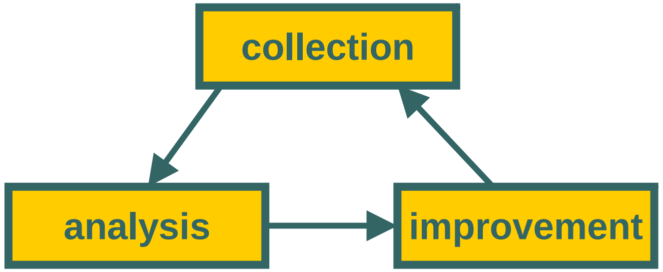
Empirical Project Monitor and Results from 100 OSS Development Projects

Masao Ohira

Empirical Software Engineering Research Laboratory,
Nara Institute of Science and Technology

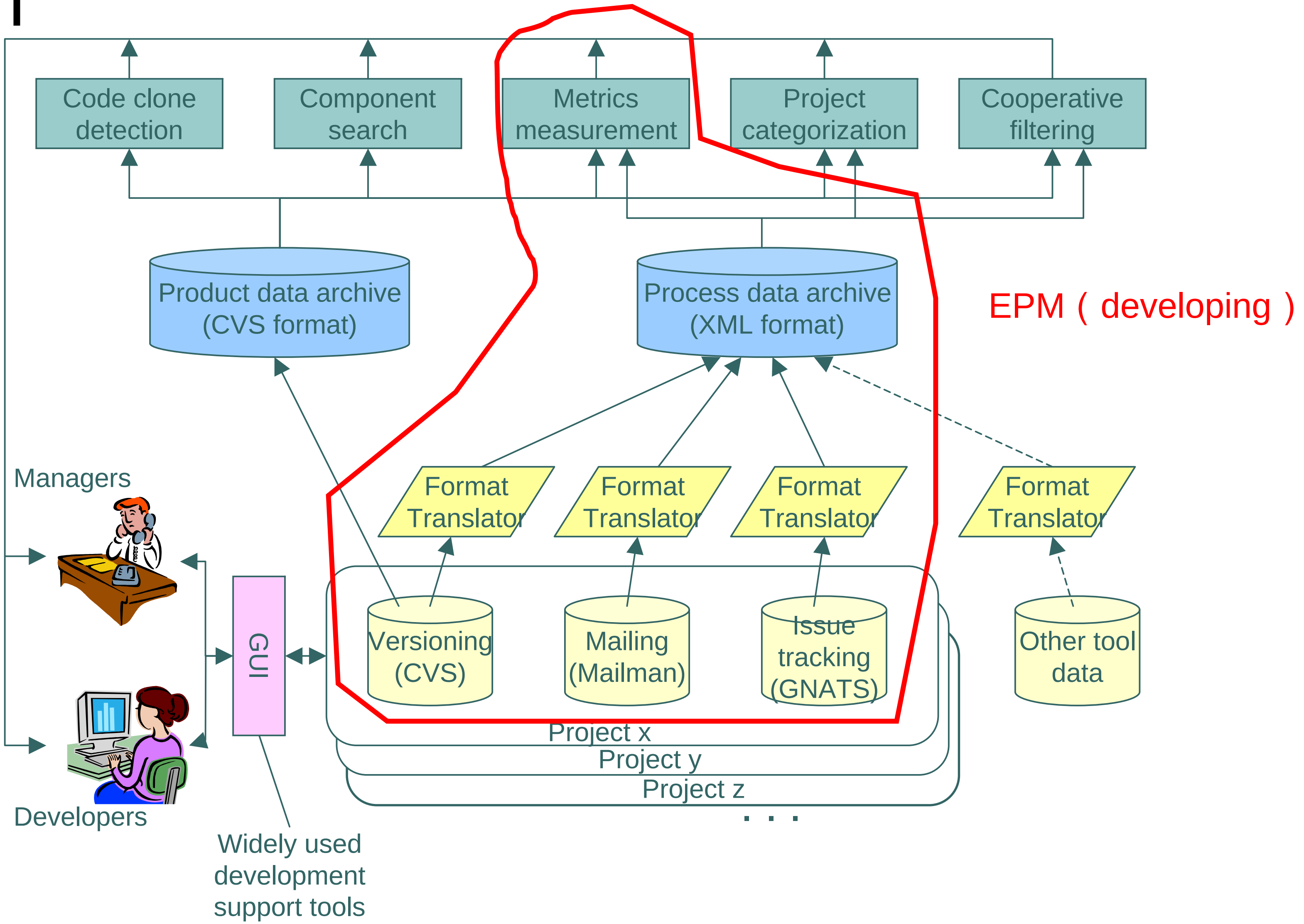
ohira@empirical.jp

EASE Project



- Empirical software development environment for tens of thousands of projects
  - Massive data collection
  - Intensive data analysis
  - Feedback for software process improvement in organizations/communities (not only a single developer/project)

