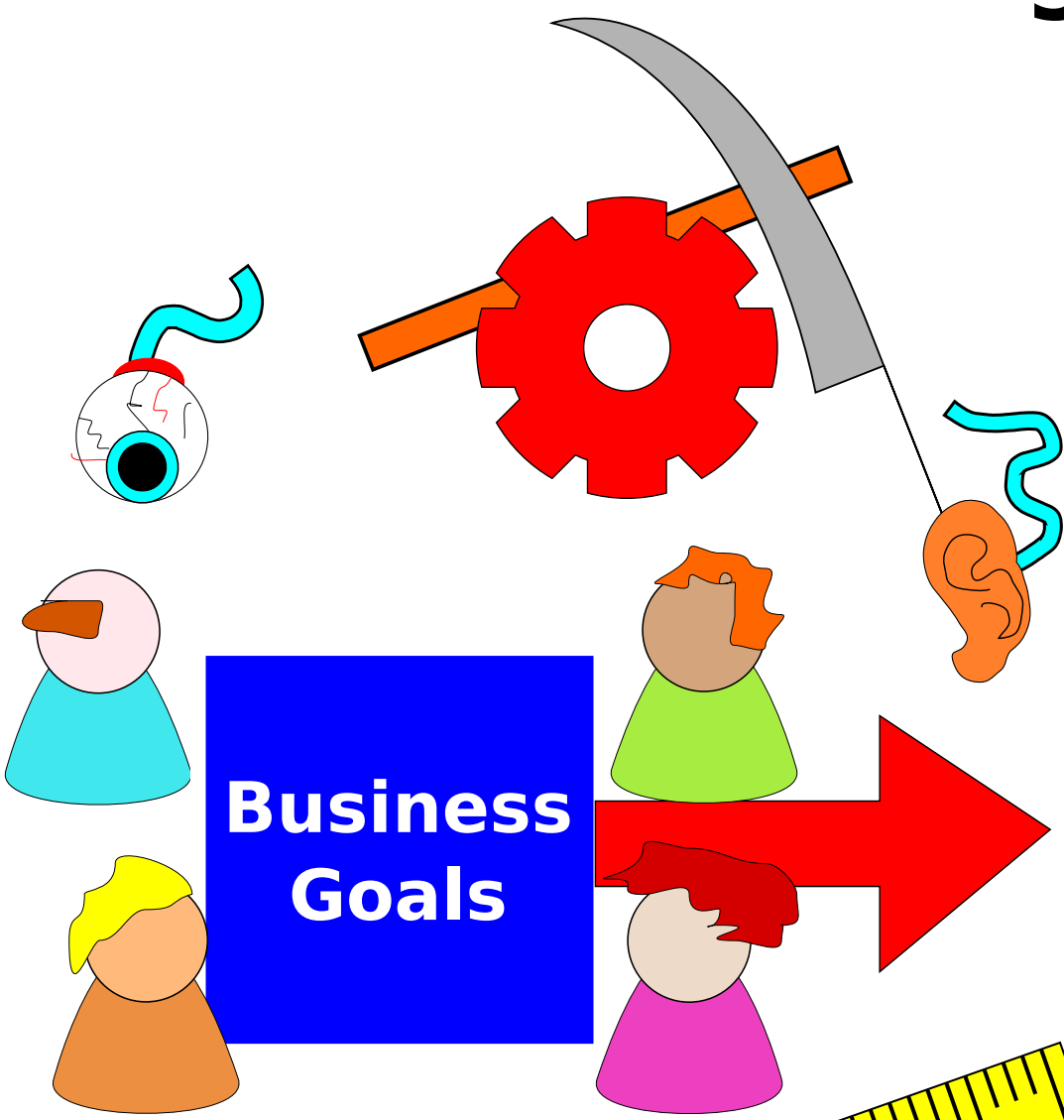
Differences between Proposed and Recovered

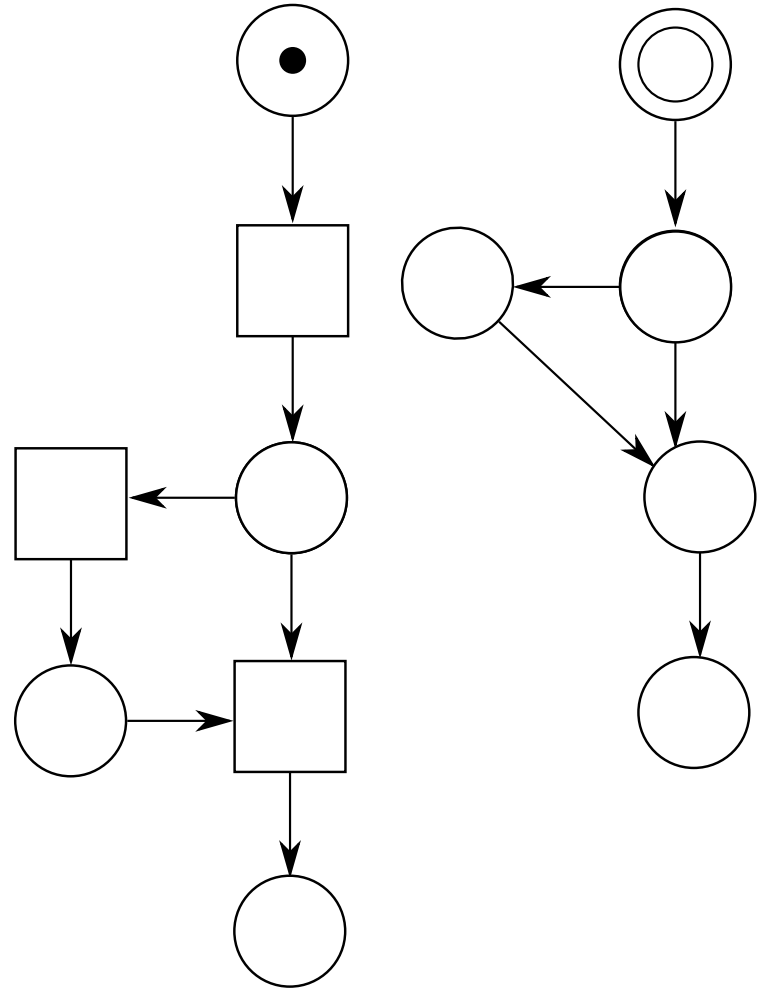




I can compare and contrast the observed process versus the expected process!

Process Mining

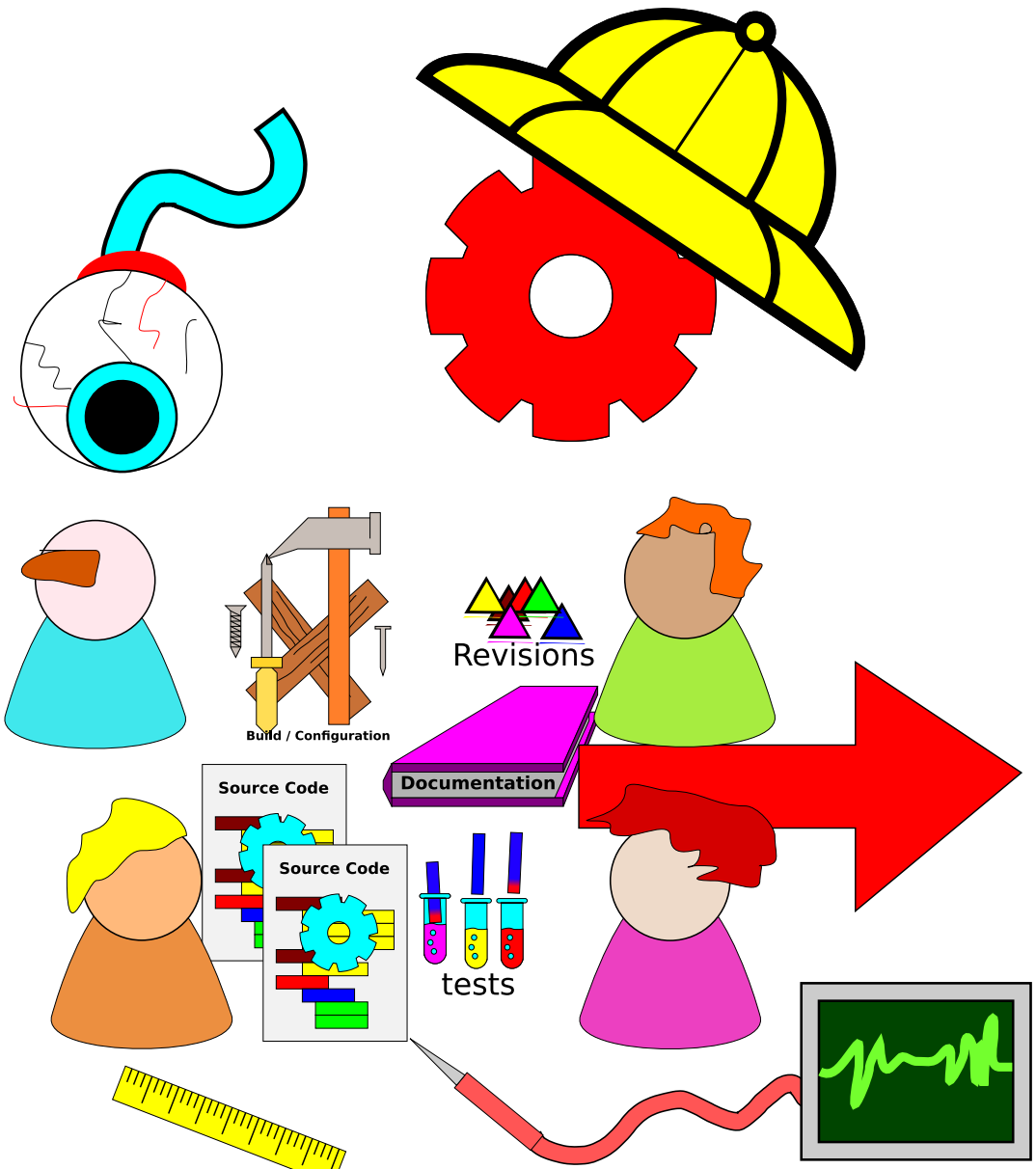




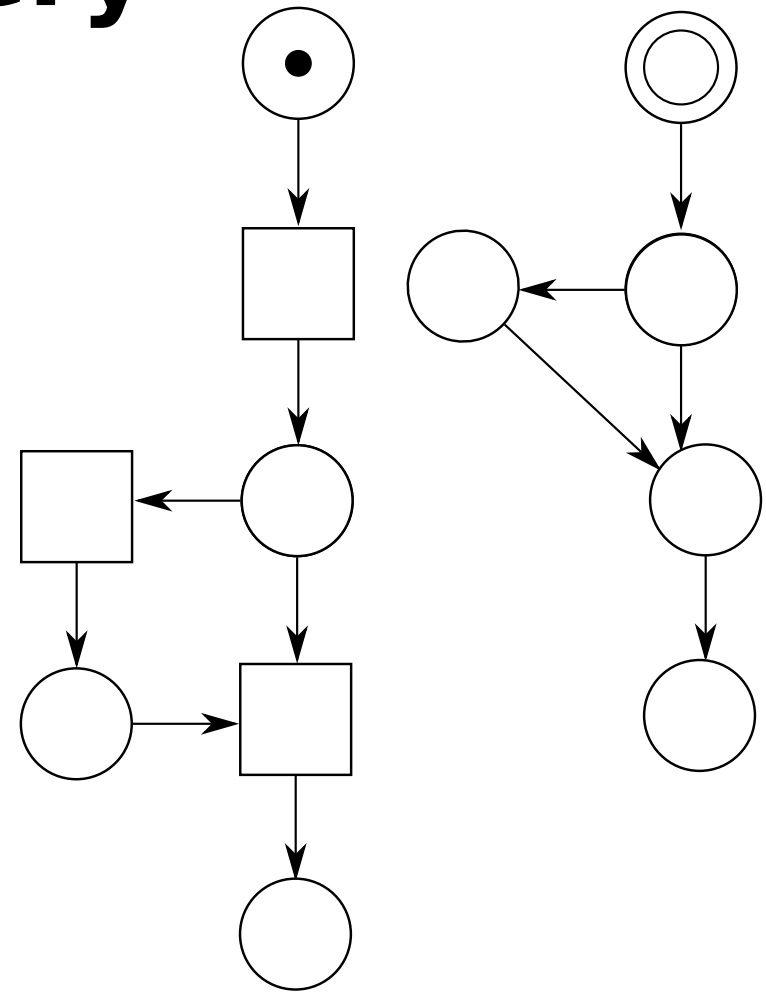
Petrinet FSM

[Aalst]

Process Discovery



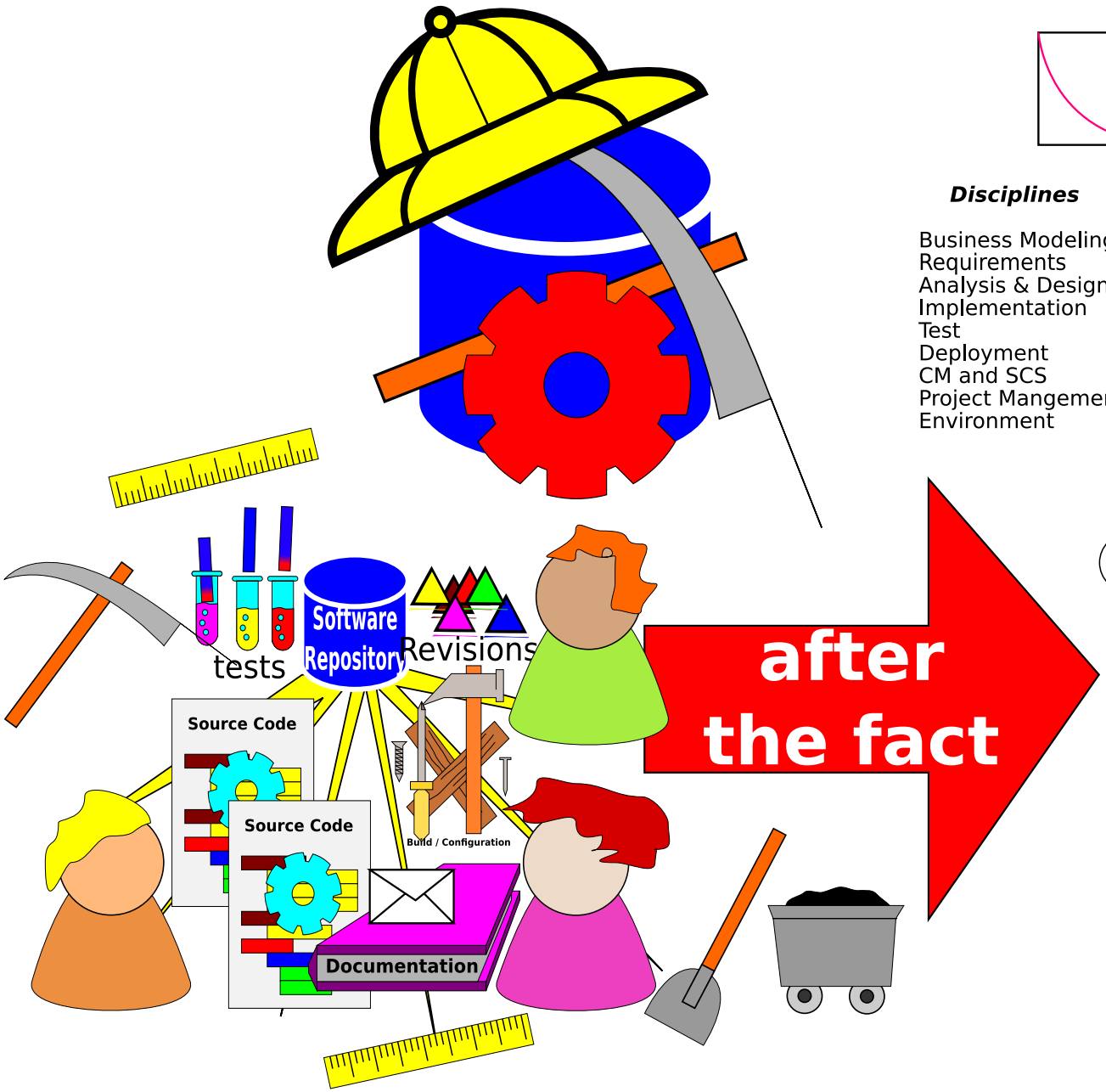
tooled process



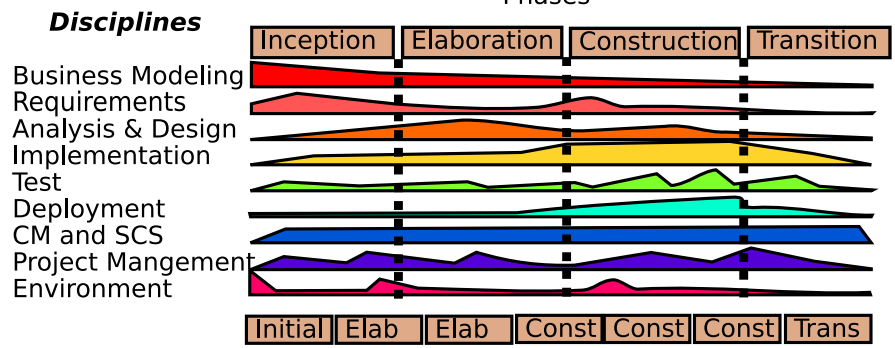
Petrinet FSM

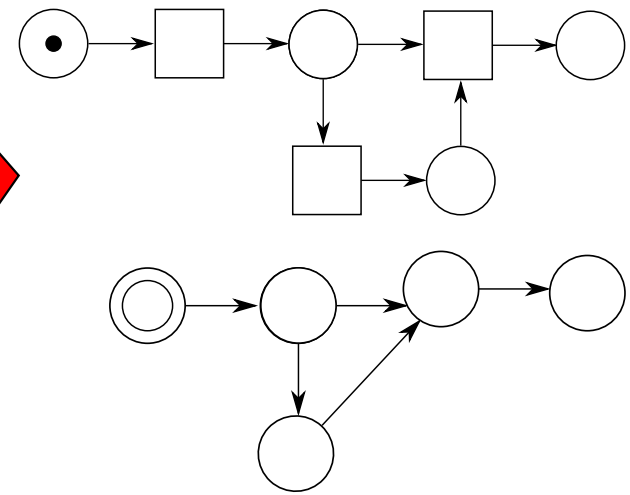
[Cook]