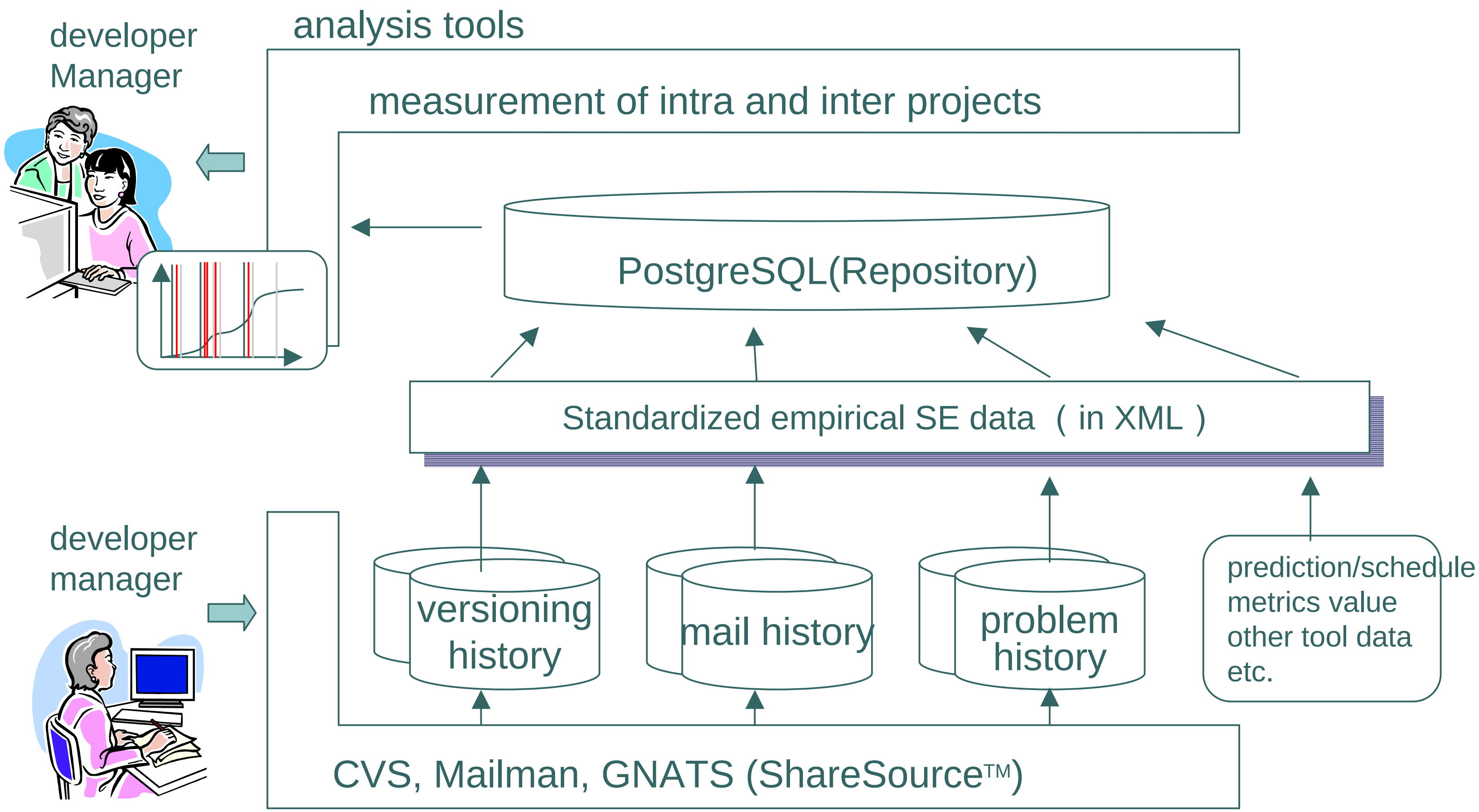
Empirical Environment



EPM: Empirical Project Monitor

- A partial implementation of Empirical Environment
- Collect, measure, and show various data for project control
- Data source from tools used in software development
  - Versioning system (e.g. CVS)
  - Mailing list manager (e.g. Mailman)
  - Issue tracking tool (e.g. GNATS)

Architecture of EPM



developer Manager

analysis tools

measurement of intra and inter projects

PostgreSQL(Repository)

Standardized empirical SE data （ in XML ）

developer manager

versioning history

mail history

problem history

prediction/schedule metrics value other tool data etc.

CVS, Mailman, GNATS (ShareSource™)

Characteristics of EPM

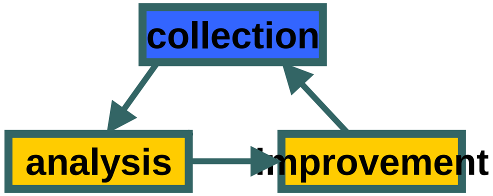
- Use open source development tools

  → Easy to introduce
- Small overhead of data collection
  - Most data from versioning history
  - Communication through e-mail, and recoding issues by tracking tool
- Easy to transform other data format to the standardized empirical SE data format

Application Area of EPM

- Large project
  - Share project status immediately
  - Reduce project management load
  - Reduce risk for tampering data
- Small project
  - Apply with small cost
  - Apply to various projects, including XP and distributed development

Data collection from OSS Development Projects


collection

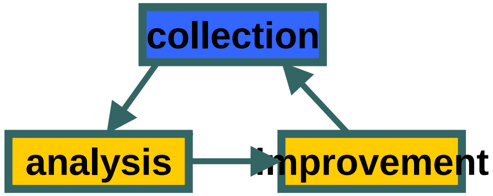
analysis

improvement

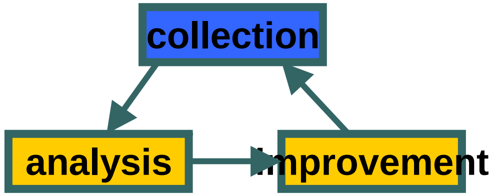
- SourceForge.net
  - hosted projects: 72,853 (Dec. 15)
  - registered Users: 753,428 (Dec. 15)
- A variety of collaboration tools
  - SourceForge Collaborative Development System (CDS) web tools
  - Project Web Server
  - Tracker: Tools for Managing Support
  - Mailing lists and discussion forums
  - MySQL Database Services
  - Project CVS Services
  - etc.

Available data source for EPM

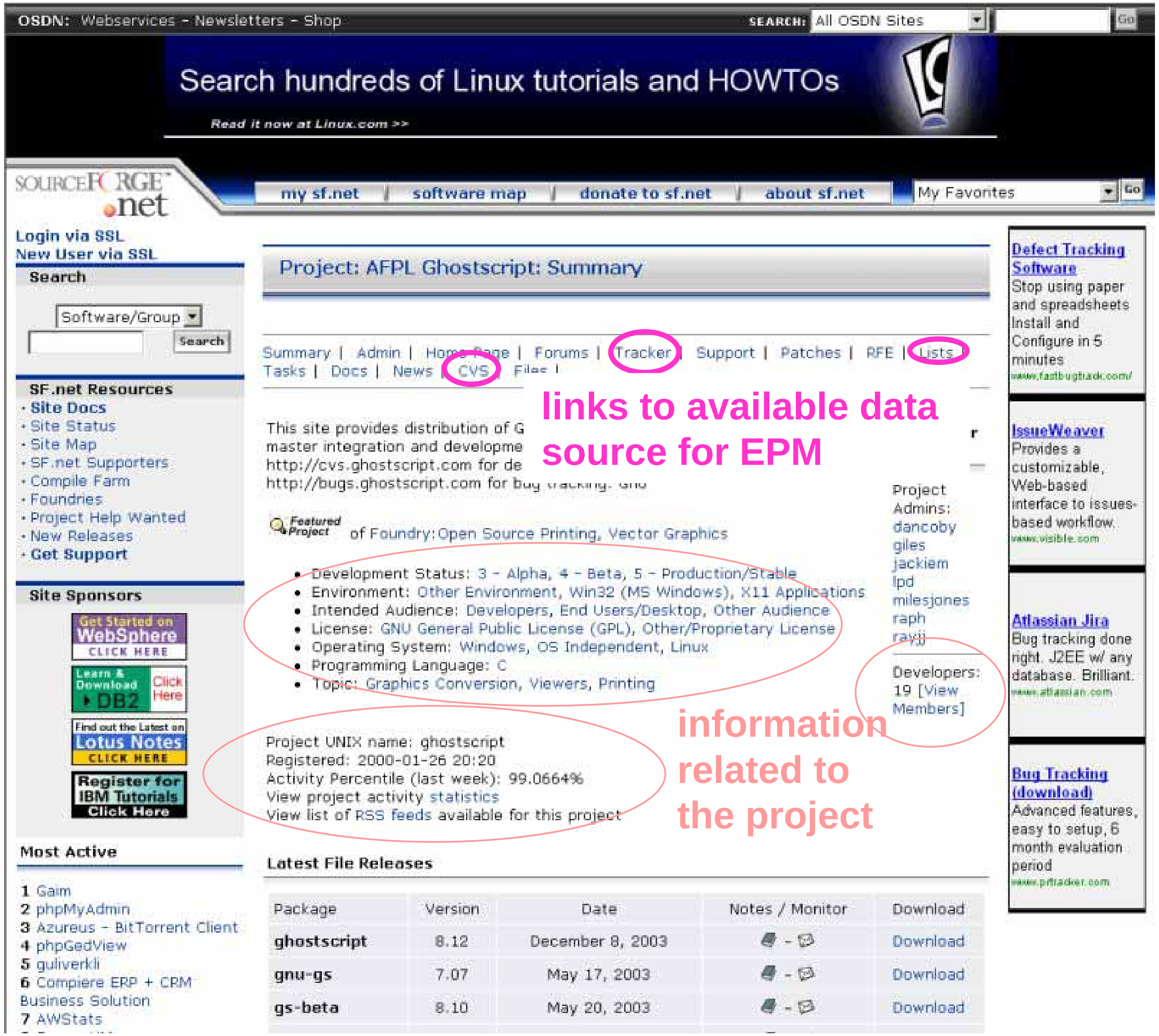
Overview of Collected Data



- 100 Active projects @ SF.net
  - Data sources for EPM
    - CVS data (only 40 projects)
    - Mailing Lists data
    - Issue (Bug) reports data
  - Project info. in a summary page
    - number of developers
    - period of a project
    - development status
    - intended audience
    - programming language
    - number of bugs
    - number of CVS commits
    - etc.



SourceForge.net



OSDN: Webservices - Newsletters - Shop   SEARCH: All OSDN Sites Go

Search hundreds of Linux tutorials and HOWTOs
Read it now at Linux.com >>

my sf.net | software map | donate to sf.net | about sf.net   My Favorites Go

Login via SSL
New User via SSL

Search

Software/Group Search

SF.net Resources

- **Site Docs**
- Site Status
- Site Map
- SF.net Supporters
- Compile Farm
- Foundries
- Project Help Wanted
- New Releases
- **Get Support**

Site Sponsors

Get Started on WebSphere CLICK HERE
Learn & Download DB2 Click Here
Find out the Latest on Lotus Notes CLICK HERE
Register for IBM Tutorials Click Here

Most Active

1 Gaim
2 phpMyAdmin
3 Azureus - BitTorrent Client
4 phpGedView
5 guliverkli
6 Compiere ERP + CRM Business Solution
7 AWStats