Process Recovery



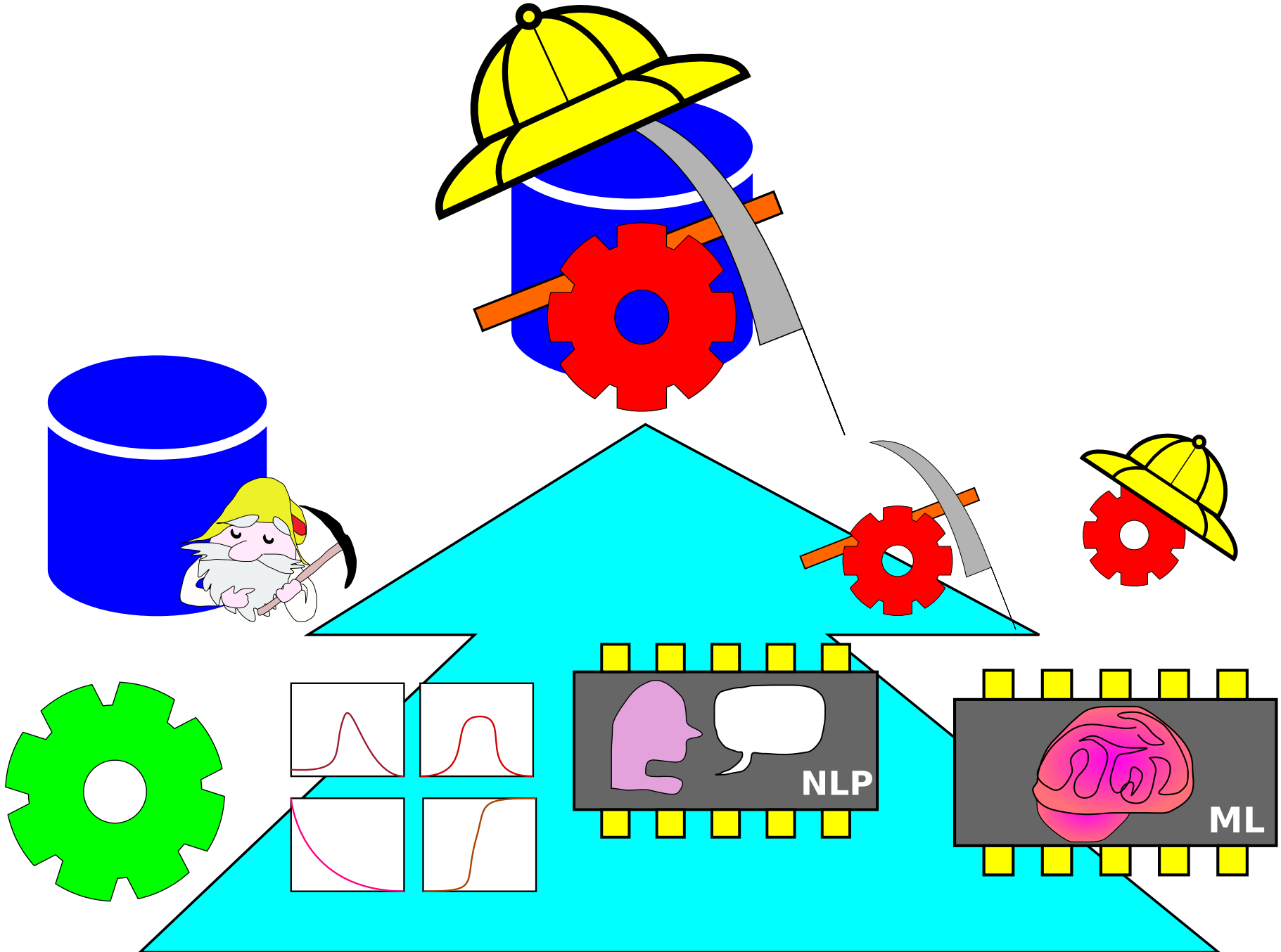
Phases



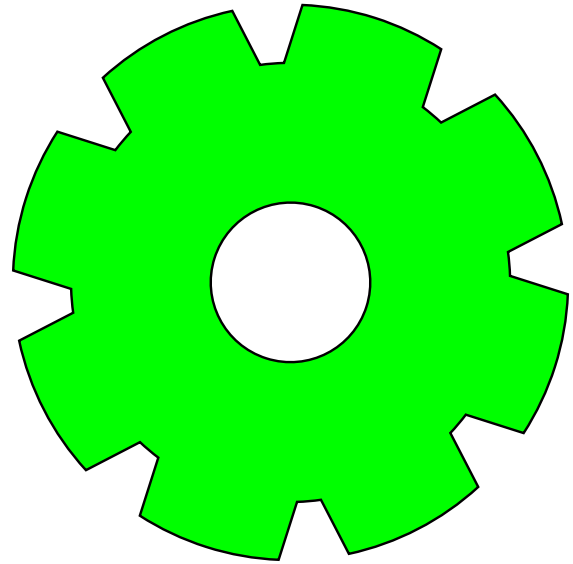


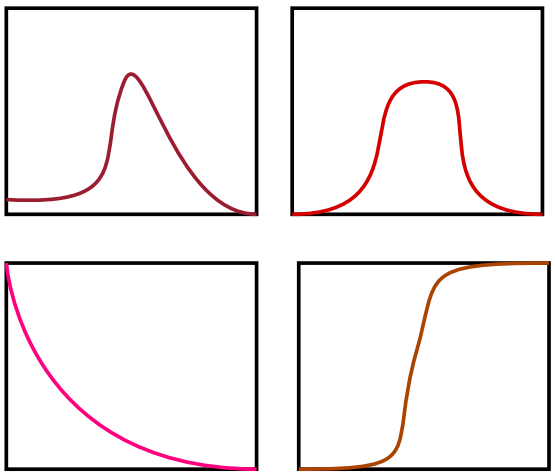
[Hindle]

Process Recovery

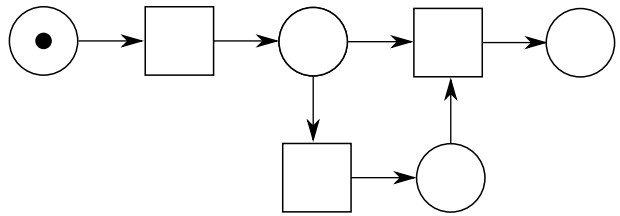


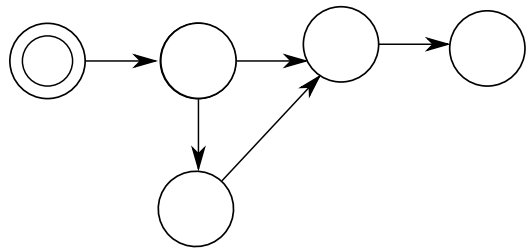
Process



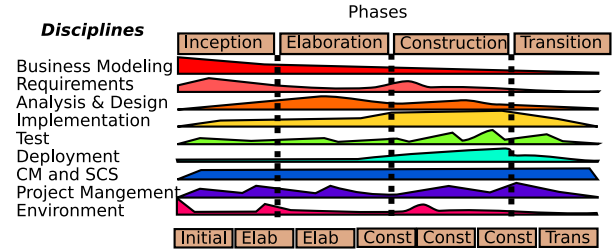


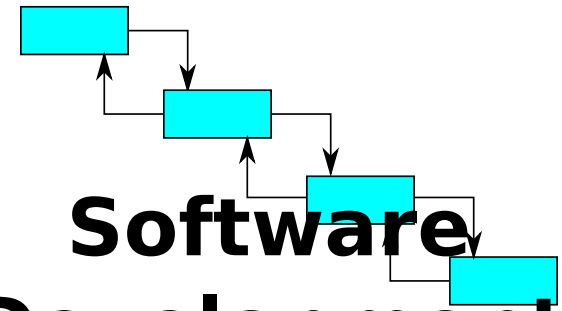
Stochastic





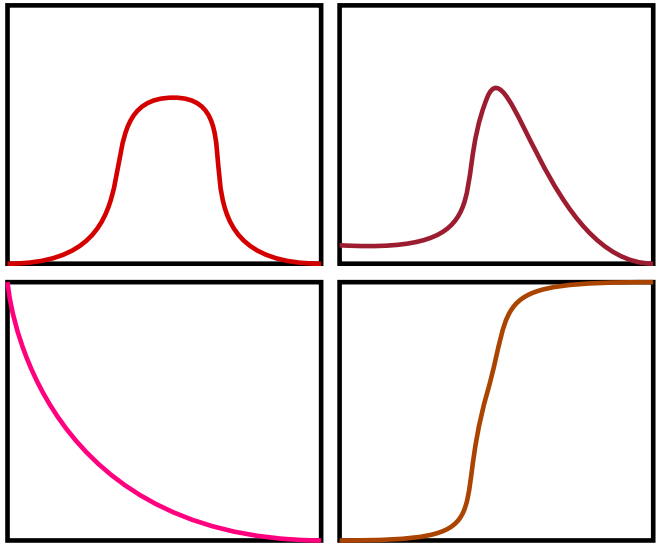
Business



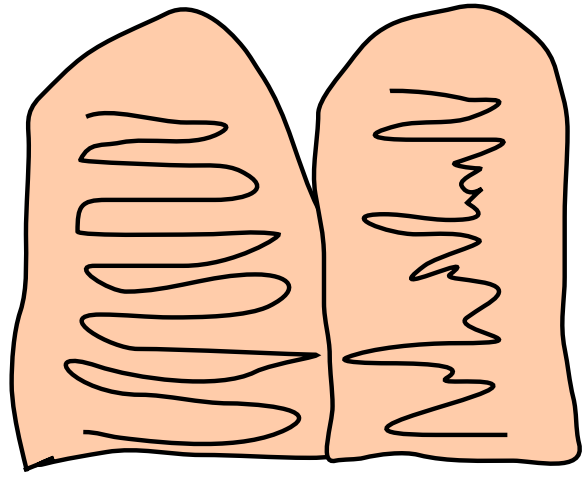


Software Development

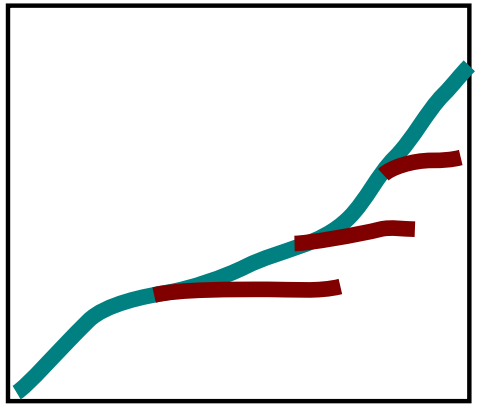
Stochastic Processes



[Herraiz]



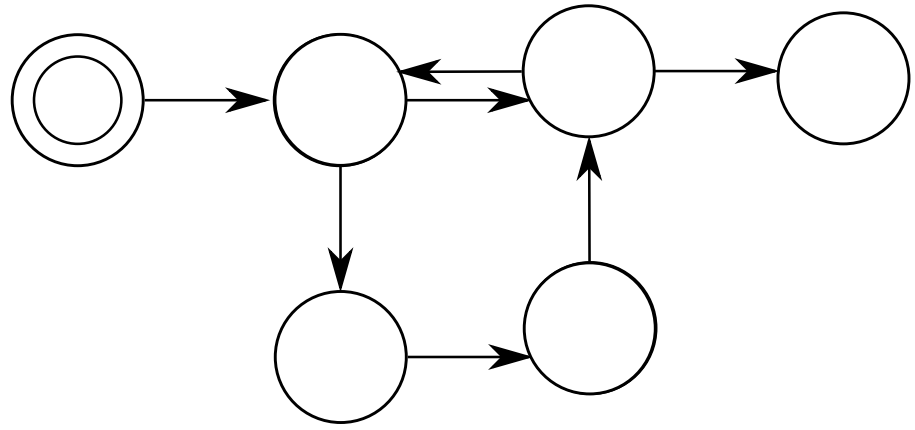
[Lehman]



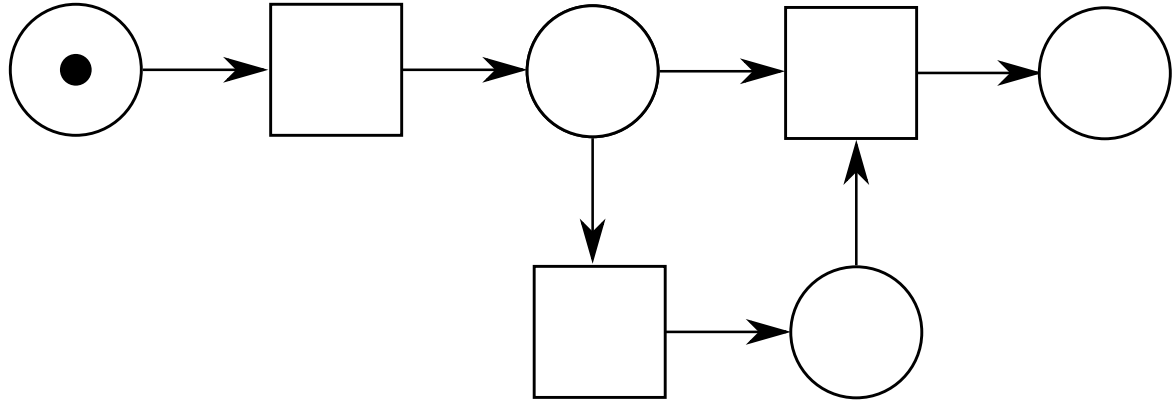
[Turksi]
[Tu]

Business Processes

Finite State Machines



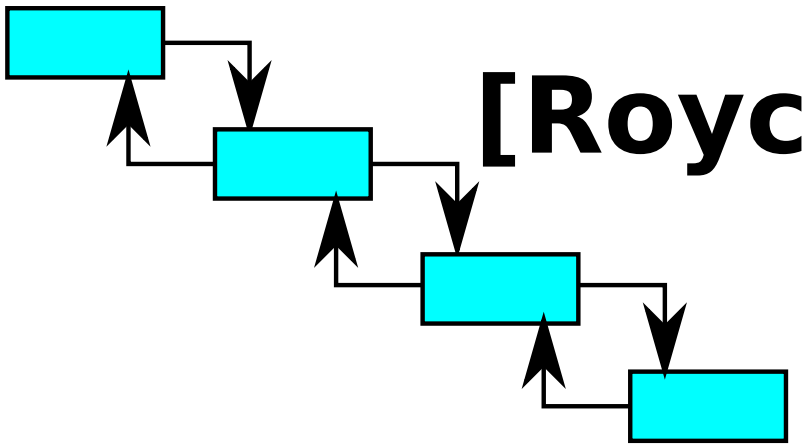
Petrinets



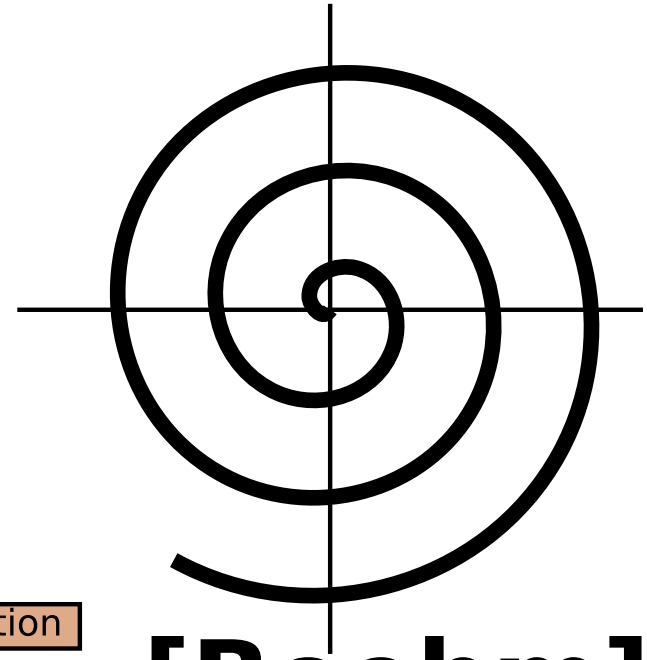
Execution of business goals

[Aalst]

Software Development Processes

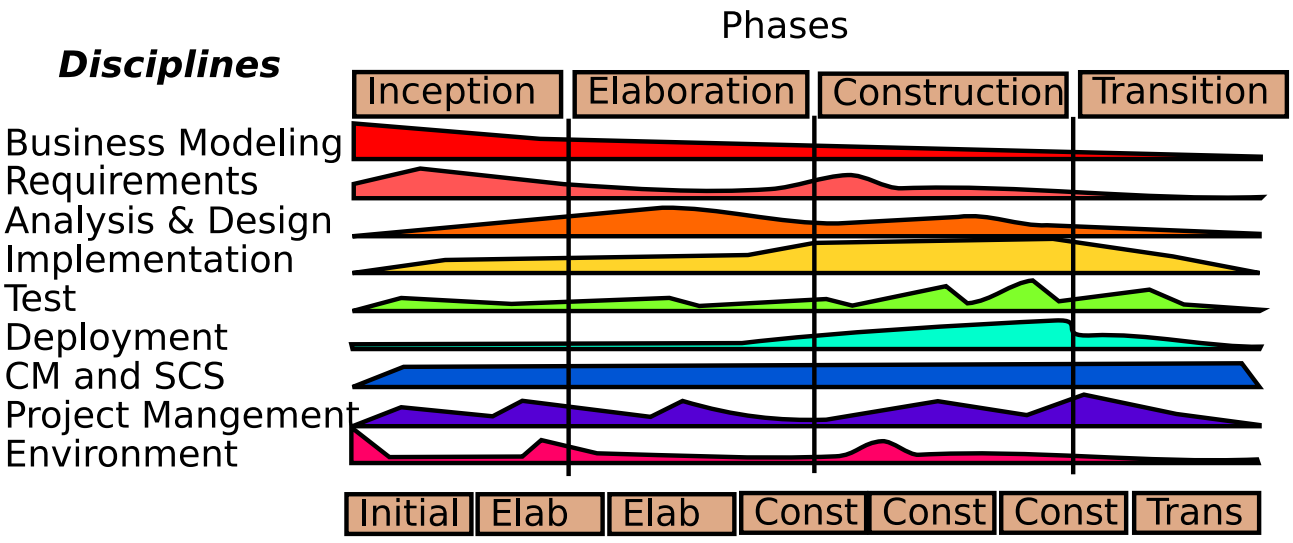


[Royce]



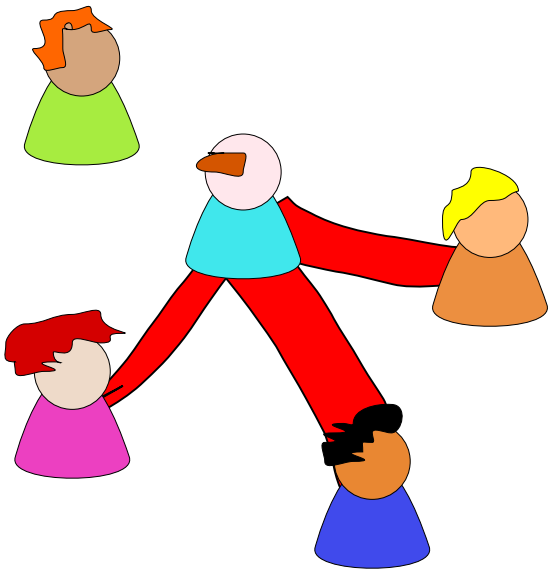
[Boehm]

* CMM
* SDLC

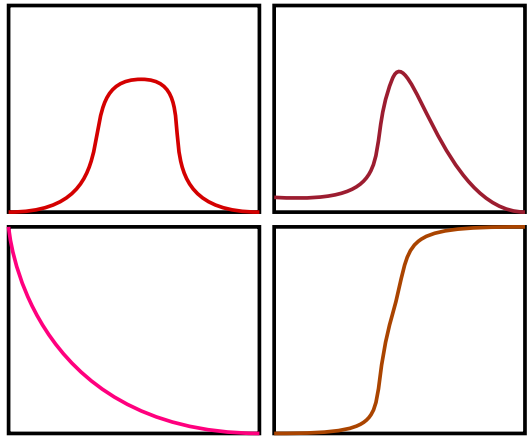


[Rational]