Project: AFPL Ghostscript: Summary

Summary | Admin | Home Page | Forums | Tracker | Support | Patches | RFE | Lists |
Tasks | Docs | News | CVS | Files |

links to available data source for EPM

This site provides distribution of G
master integration and developme
http://cvs.ghostscript.com for de
http://bugs.ghostscript.com for bug tracking. Gnu

Featured Project of Foundry: Open Source Printing, Vector Graphics

- Development Status: 3 - Alpha, 4 - Beta, 5 - Production/Stable
- Environment: Other Environment, Win32 (MS Windows), X11 Applications
- Intended Audience: Developers, End Users/Desktop, Other Audience
- License: GNU General Public License (GPL), Other/Proprietary License
- Operating System: Windows, OS Independent, Linux
- Programming Language: C
- Topic: Graphics Conversion, Viewers, Printing

Project Admins:
dancoby
giles
jackiem
lpd
milesjones
raph
rayjj

Developers:
19 [View Members]

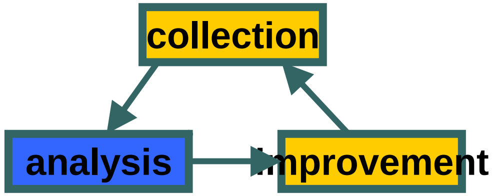
Project UNIX name: ghostscript
Registered: 2000-01-26 20:20
Activity Percentile (last week): 99.0664%
View project activity statistics
View list of RSS feeds available for this project

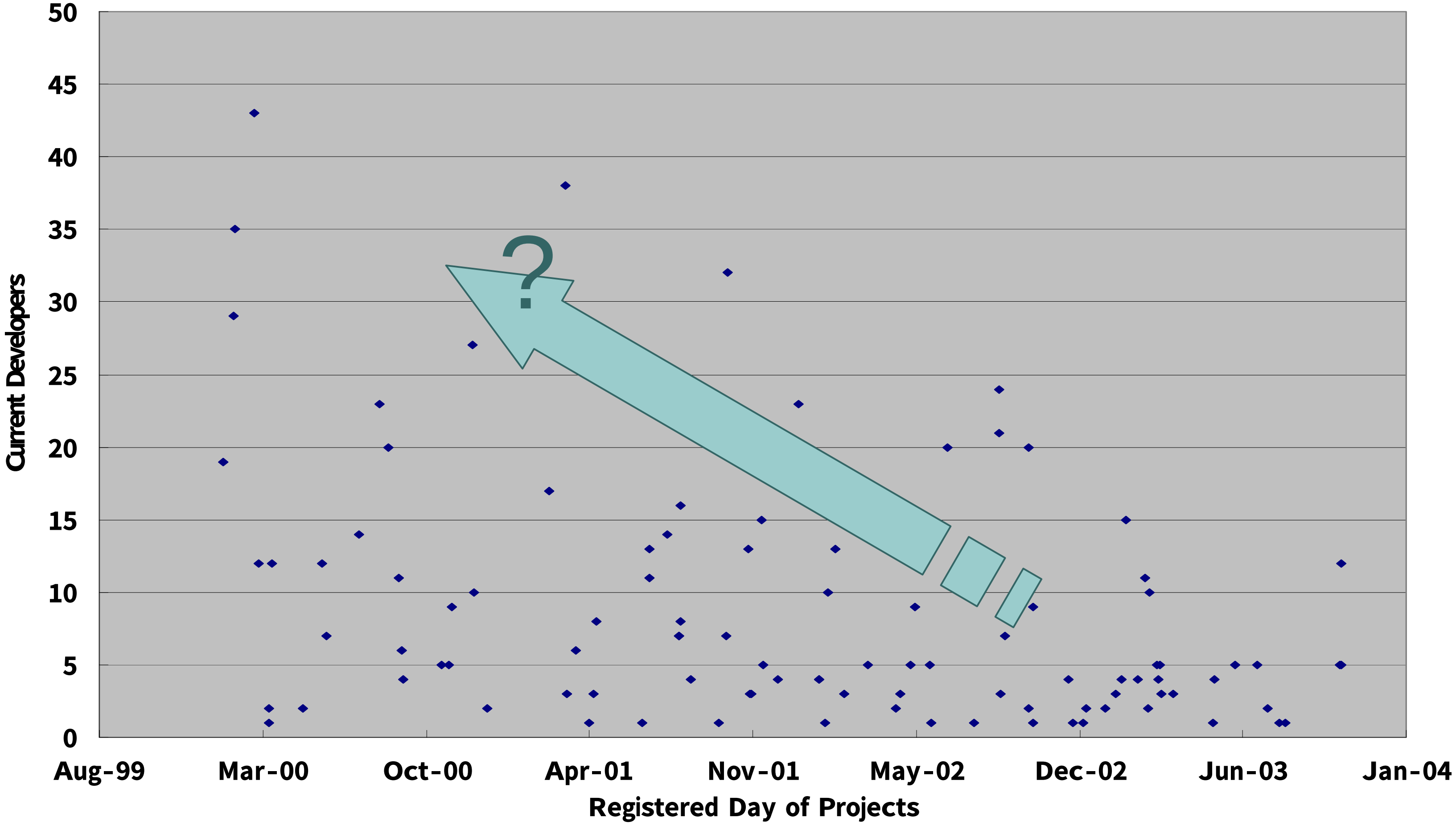
information related to the project

Latest File Releases

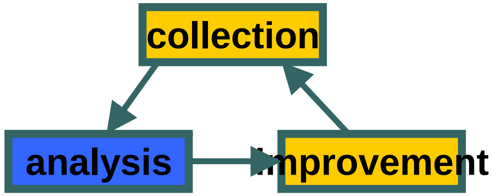
Package	Version	Date	Notes / Monitor	Download
ghostscript	8.12	December 8, 2003	-	Download
gnu-gs	7.07	May 17, 2003	-	Download
gs-beta	8.10	May 20, 2003	-	Download

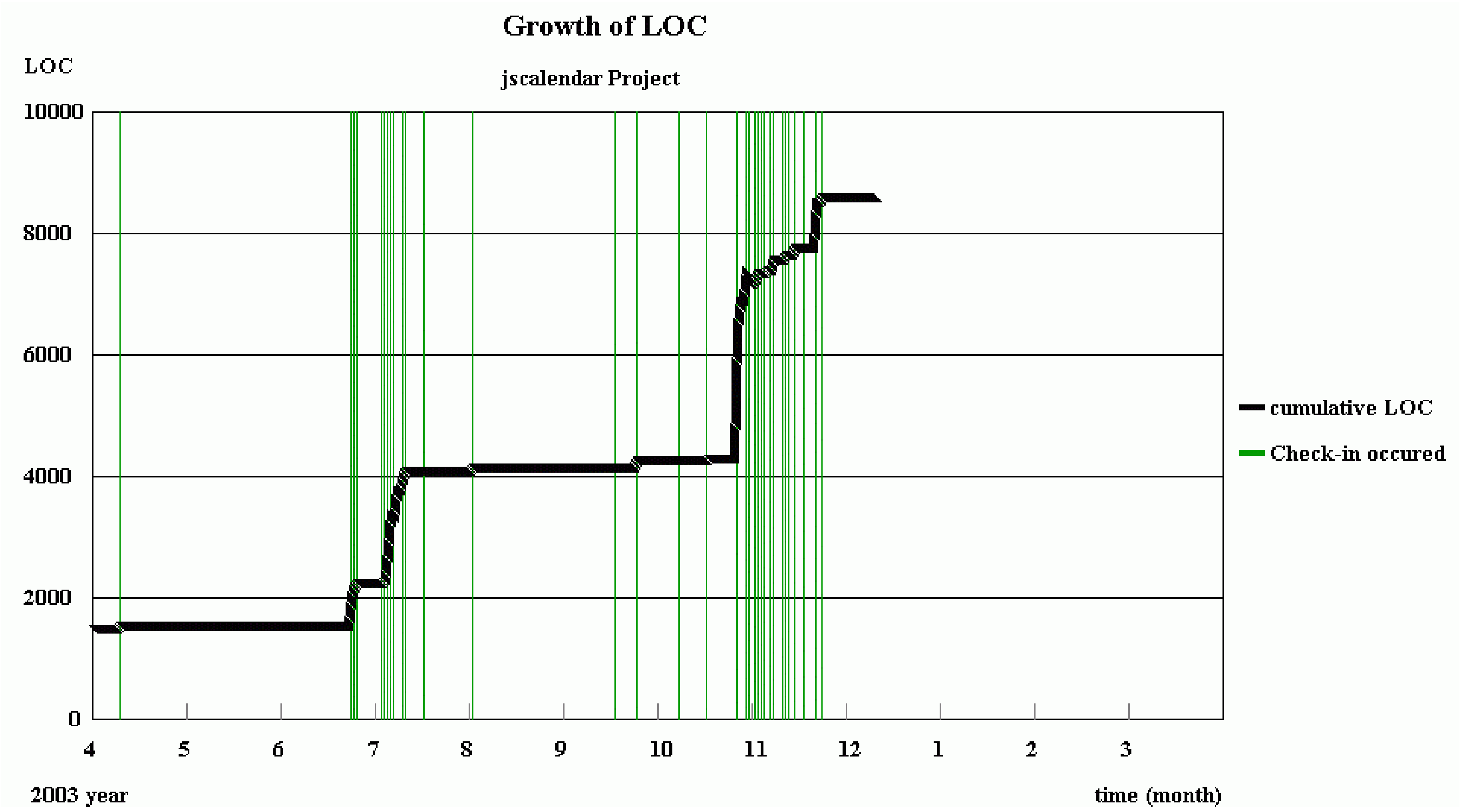
Summary of 100 OSS projects@SF.net: Evolution?