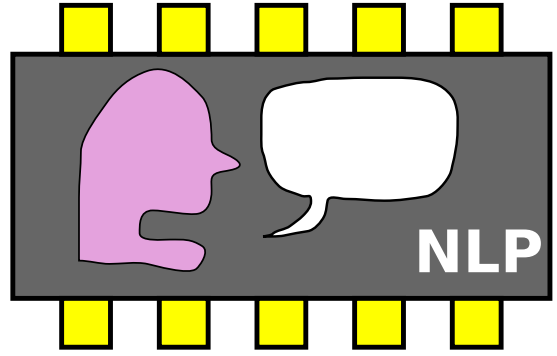
Analysis



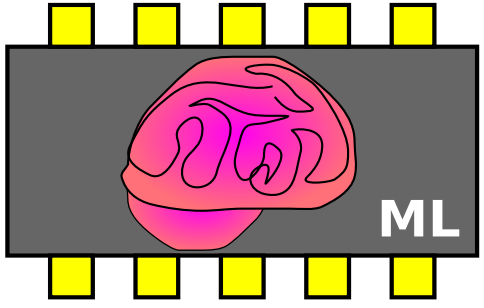
SNA



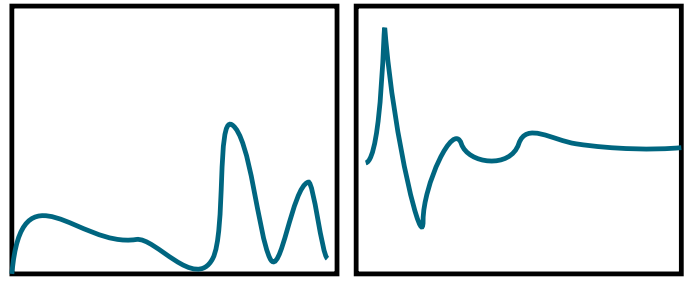
Statistics



NLP

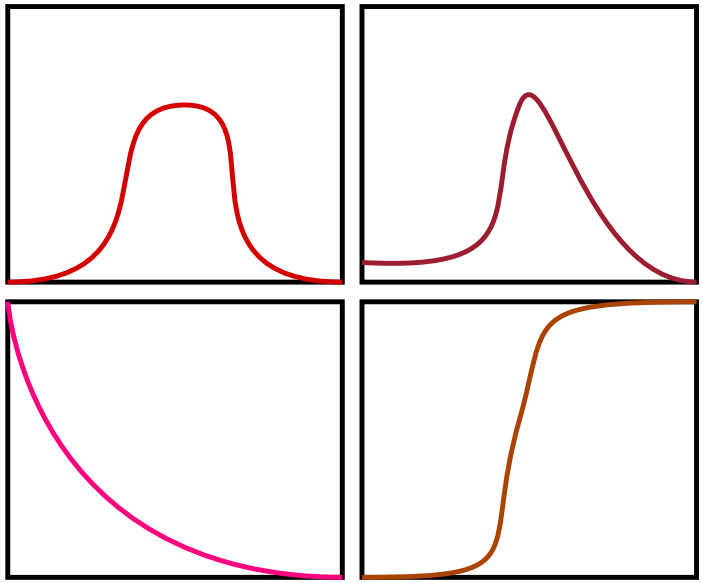


ML

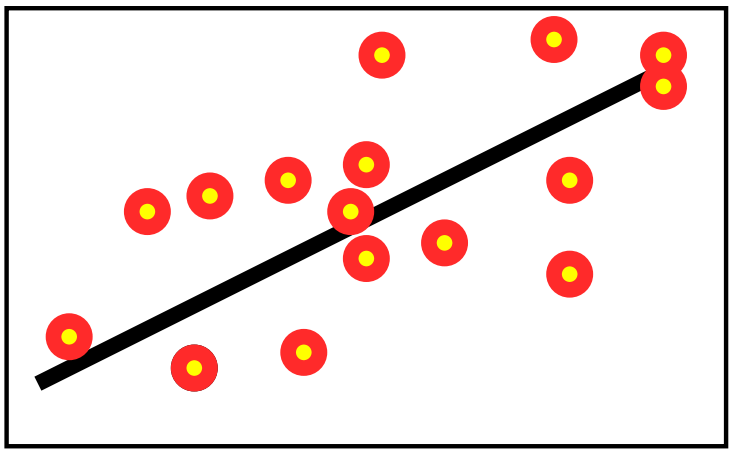


Timeseries

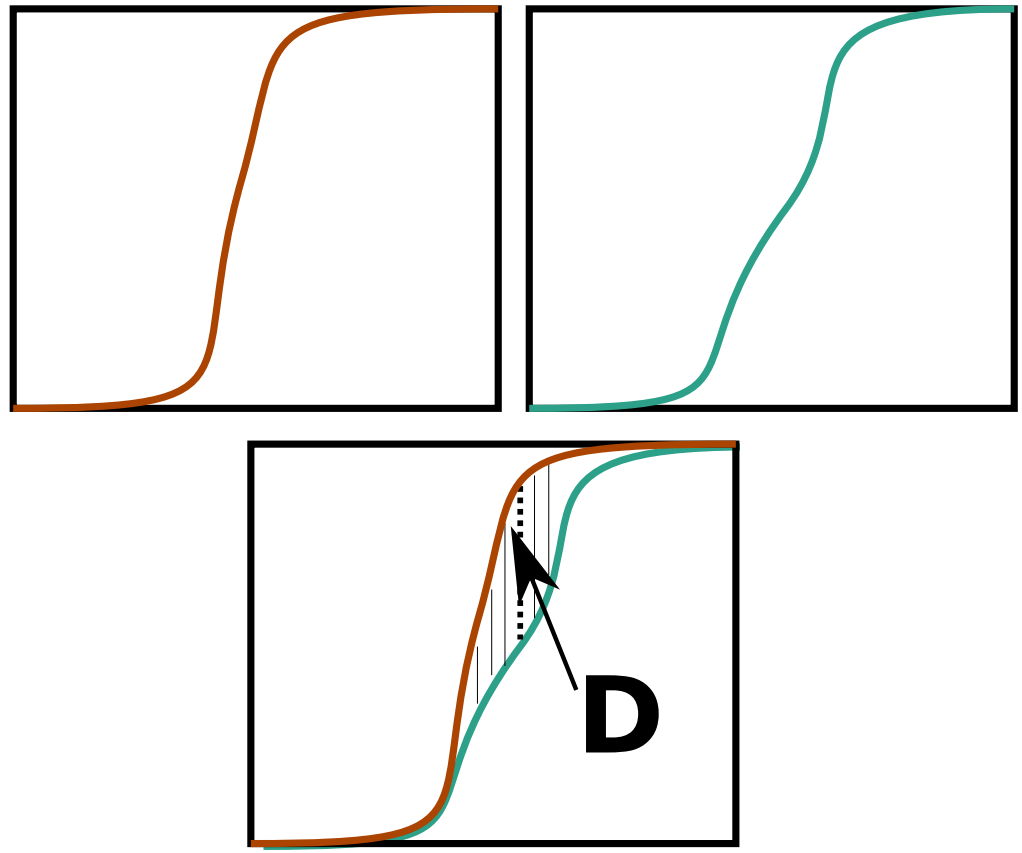
Statistics



Distributions

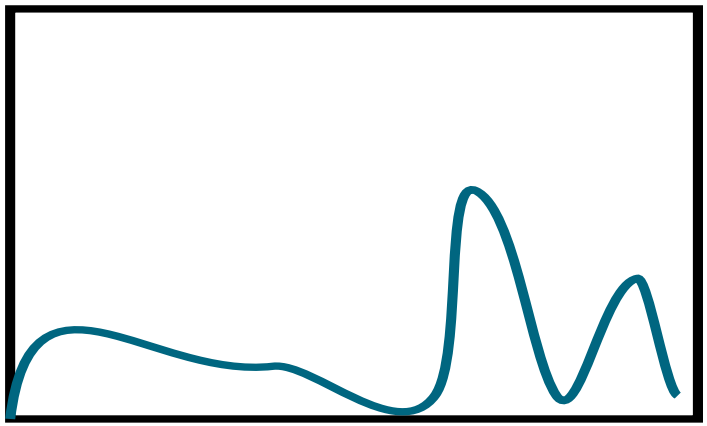


Linear Regression

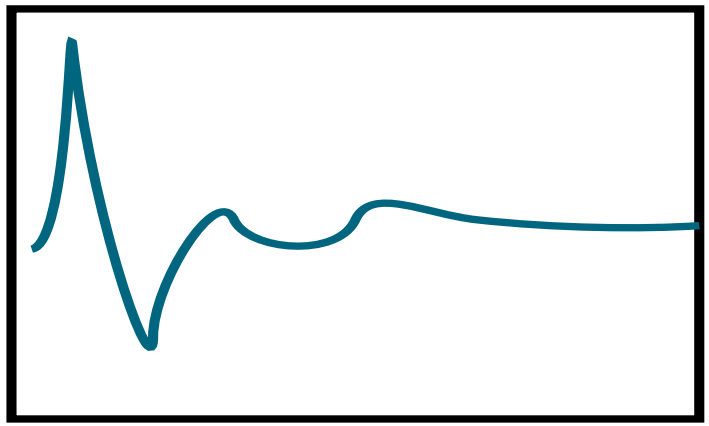


Compare
distributions

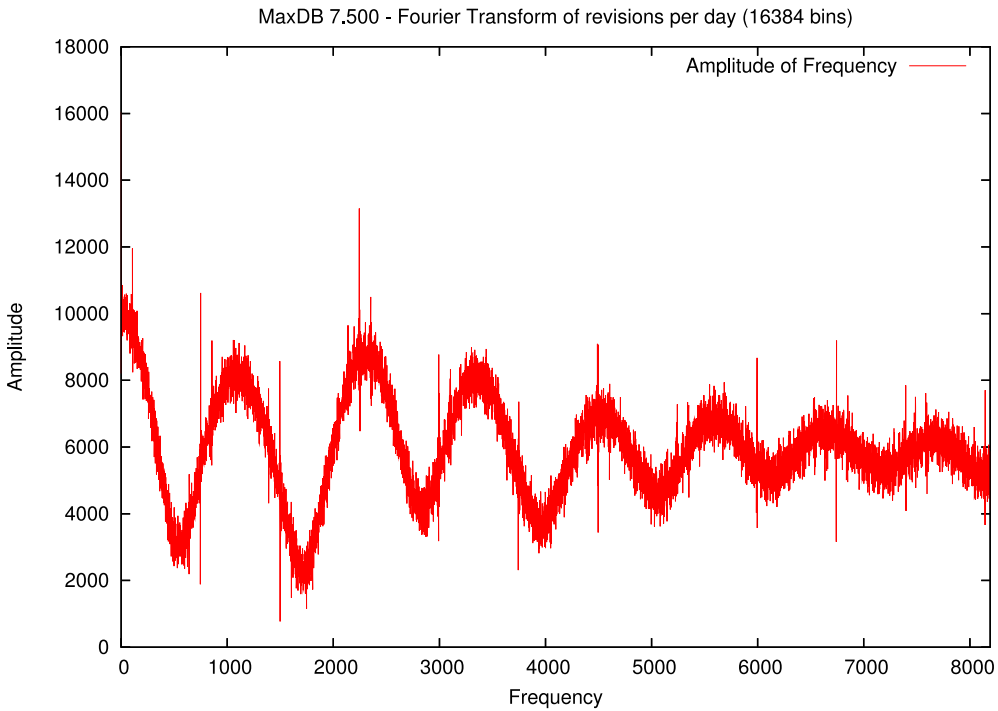
Timeseries



Timeseries

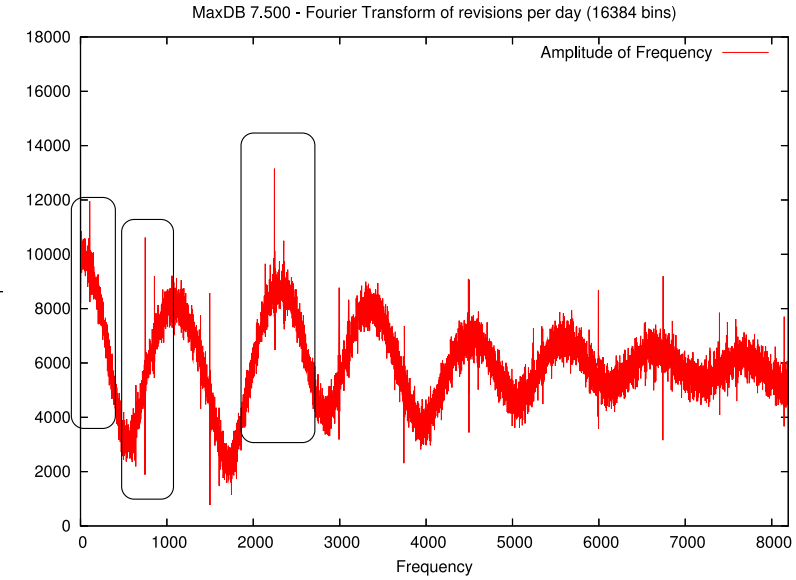


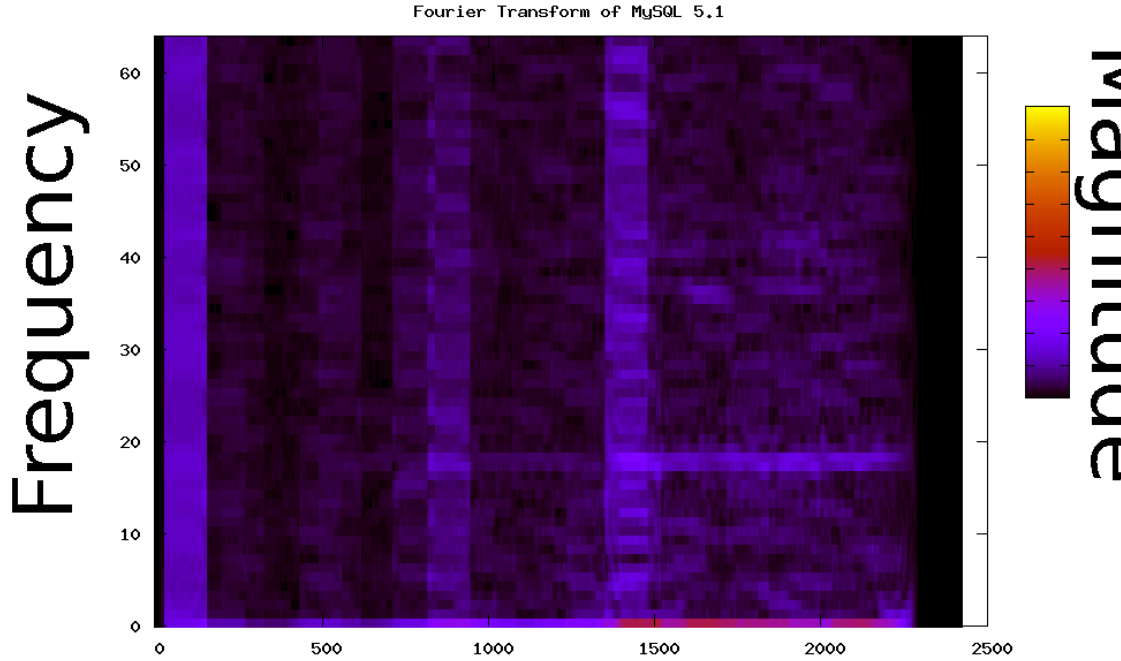
Autocorrelation

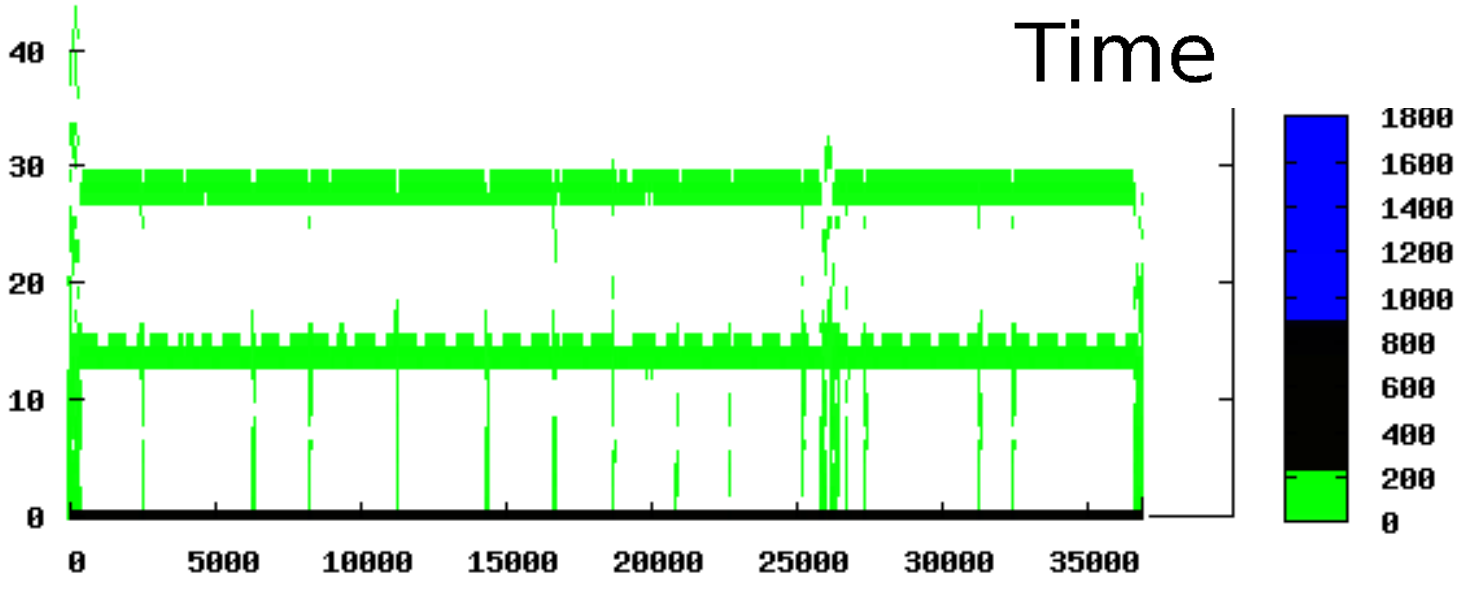


[Herraiz]
[Hindle]

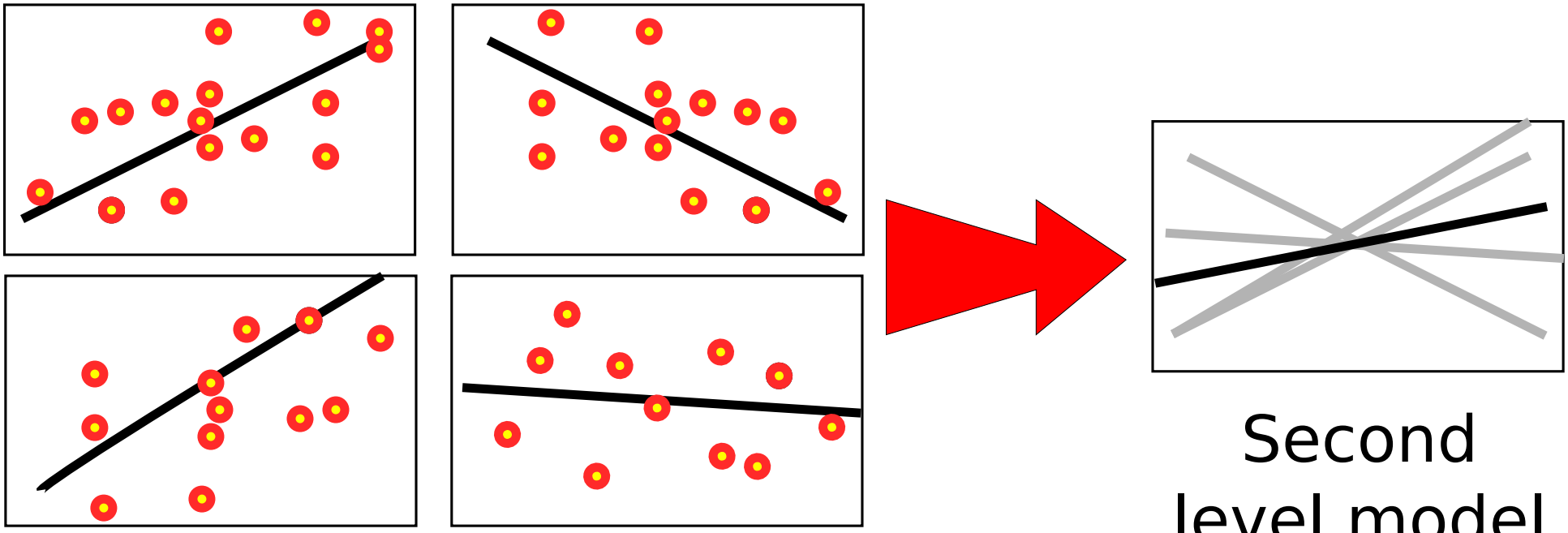
Mining Recurrent Activities: Fourier Analysis of Change Events