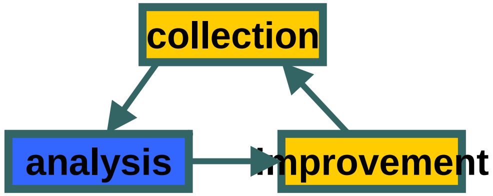


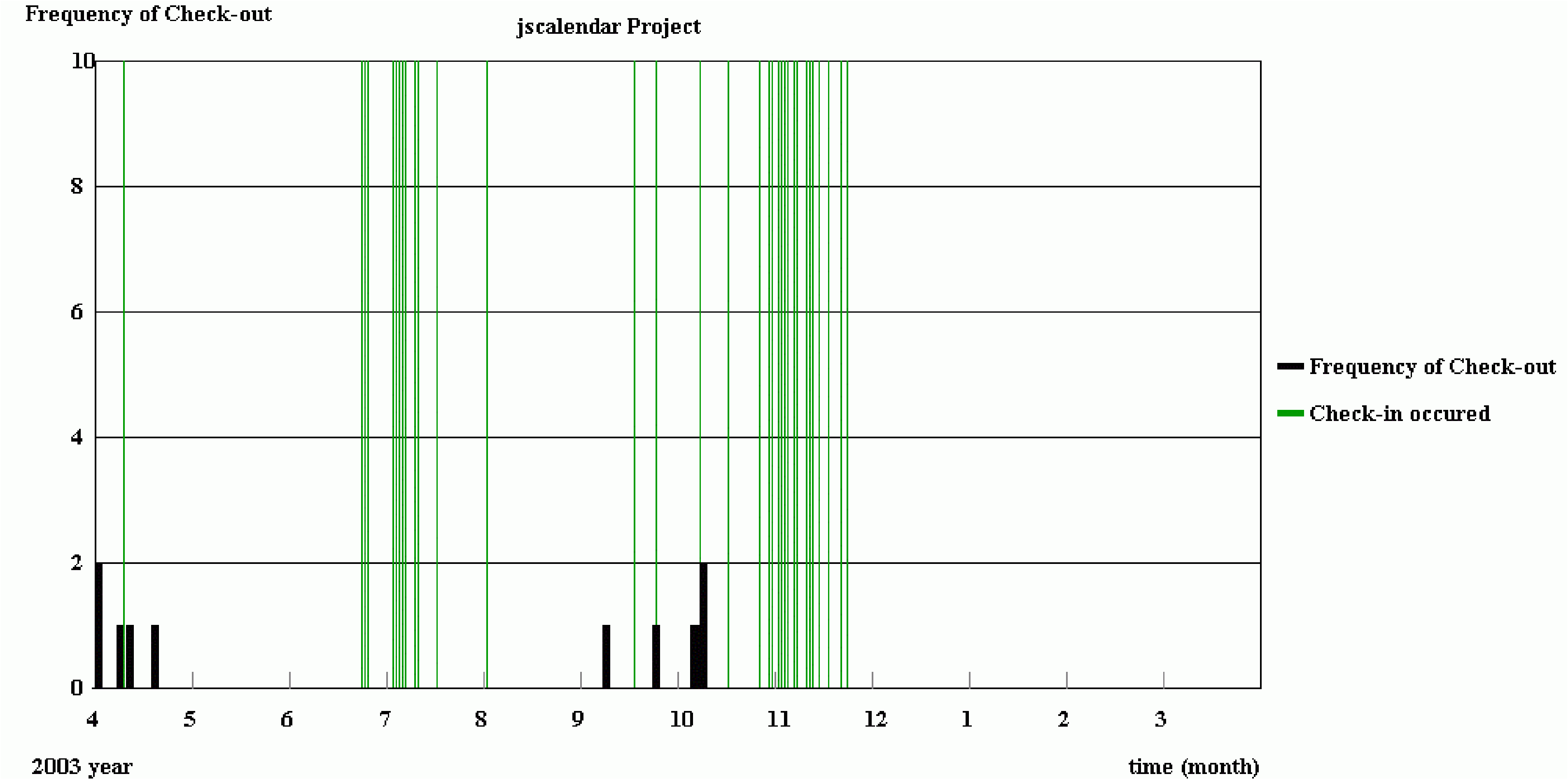
Result of CVS Product Data: Lines of Code (history of software growth)



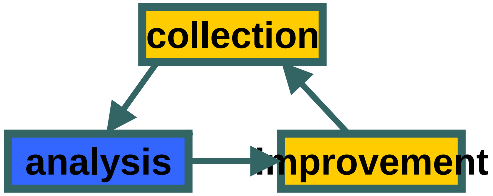


Result of CVS Process Data: Check in/out (history of developer’s activities)



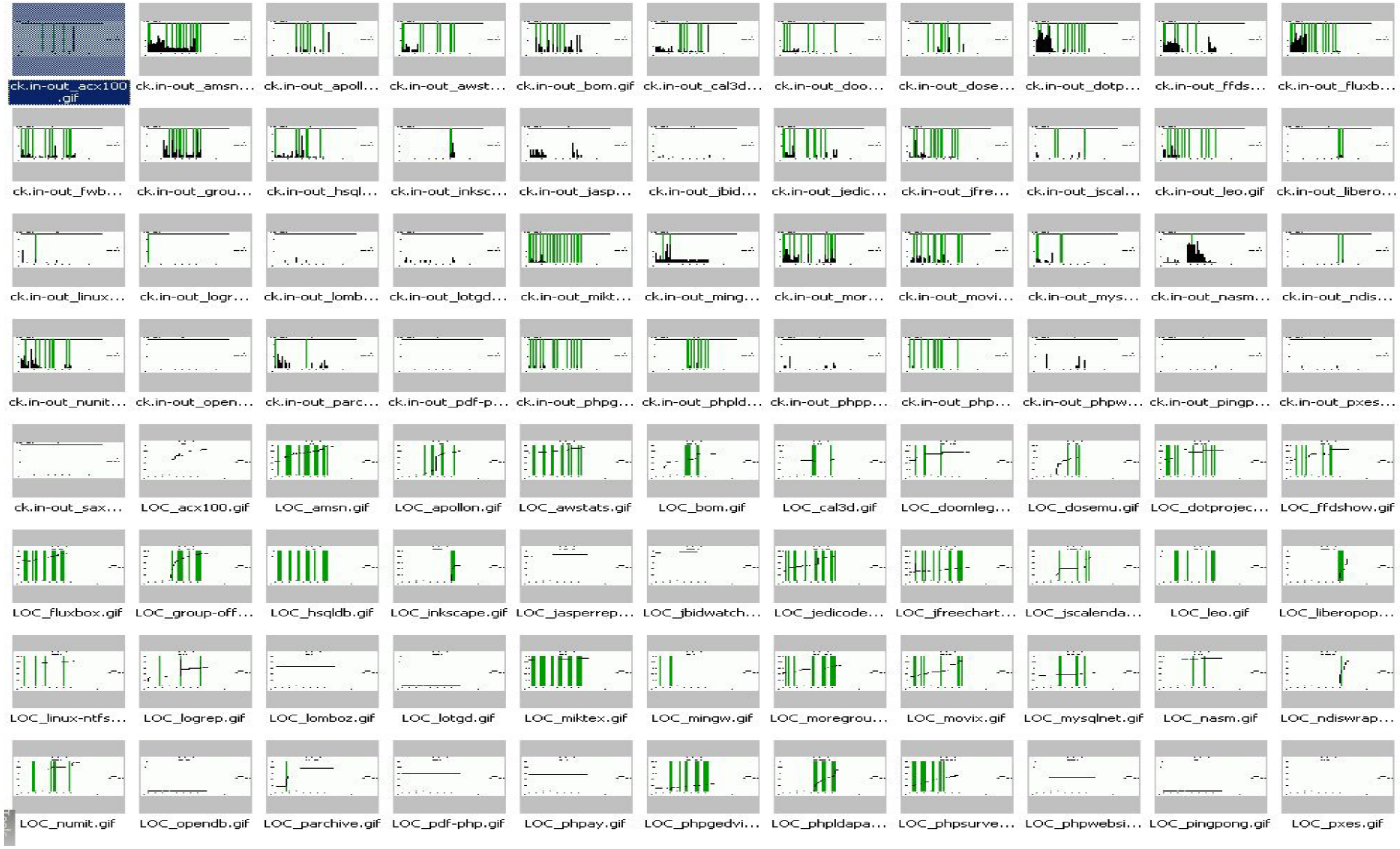


How can we use such a lot of data?


collection

analysis

improvement


ck.in-out_acx100.gif ck.in-out_amsn... ck.in-out_apoll... ck.in-out_awst... ck.in-out_bom.gif ck.in-out_cal3d... ck.in-out_doo... ck.in-out_dose... ck.in-out_dotp... ck.in-out_ffds... ck.in-out_fluxb...

ck.in-out_fwb... ck.in-out_grou... ck.in-out_hsql... ck.in-out_inksc... ck.in-out_jasp... ck.in-out_jbid... ck.in-out_jedic... ck.in-out_jfre... ck.in-out_jscal... ck.in-out_leo.gif ck.in-out_libero...

ck.in-out_linux... ck.in-out_logr... ck.in-out_lomb... ck.in-out_lotgd... ck.in-out_mikt... ck.in-out_ming... ck.in-out_mor... ck.in-out_movi... ck.in-out_mys... ck.in-out_nasm... ck.in-out_ndis...

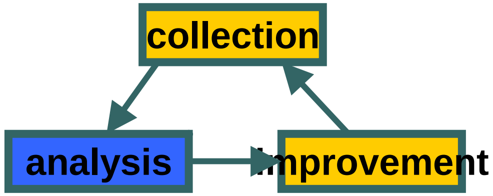
ck.in-out_nunit... ck.in-out_open... ck.in-out_parc... ck.in-out_pdf-p... ck.in-out_phpg... ck.in-out_phpld... ck.in-out_phpp... ck.in-out_php... ck.in-out_phpw... ck.in-out_pingp... ck.in-out_pxes...

ck.in-out_sax... LOC_acx100.gif LOC_amsn.gif LOC_apollon.gif LOC_awstats.gif LOC_bom.gif LOC_cal3d.gif LOC_doomleg... LOC_dosemu.gif LOC_dotprojec... LOC_ffdshow.gif

LOC_fluxbox.gif LOC_group-off... LOC_hsqldb.gif LOC_inkscape.gif LOC_jasperrep... LOC_jbidwatch... LOC_jedicode... LOC_jfreechart... LOC_jscalenda... LOC_leo.gif LOC_liberopop...

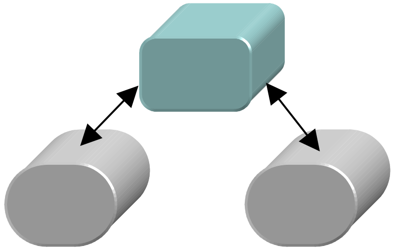
LOC_linux-ntfs... LOC_logrep.gif LOC_lomboz.gif LOC_lotgd.gif LOC_miktex.gif LOC_mingw.gif LOC_moregrou... LOC_movix.gif LOC_mysqlnet.gif LOC_nasm.gif LOC_ndiswrap...