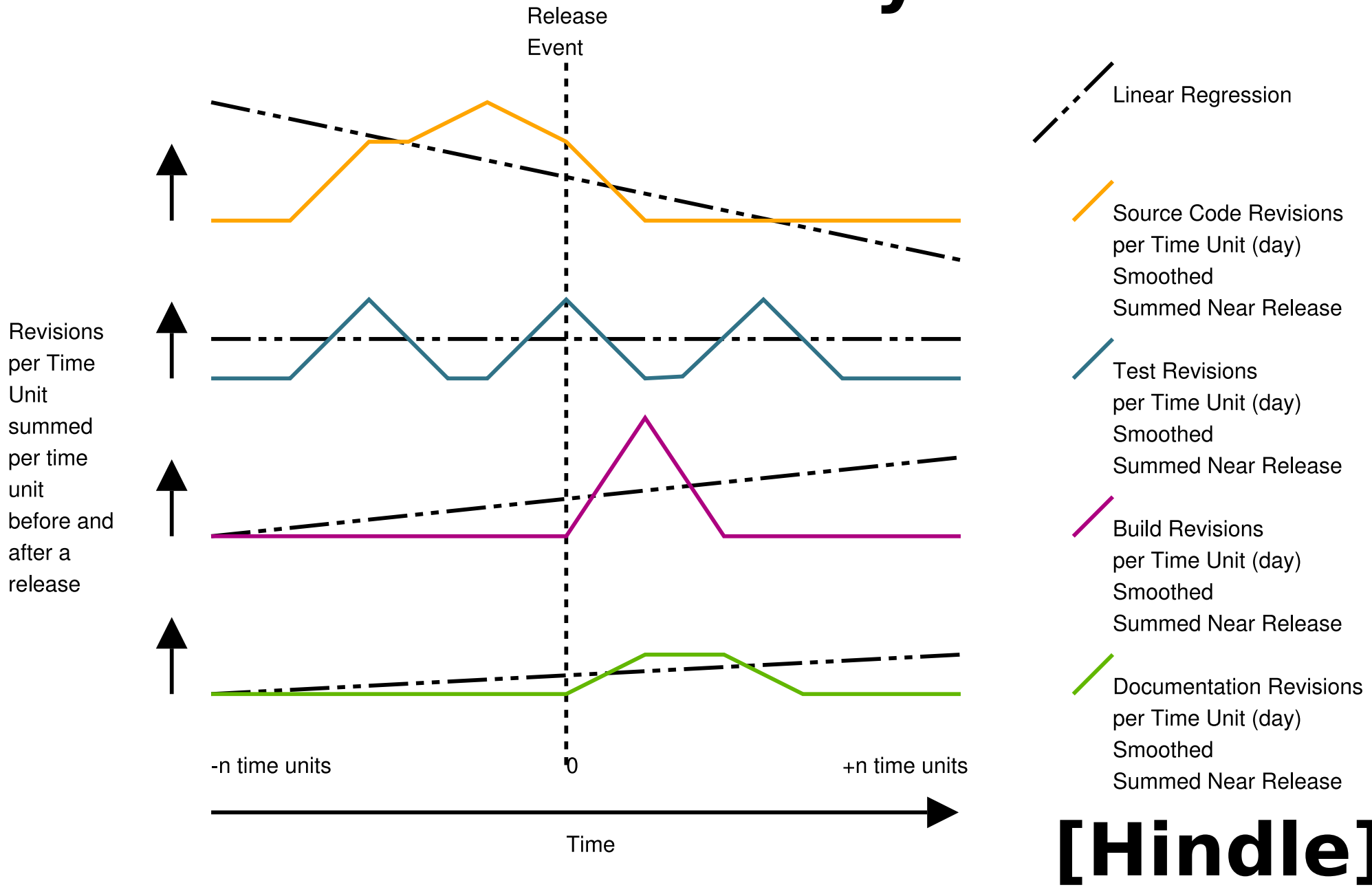
Longitudinal Studies and Multilevel Modelling



First level model

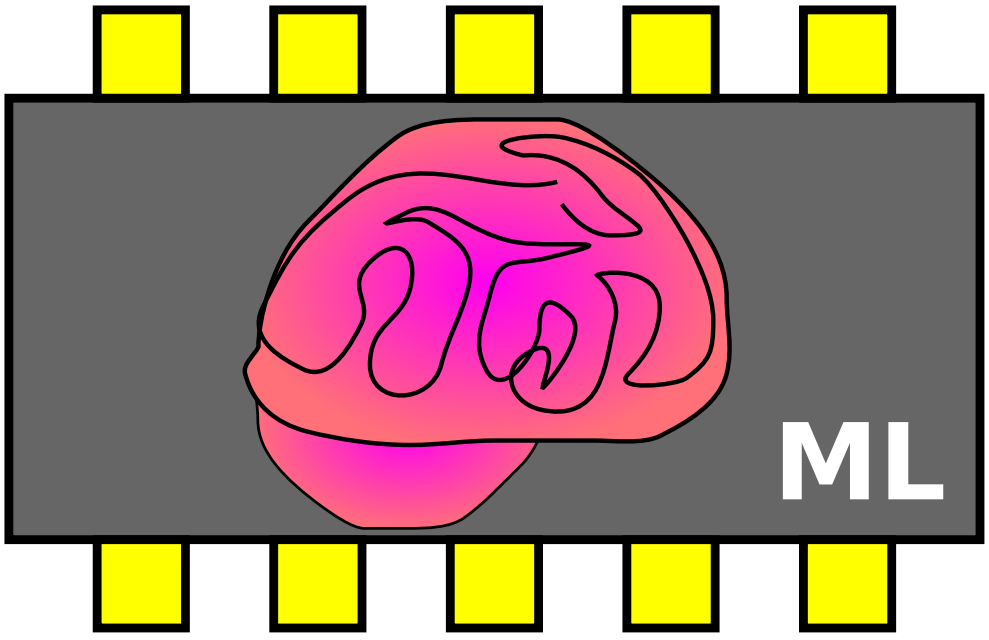
Second level model

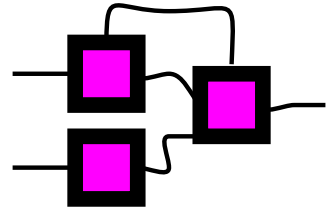
Release Discovery

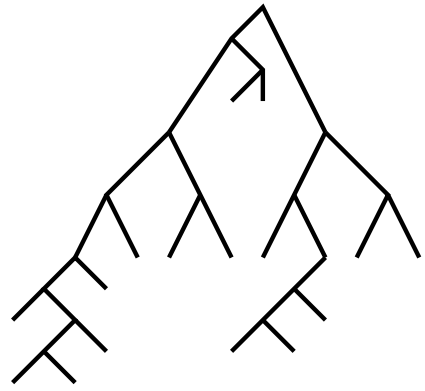


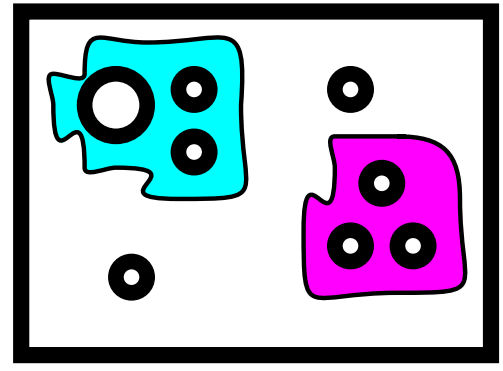
[Hindle]

Machine Learning

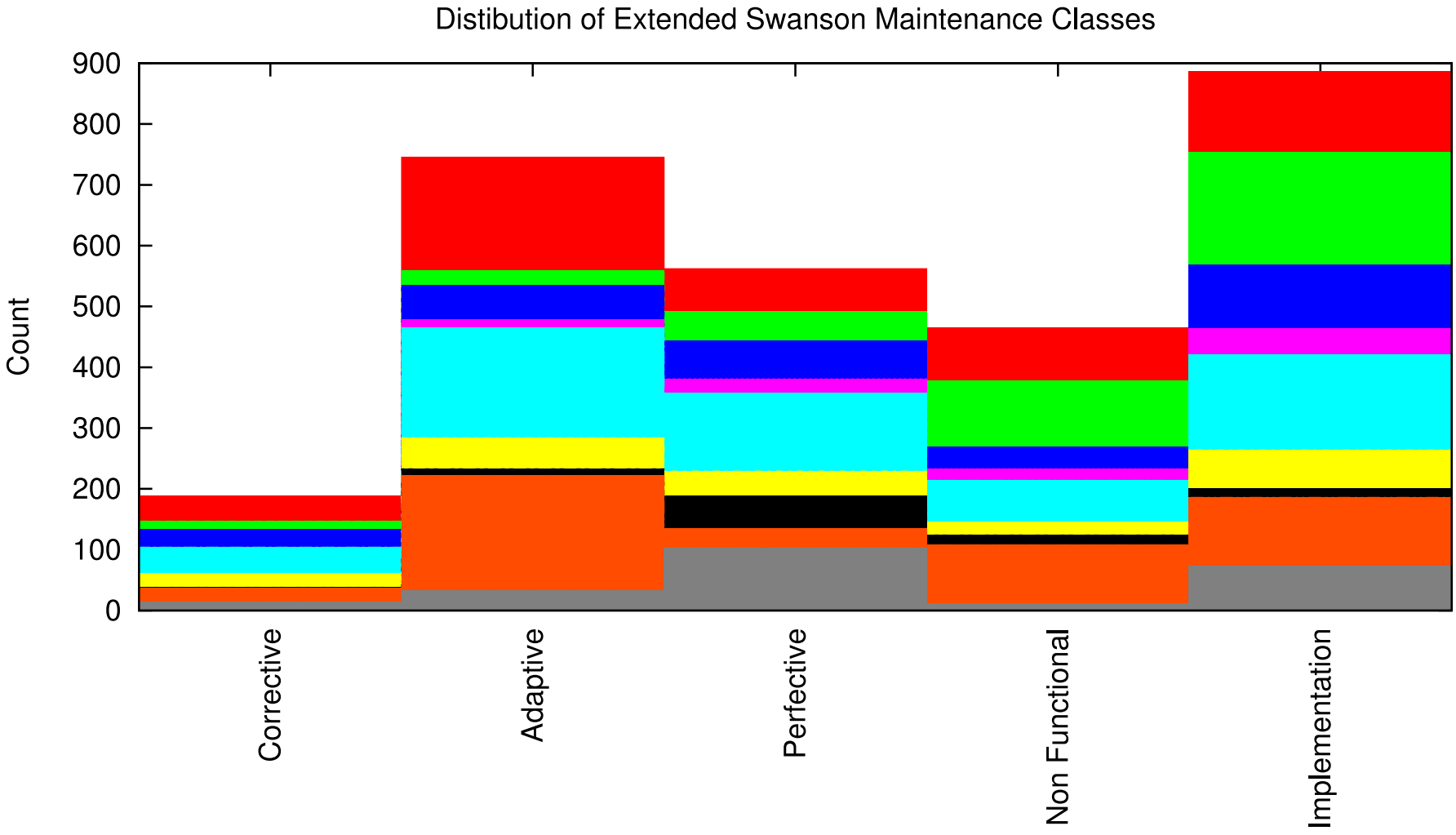






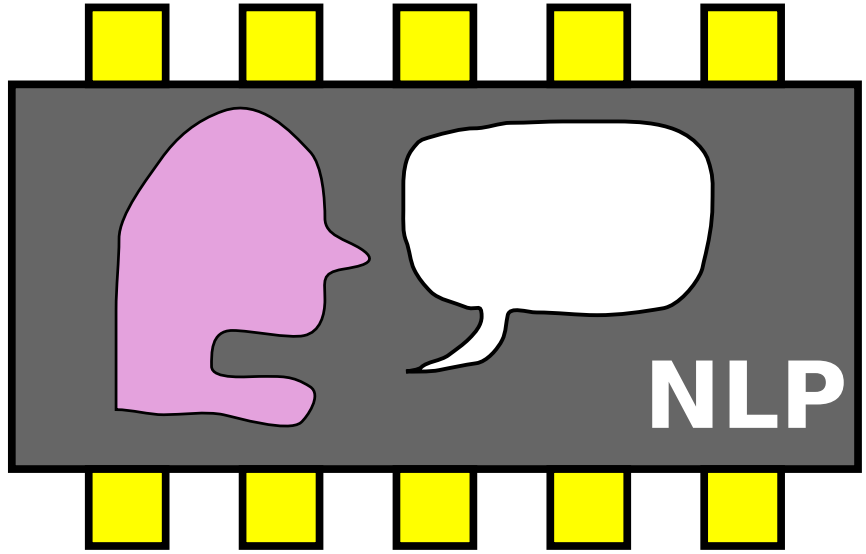


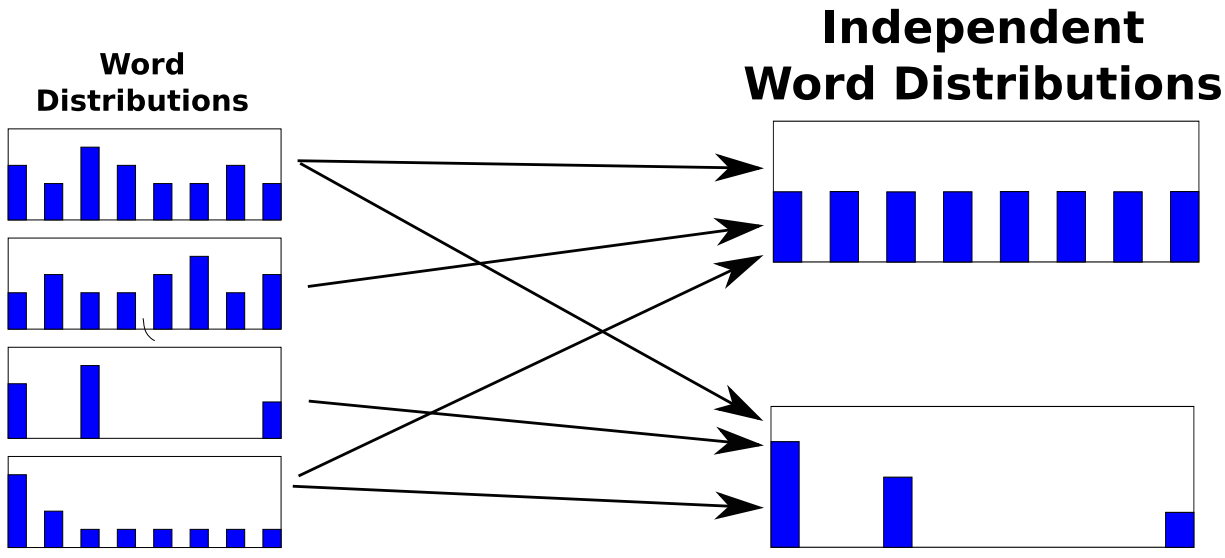
Automatic Classification of Large Changes into Maintenance Categories



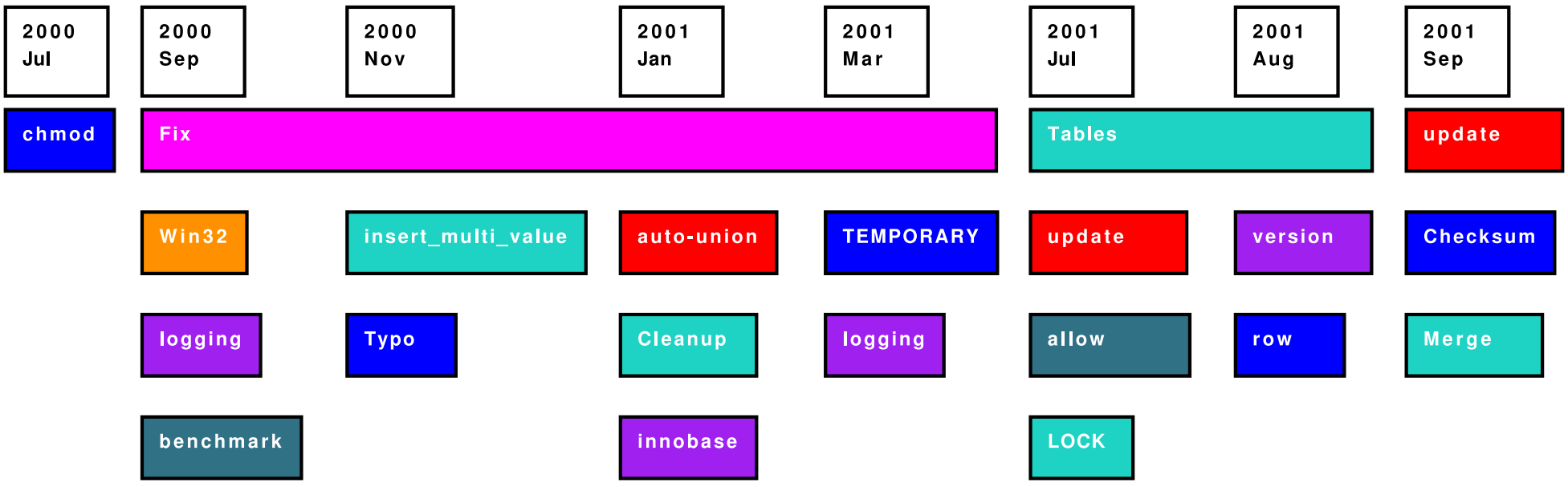
[Hindle]

Natural Language Processing





Topic Analysis: MySQL 3.23

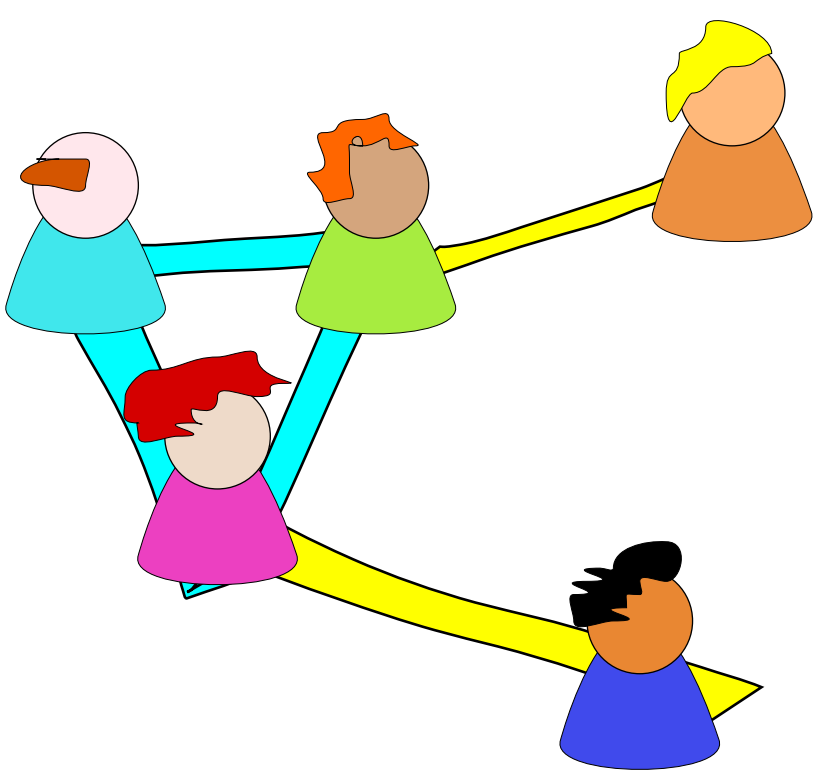


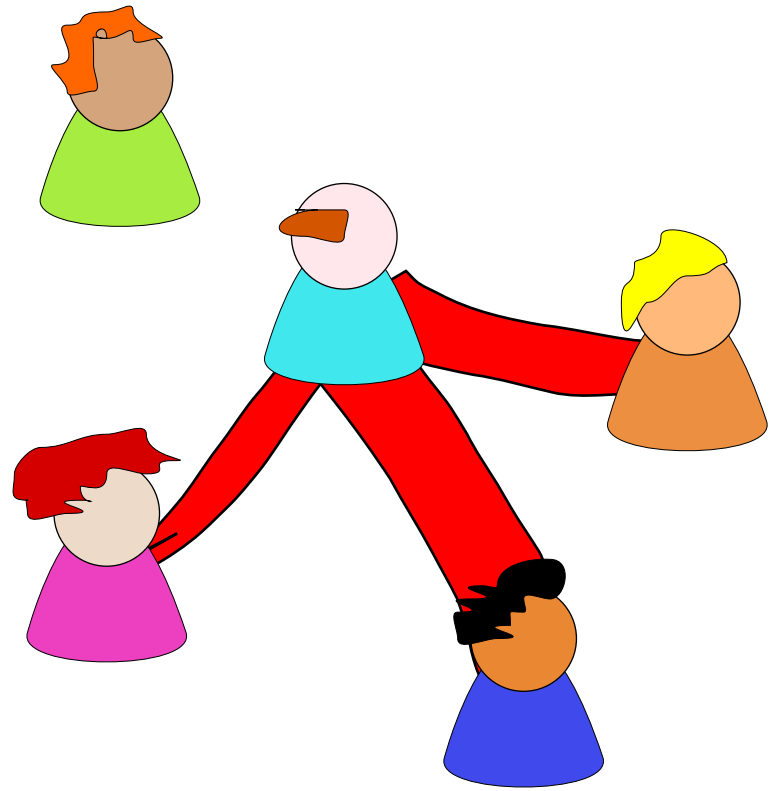
Extracted topics across time

What's hot what's not: developer topic analysis

[Hindle]