LOC_numit.gif LOC_opendb.gif LOC_parchive.gif LOC_pdf-php.gif LOC_phpay.gif LOC_phpgedvi... LOC_phpldapa... LOC_phpsurve... LOC_phpwebsi... LOC_pingpong.gif LOC_pxes.gif



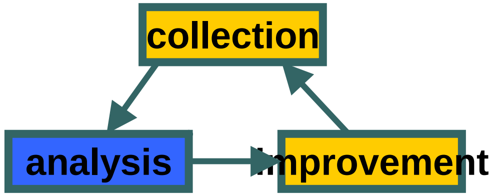
Gross Classification using EVIDII

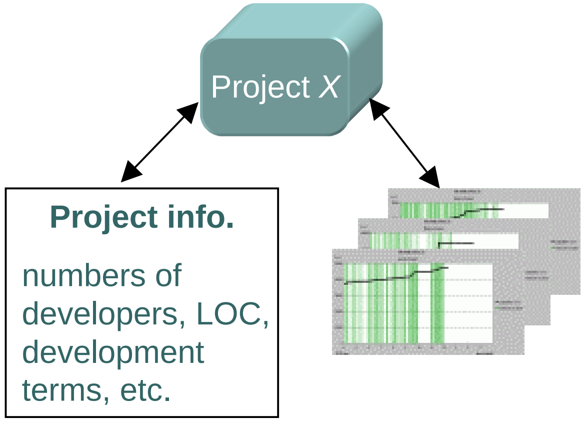
- EVIDII: Interactive interfaces that visualize relationships among three sets of data



(original application domain: face-to-face communication support between clients and designers)

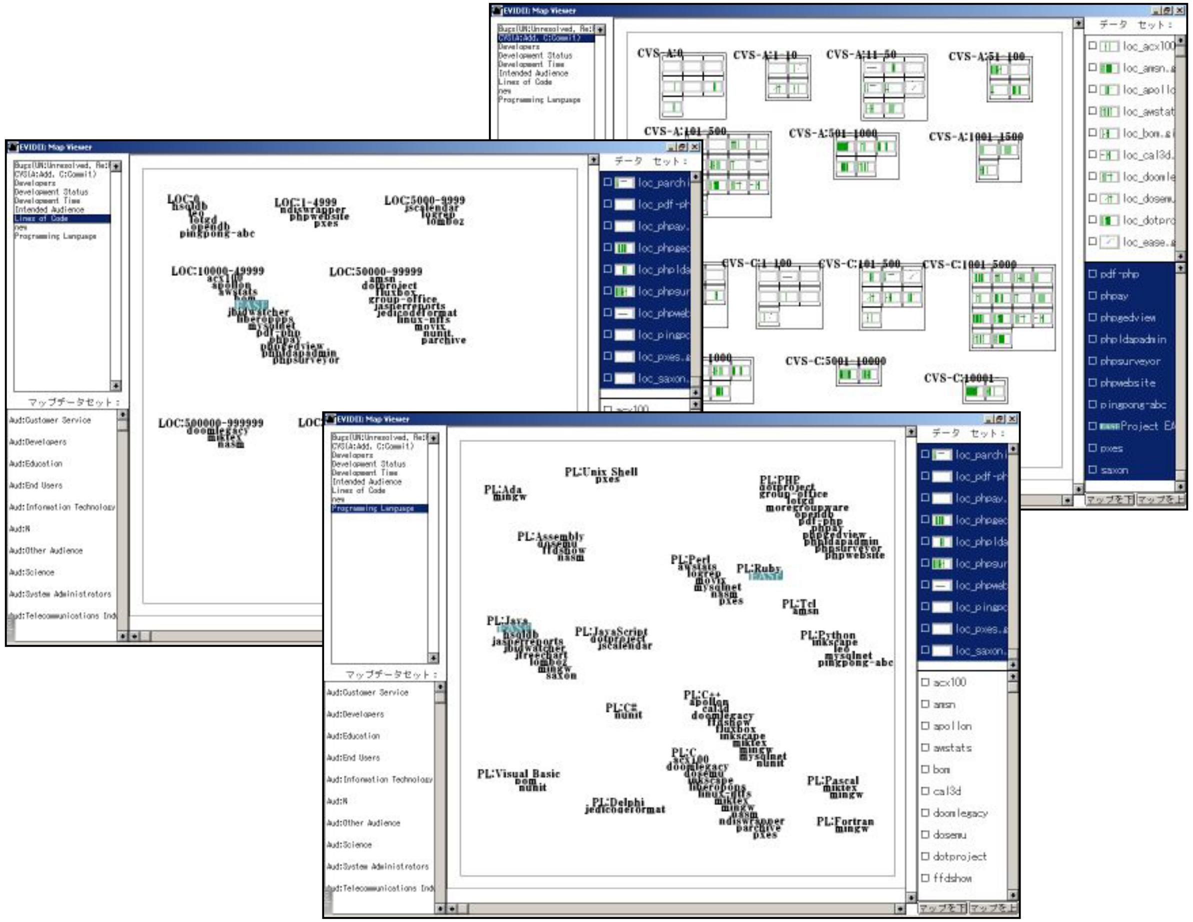
Demo: organizing dynamic community?



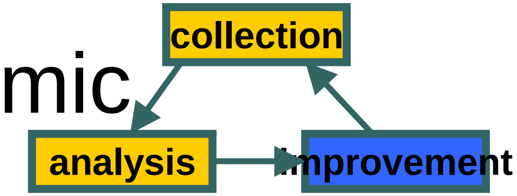


Project info.

numbers of developers, LOC, development terms, etc.



Scenario: organizing a dynamic community / providing feedback for improvement



1. Comparing other projects with a target project
2. Finding similarities and differences between them

DynC approach

EASE approach

3-a. Notifying to related project leaders of the existence of communities

4-a. Asking them help/ advices for improvement

3-b. Identifying factors of the similarities and differences

4-b. Providing suggestions for improvement

Summary and Future Work

- EPM: Empirical Project Monitor
- Data Collection from 100 OSS projects (only 40 CVS data…)
- Two scenarios using EVIDII

- More data collection (mails and bug issues) and analysis using EPM/EVIDII