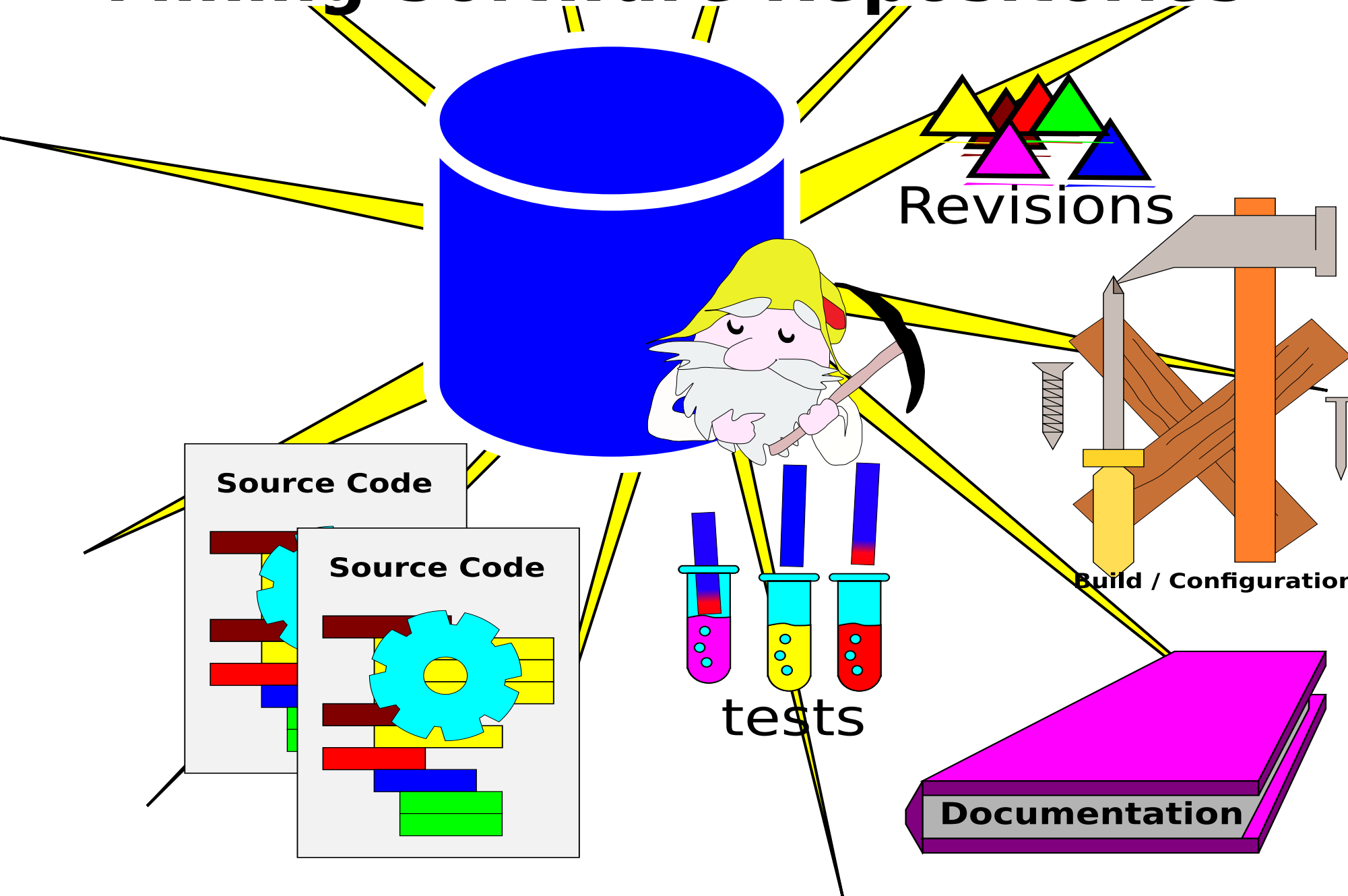
Social Network Analysis



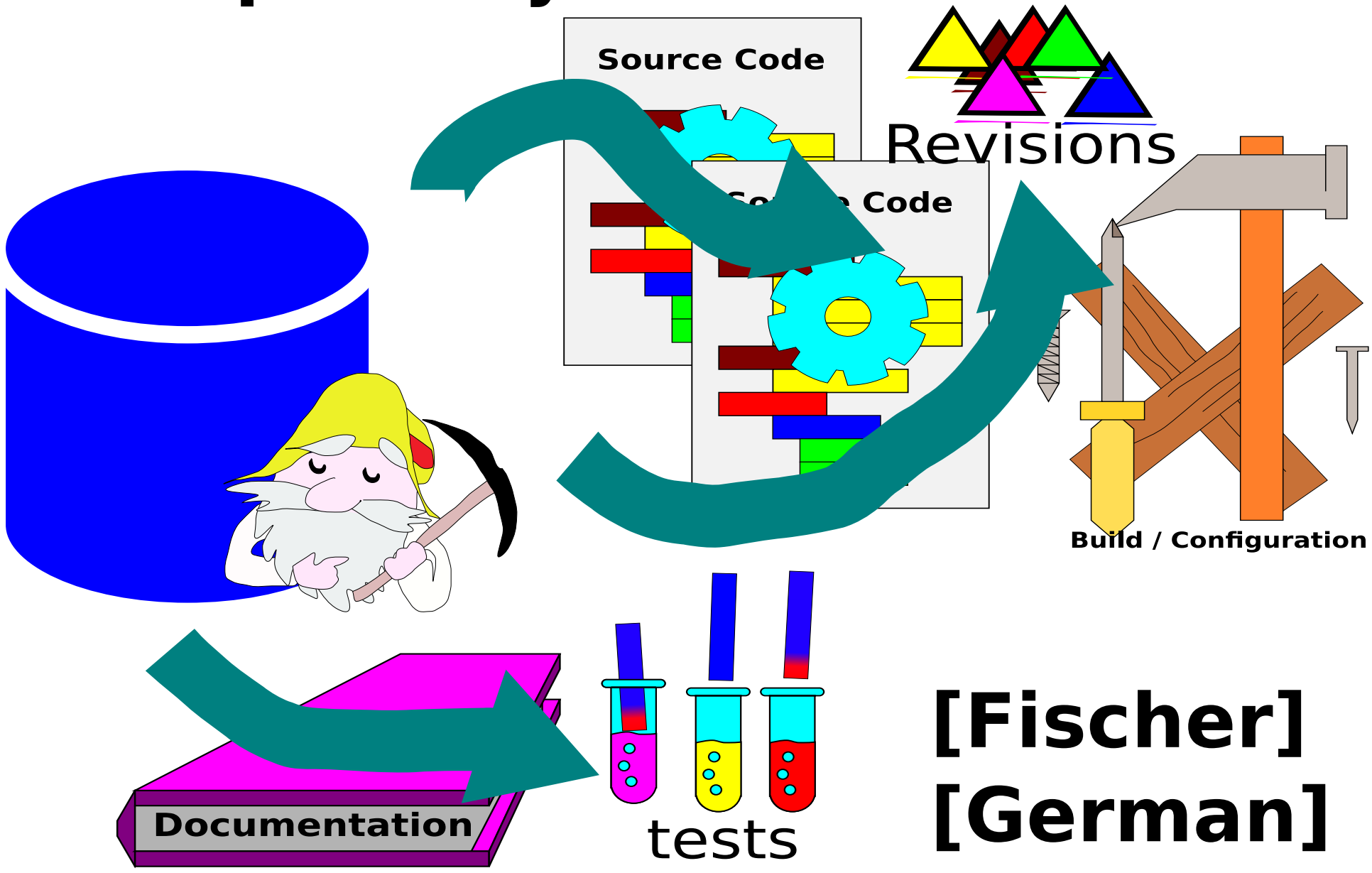


[Bird]

Mining Software Repositories

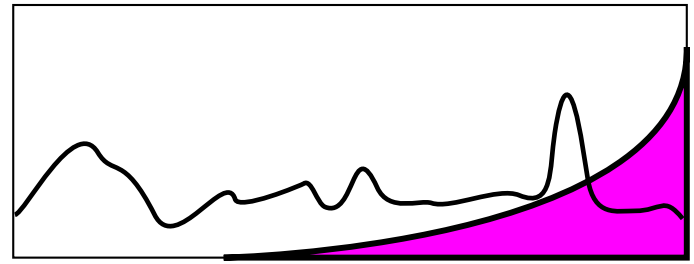


Repository Extraction

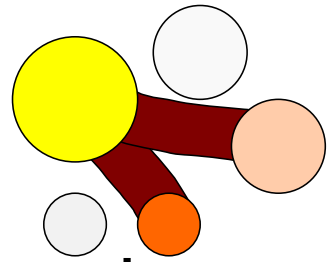


[Fischer]
[German]

Prediction



window decay

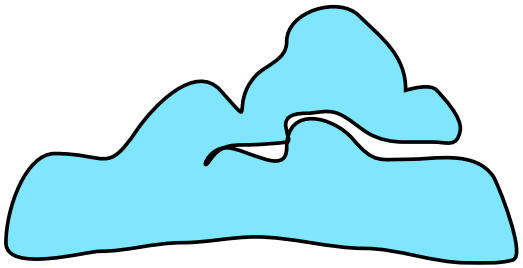


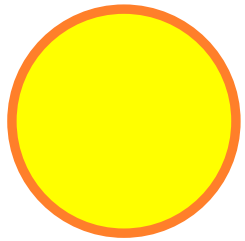
co-change

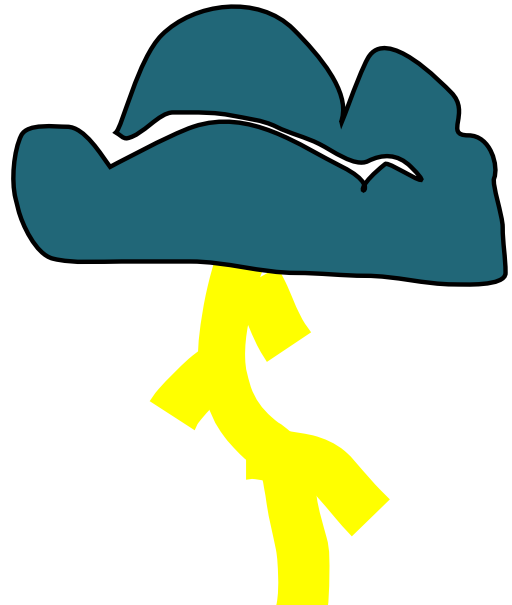
[Girba]

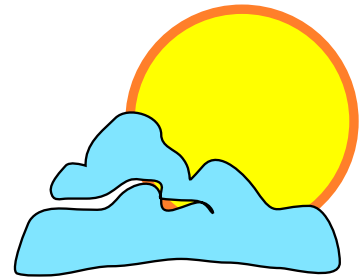
[Askari]

[Hassan]

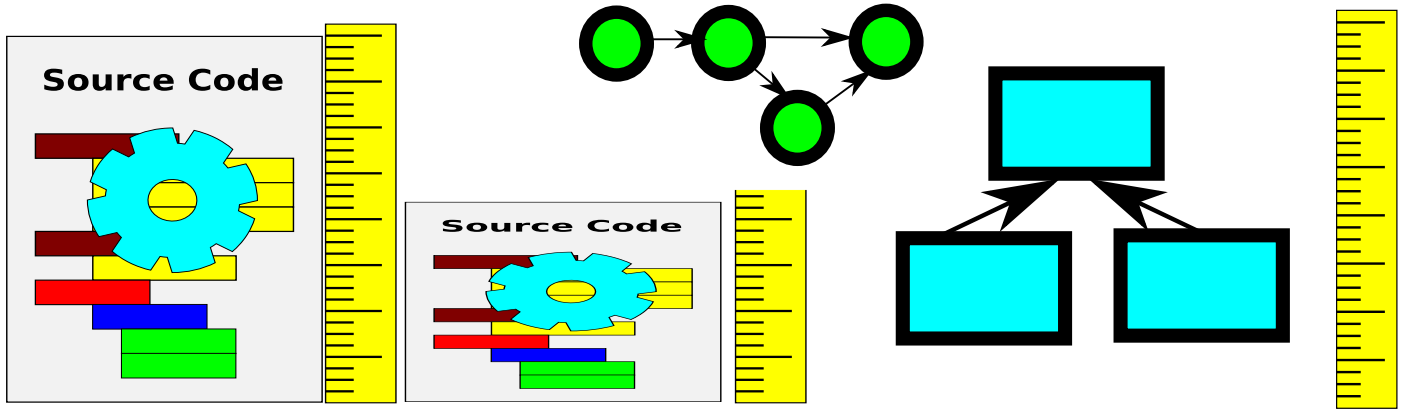




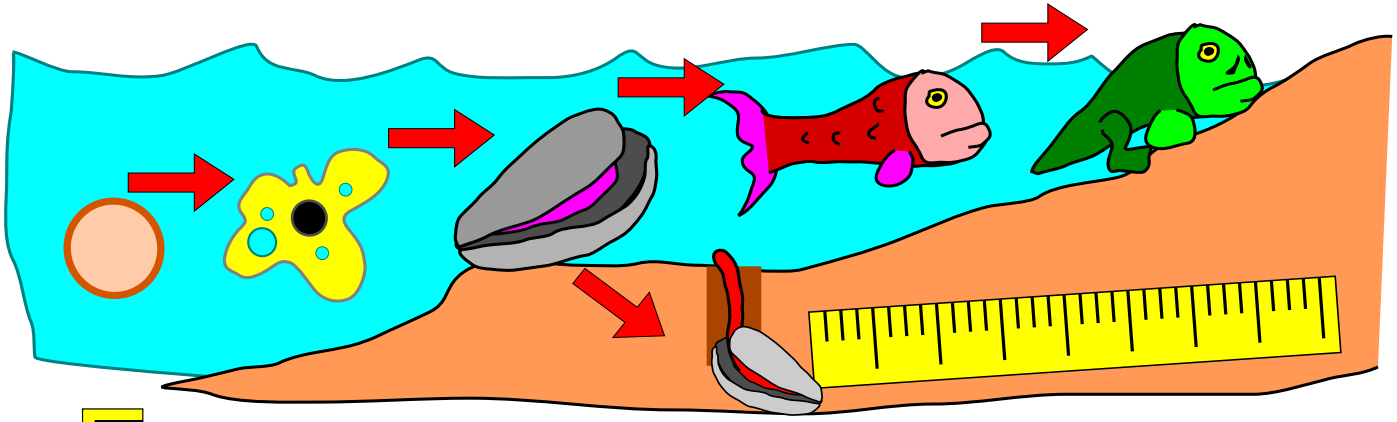




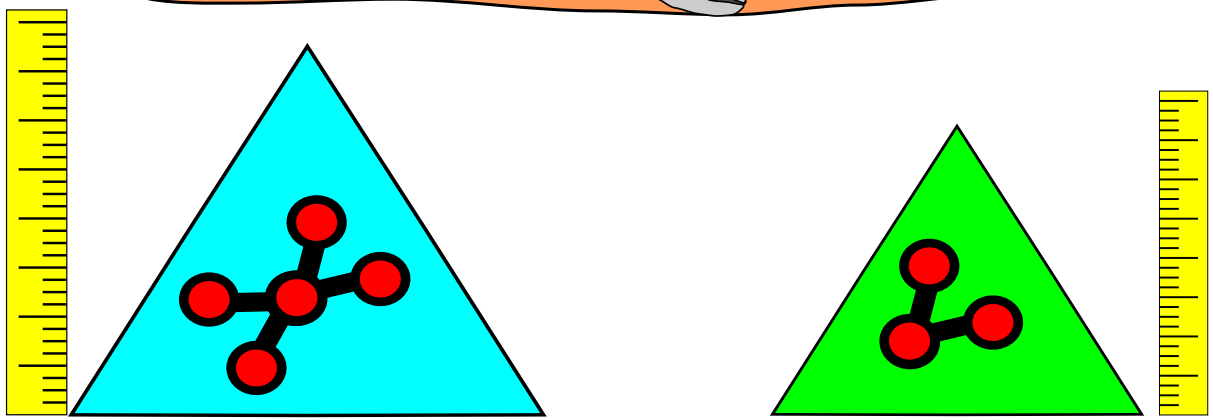
Metrics



Source code Metrics

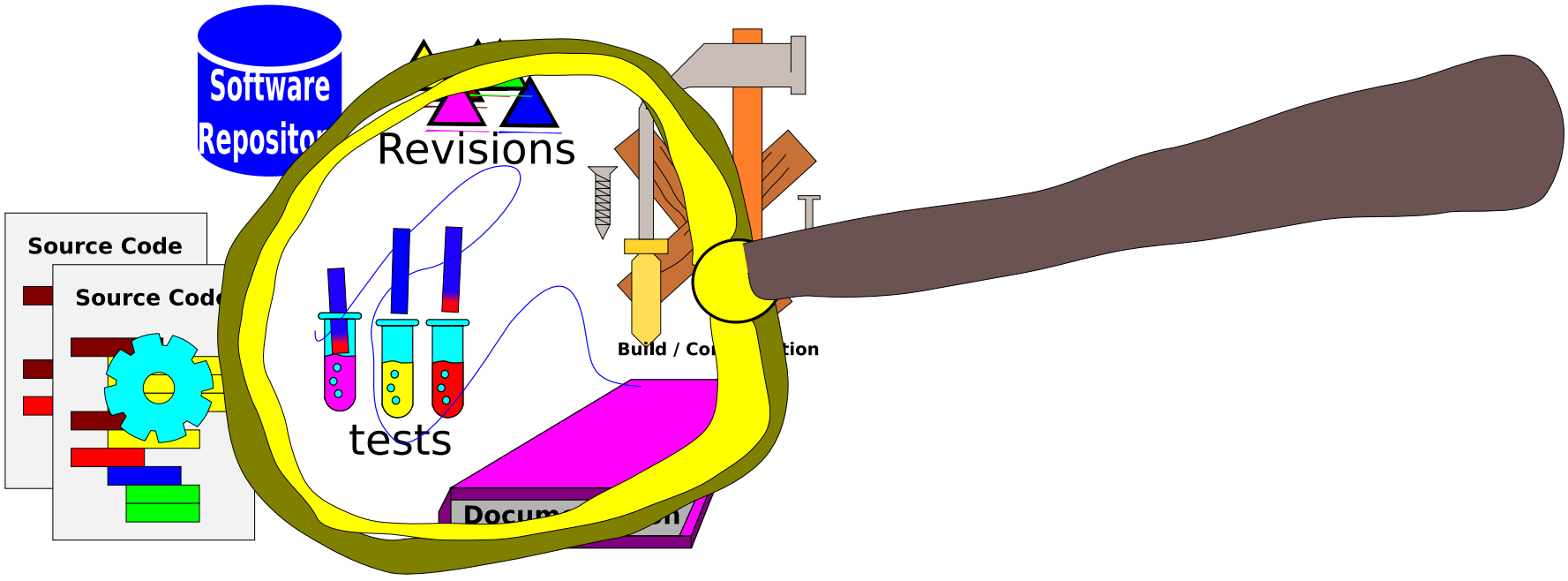


Software Evolution Metrics



Coupling Metrics

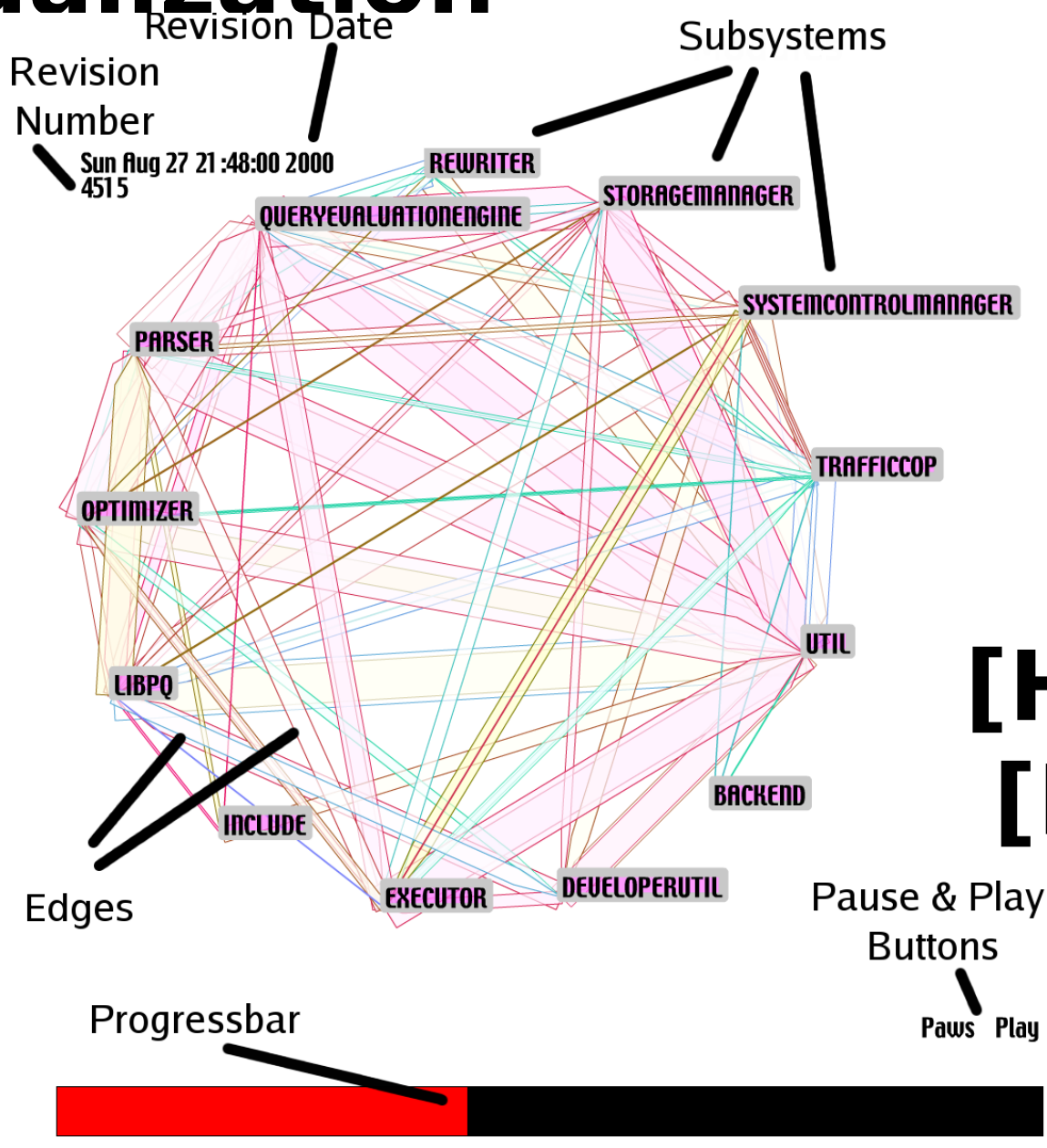
Querying



* 1st Order Logic
* Temporal Logic
* Unification

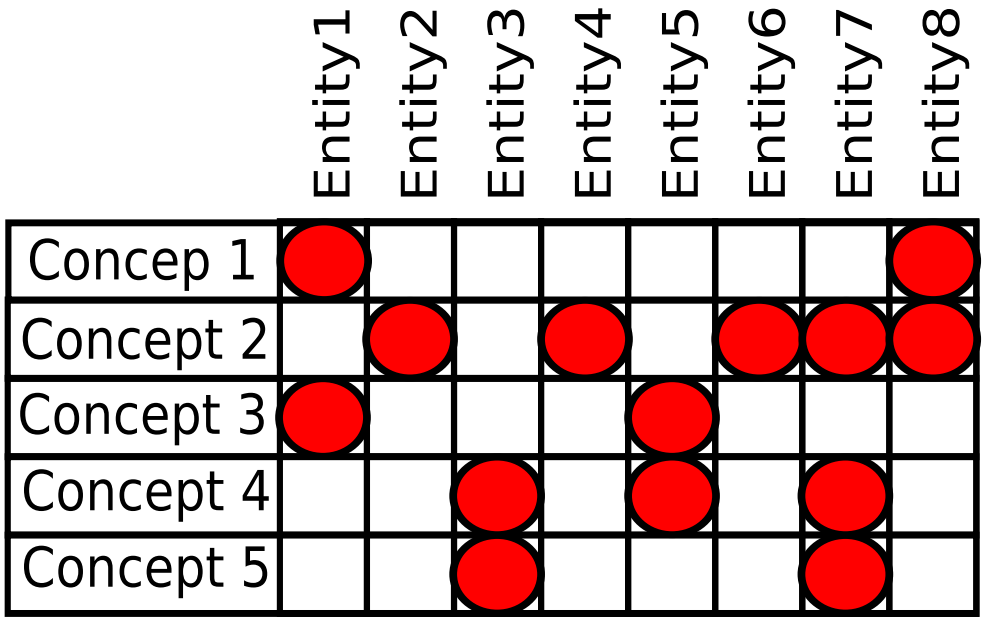
[Cubranic]
[Kim]
[Hindle]

Visualization

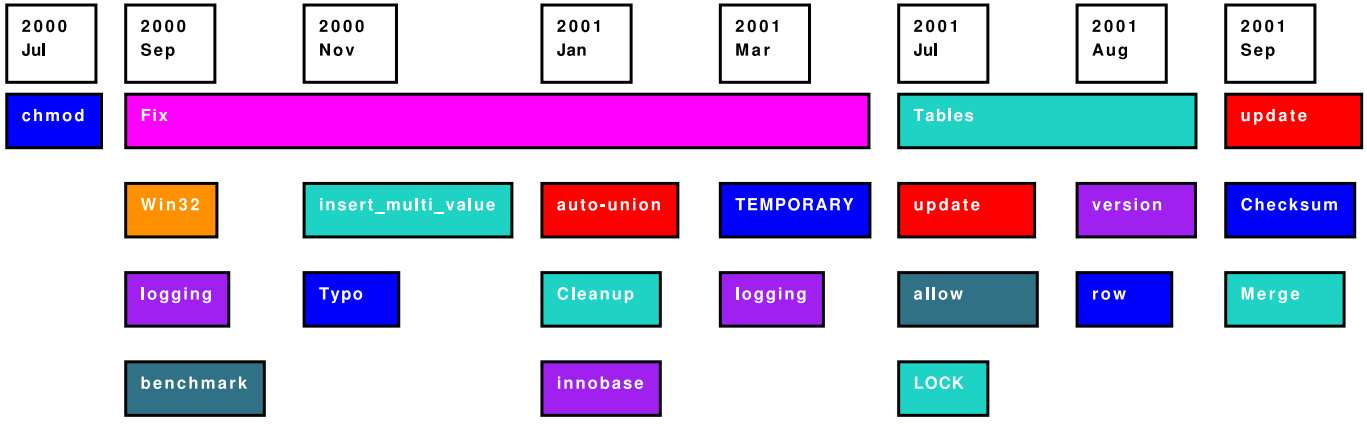


[Hindle]
[Lanza]

Topic/Concept Analysis

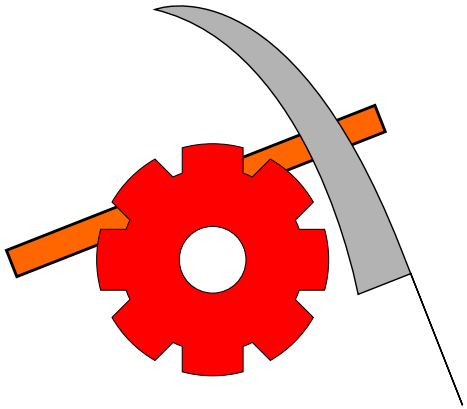


[Lukins]
[Linstead]
[Maletic]
[Poshyvanik]
[Marcus]
[Hindle]

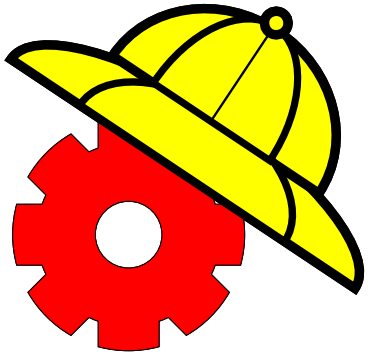


Software Processes & MSR

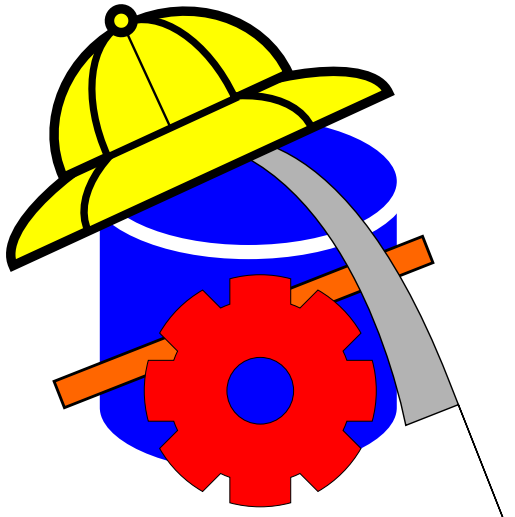
[Aalst]
[Cook]
[Jensen]
[German]
[Hindle]



Process Mining

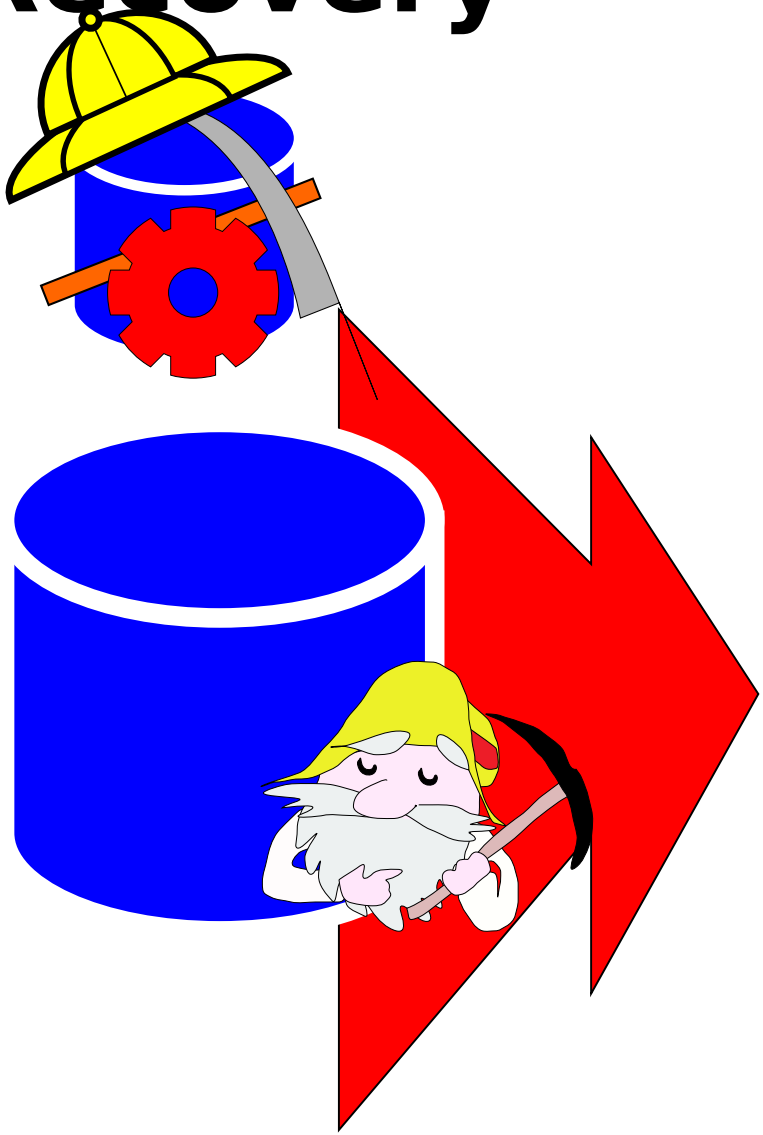


Process Discovery

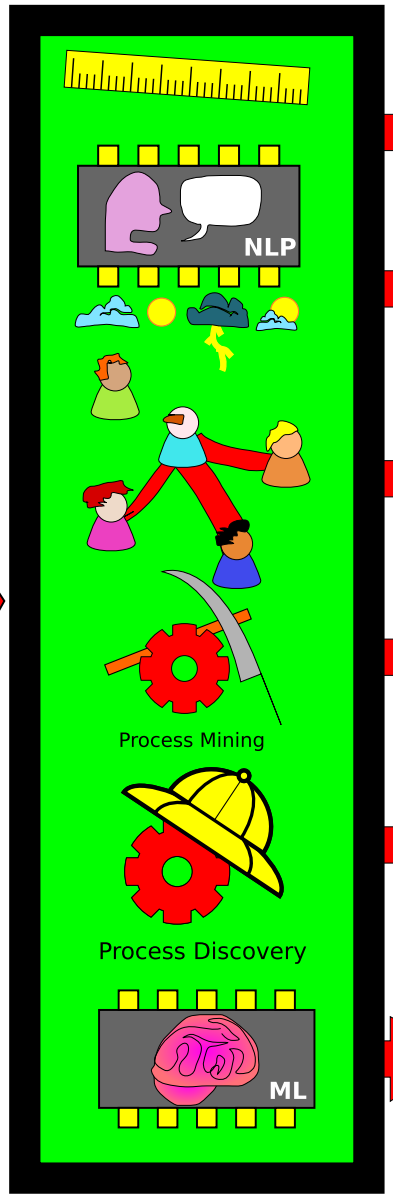


Process Recovery

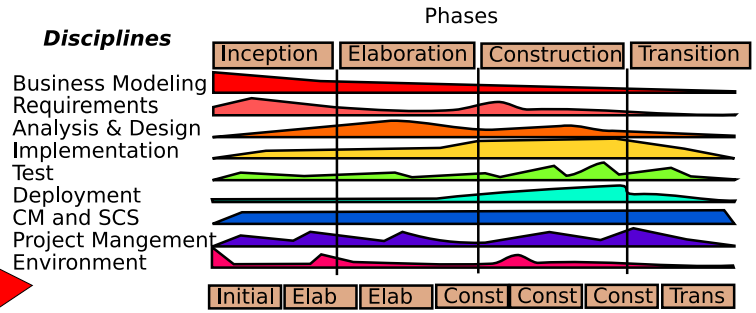
Process Recovery

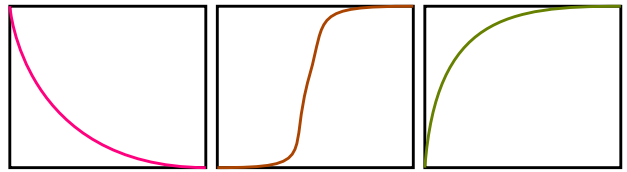


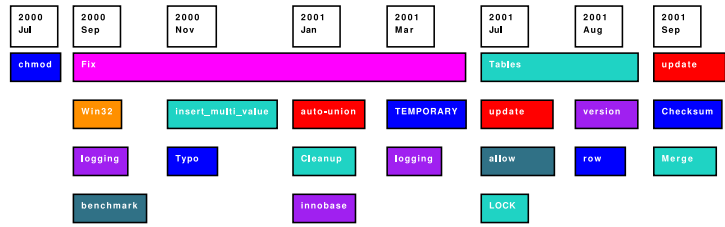
Tools

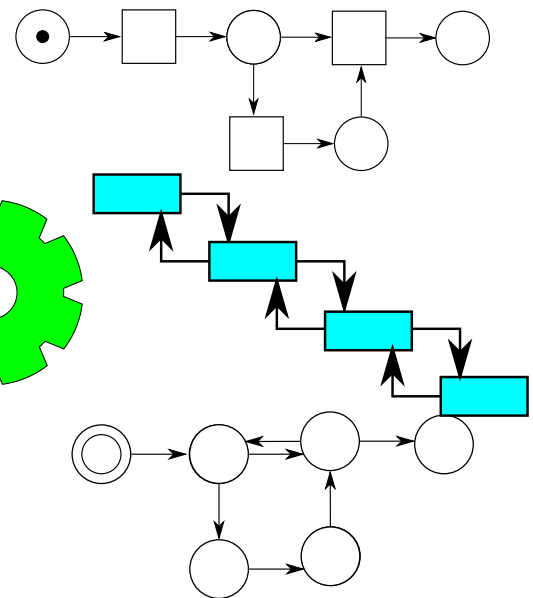


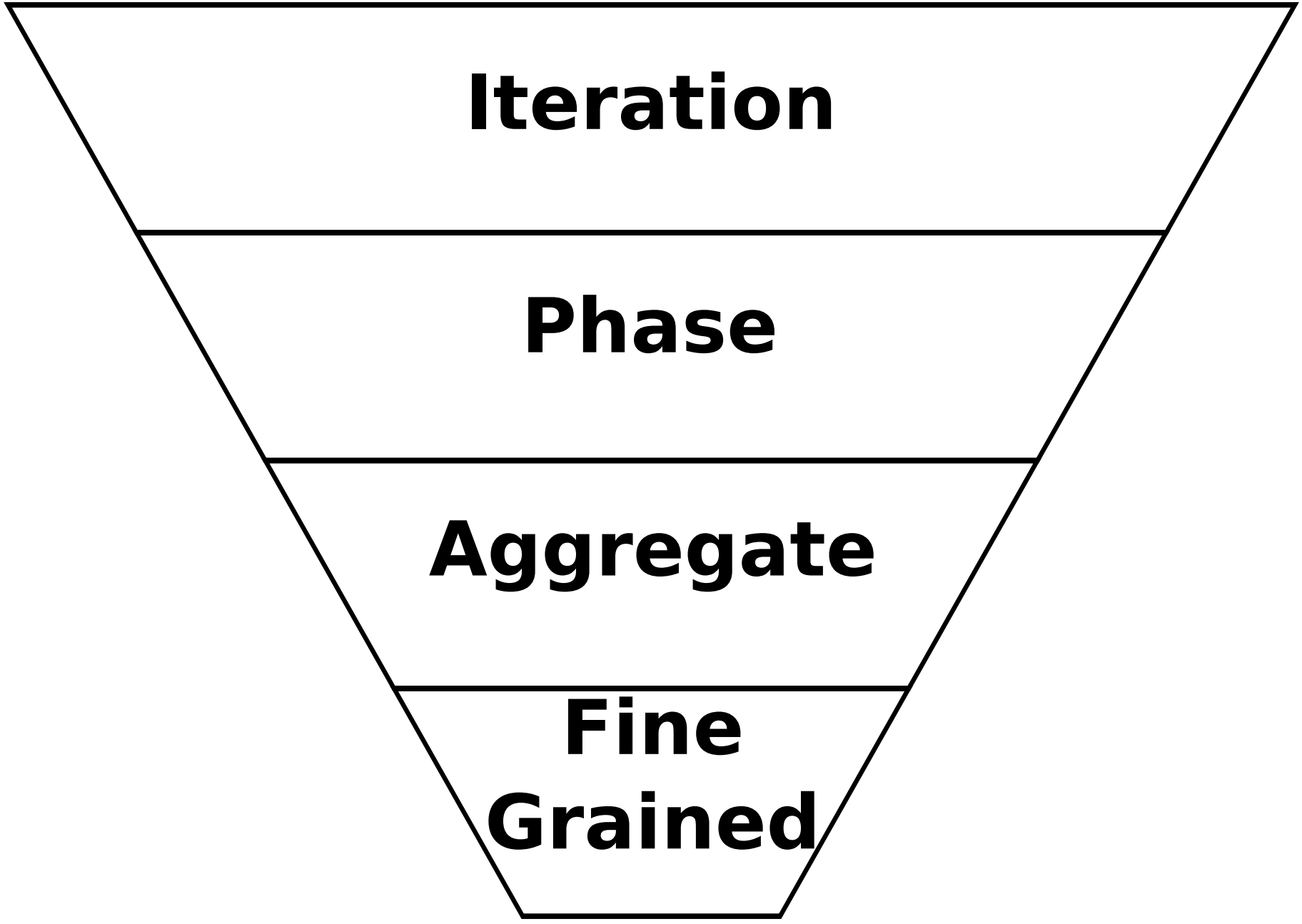
Processes

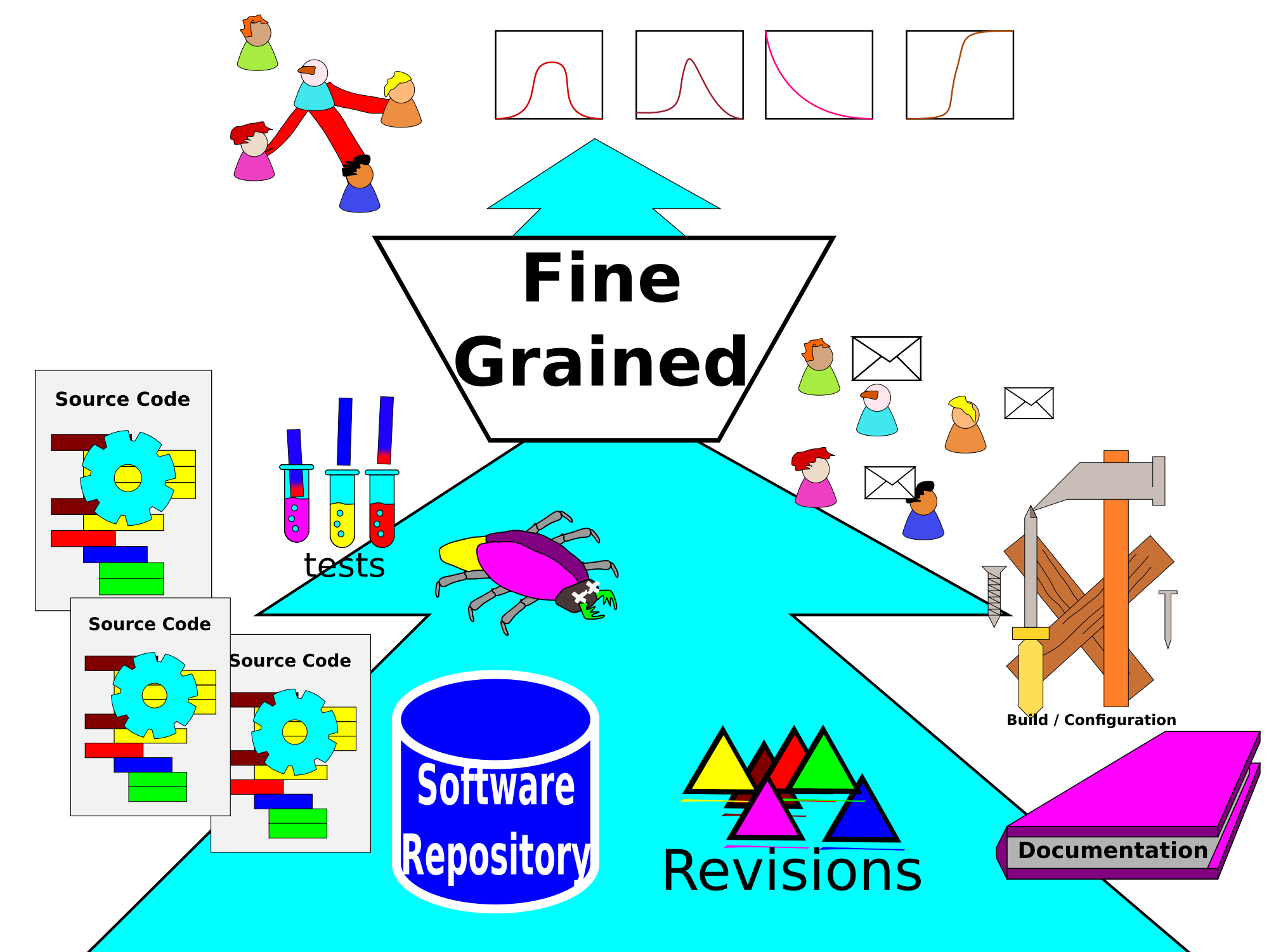


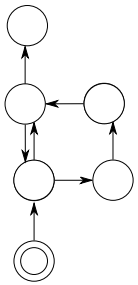


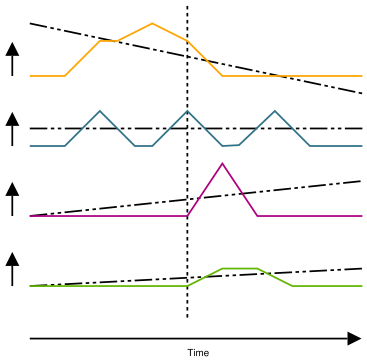


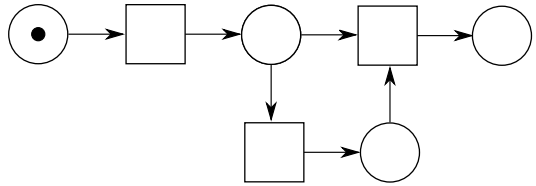


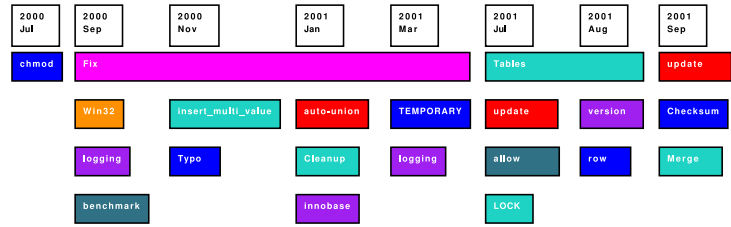

Iteration
Phase
Aggregate
Fine
Grained


Fine Grained
Source Code
Source Code
Source Code
tests
Software Repository
Revisions
Build / Configuration
Documentation

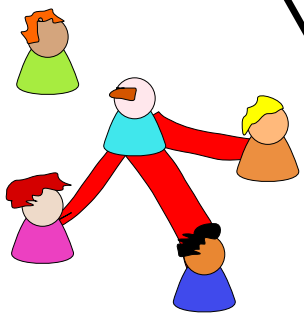


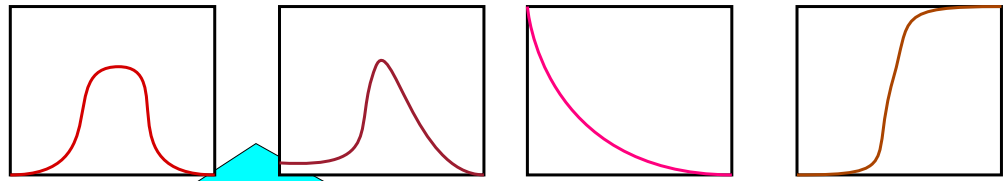




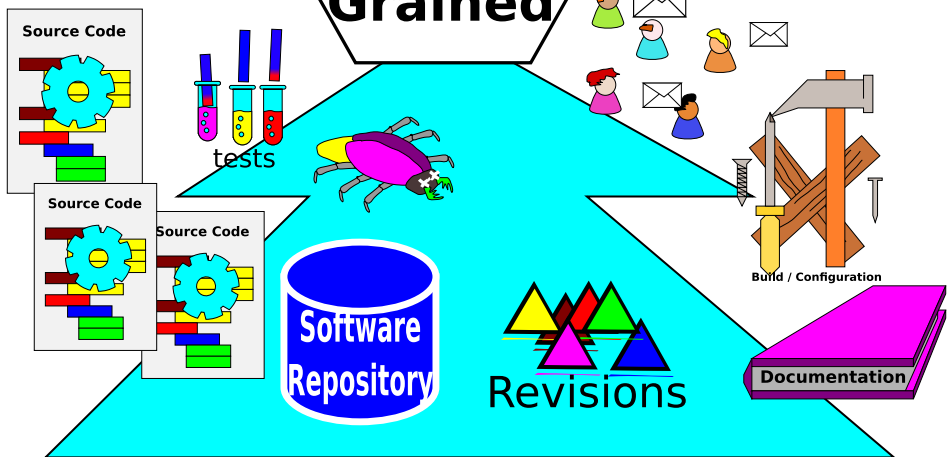


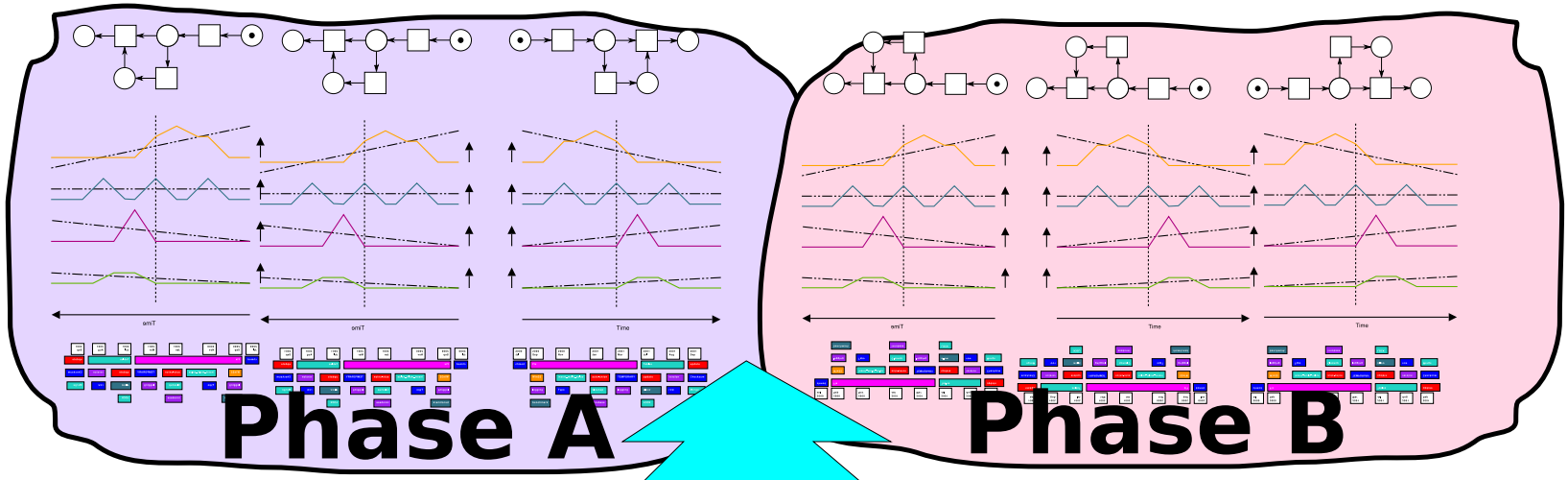
Aggregate



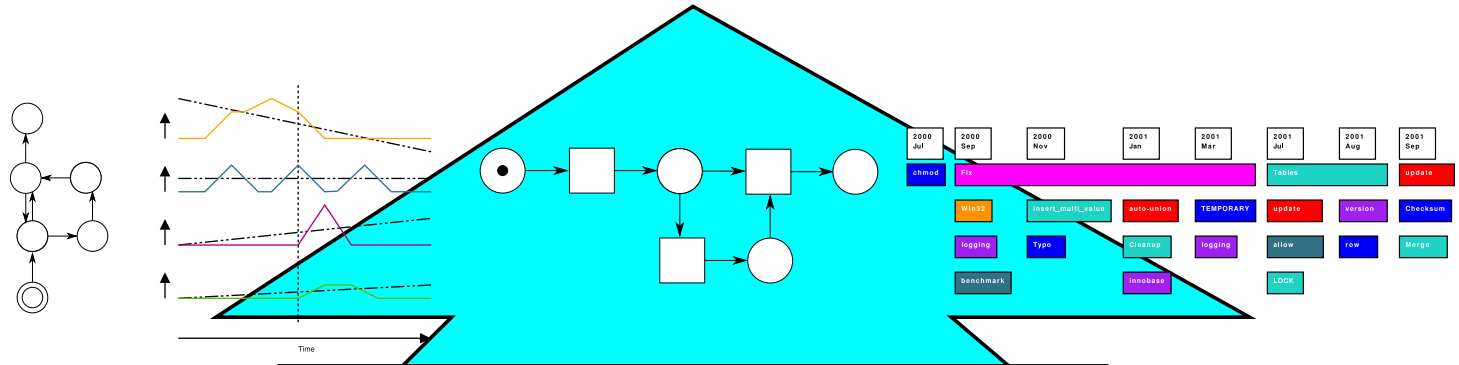


Fine Grained

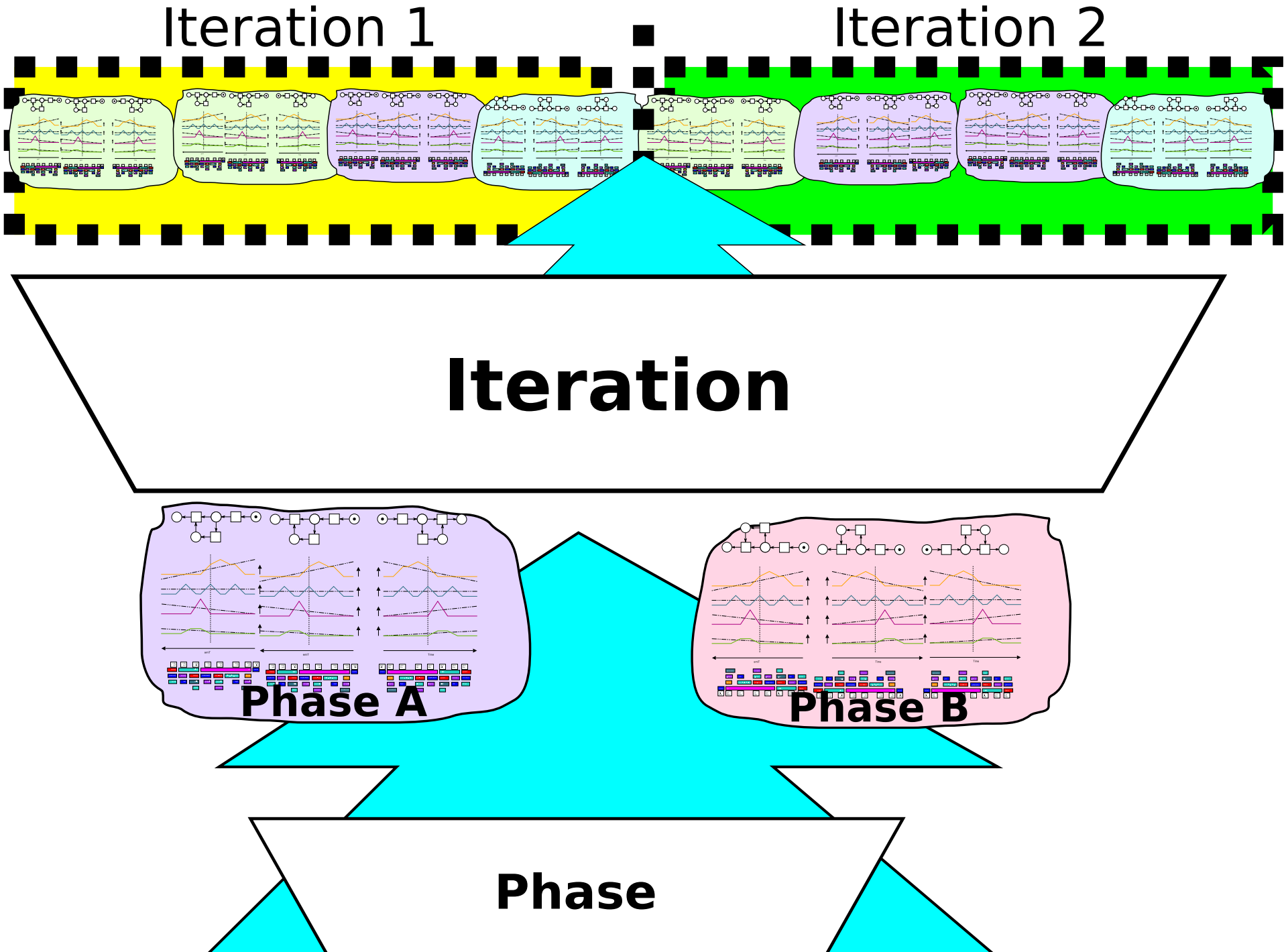




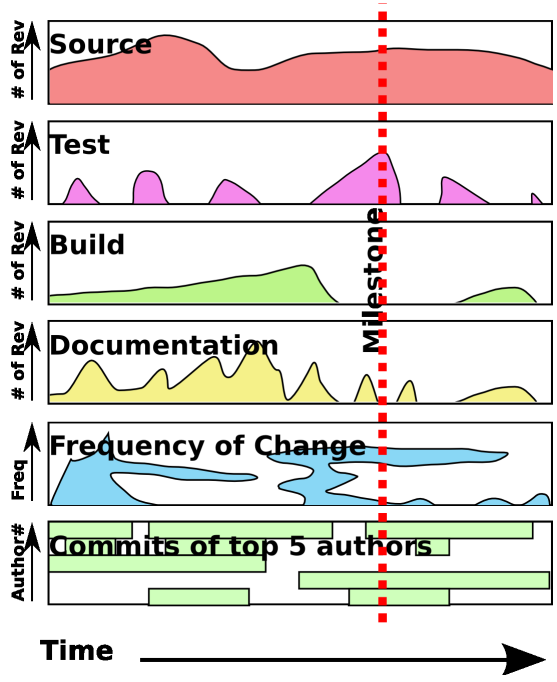
Phase

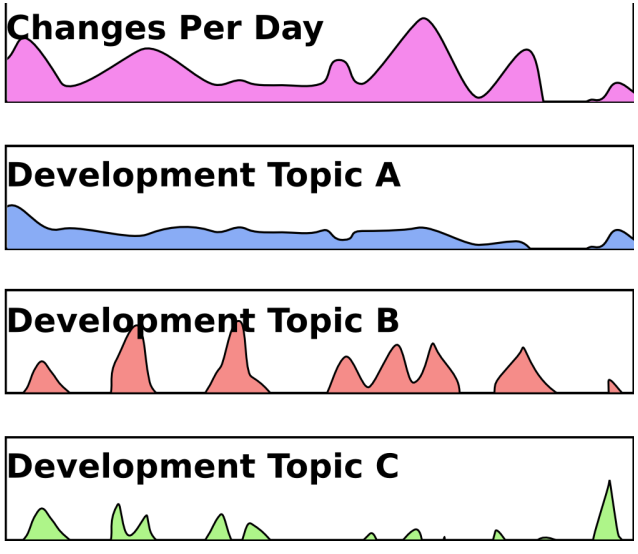


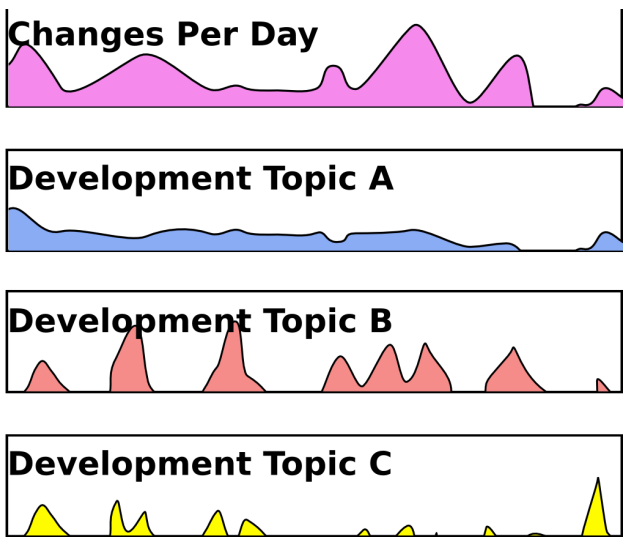
Aggregate

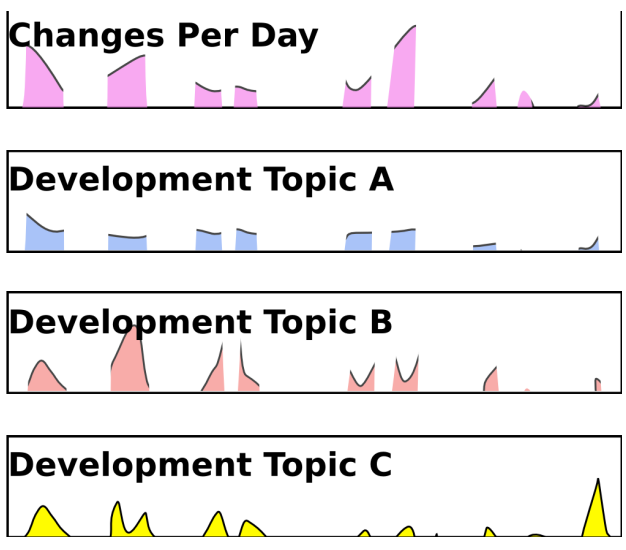

Iteration 1
Iteration 2
Iteration
Phase A
Phase B
Phase

Time Correlation / Slicing

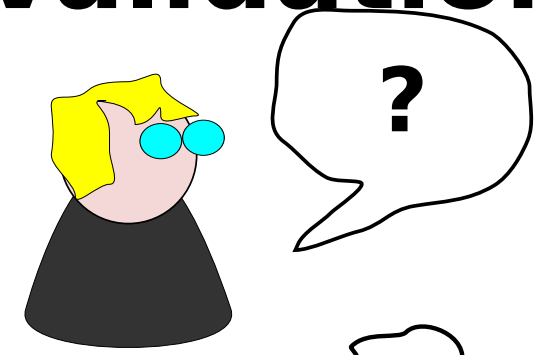


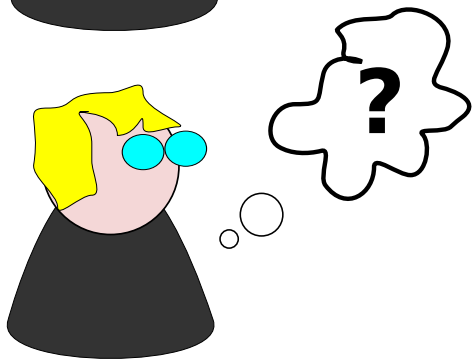


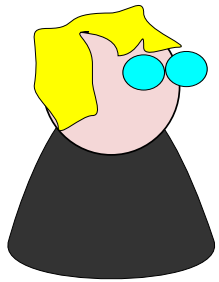


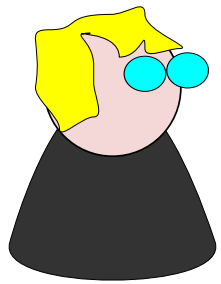


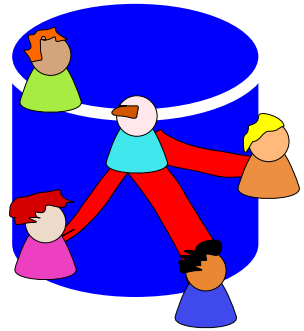
Validation

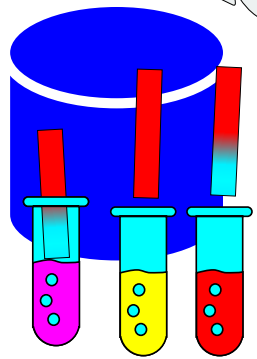


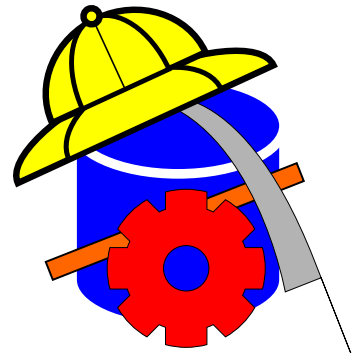


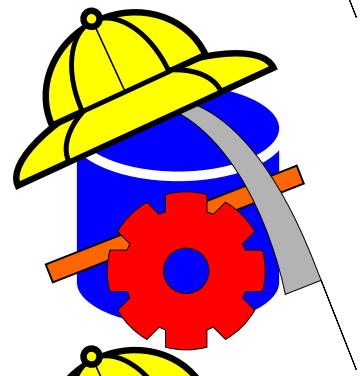


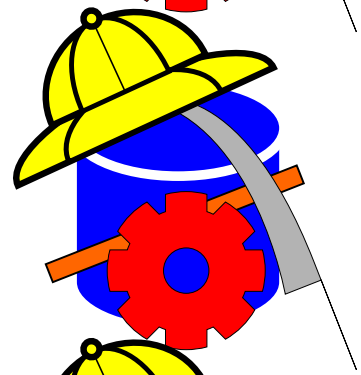


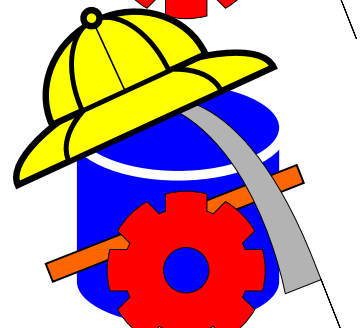




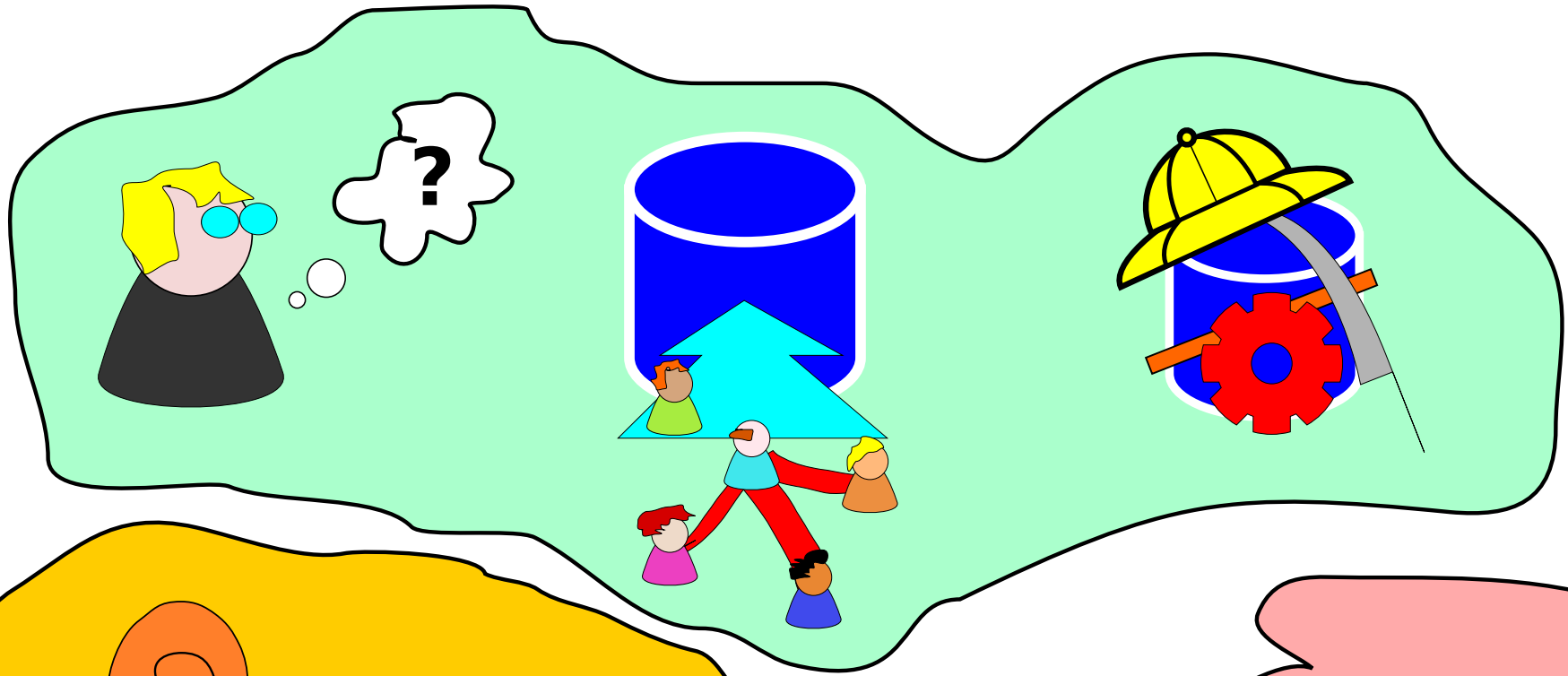


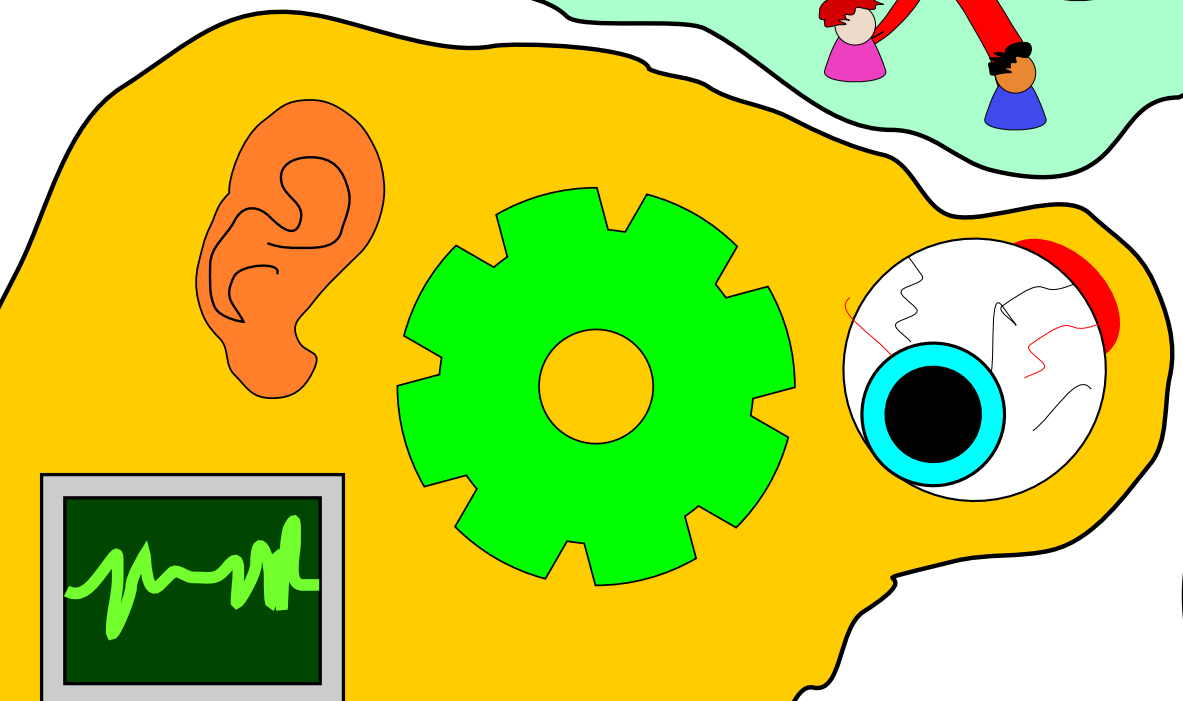


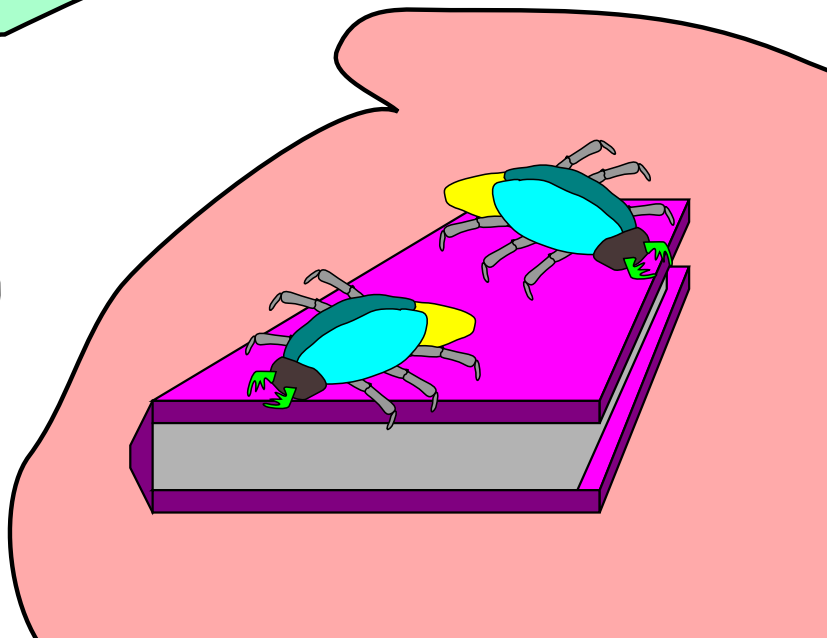




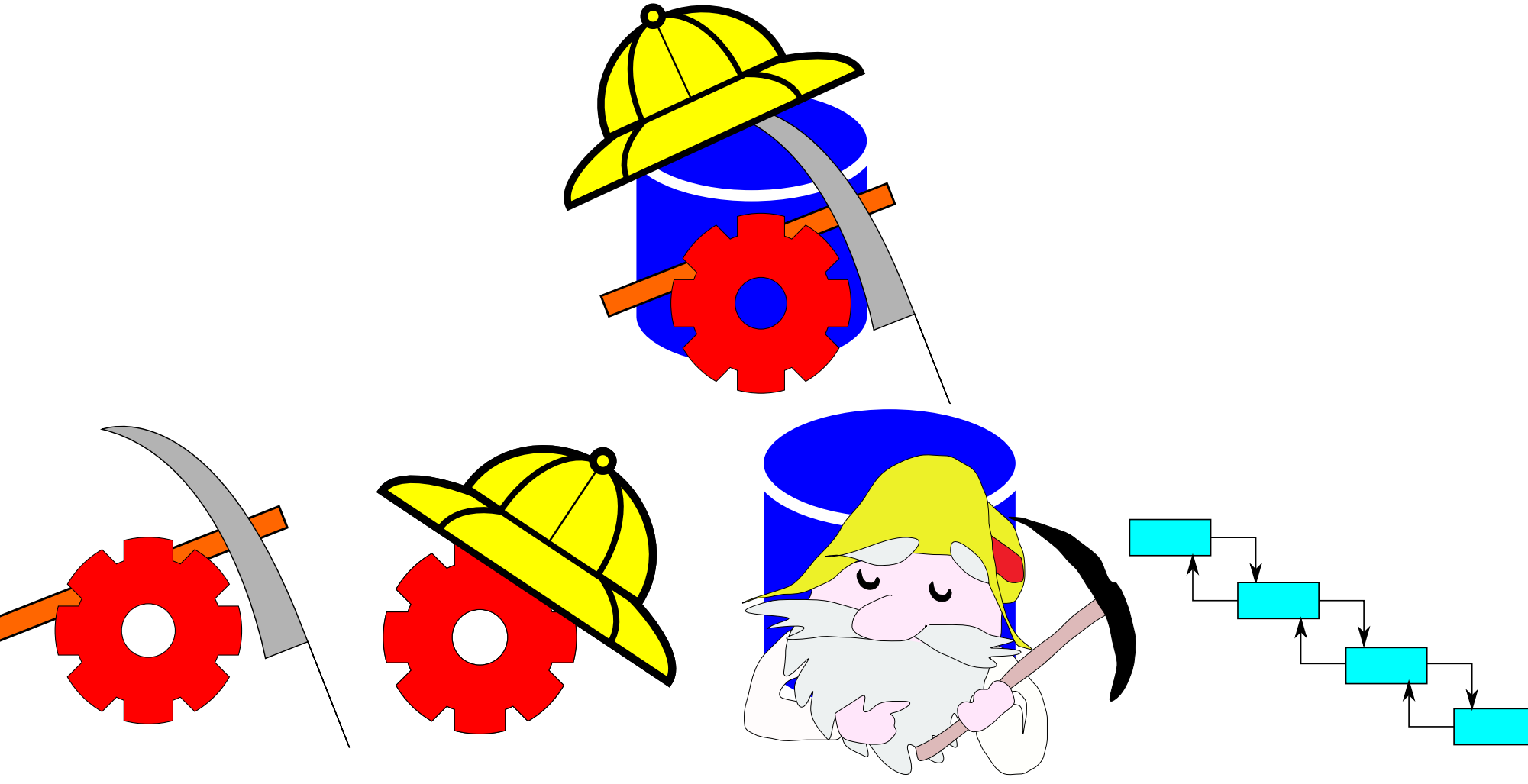
Assumptions







Conclusions

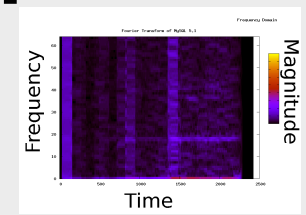


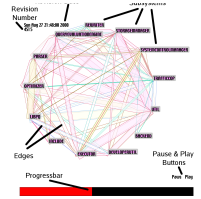
Research Timeline

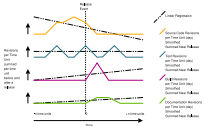
The Past

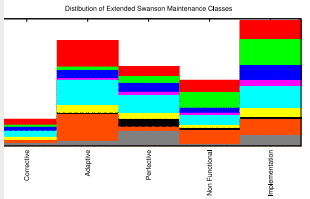
Published

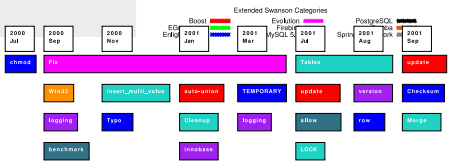
- YARN - visualization [VISSOFT 2007]
- Release Discovery (MLM) [MSR & ICSM 2007]
- evolution metrics [ICPC/SCAM 2008, SSP]
- study of large changes [MSR 2008]
- change classification [ICPC 2009]
- recurrent behaviour [ICSE 2009]
- topic analysis [ICSM 2009]

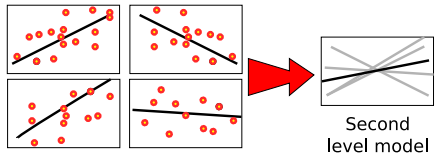


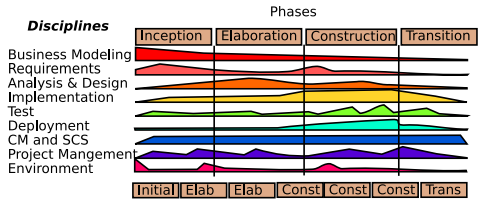


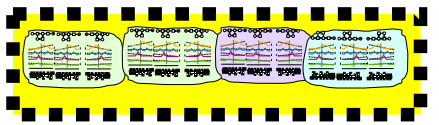












Present and Future

In progress

- finish Topic Naming paper [writeup]
- finished MLM journal paper [casestudy]
- slicing [writeup]
- multi-timeline correlation [some imp]

To do

- Unified process diagram summary [imp]
- phase & iteration identification [imp]
- Thesis