



## Empirical Project Monitor: Automatic Data Collection and Analysis toward Software Process Improvement

大平 雅雄<sup>1</sup> 横森 励士<sup>2</sup> 阪井 誠<sup>3</sup>  
松本 健一<sup>1</sup> 井上 克郎<sup>2</sup> 鳥居 宏次<sup>1</sup>

<sup>1</sup>奈良先端科学技術大学院大学  
<sup>2</sup>大阪大学  
<sup>3</sup>株式会社SRA先端技術研究所

### 発表の構成

- プロセス改善活動の現状と実践における問題点
- 定量的データ収集に基づくプロセス改善のためのアプローチ
  - 基本モデル(データ収集・分析・フィードバック)
  - 支援環境の設計方針
  - エンピリカルソフトウェア工学環境 (ESEE: Empirical Software Engineering Environment)
- 開発データ自動収集・分析支援システム Empirical Project Monitor (EPM)
  - 開発データの自動収集
  - 分析結果の可視化

2004/2/25

DSW04

2

### 研究の背景

- ソフトウェアの信頼性・生産性の向上を目的としたソフトウェアプロセス改善への期待
  - ソフトウェアの信頼性
    - 多数のバグを含んだソフトの流通
    - 一度ダウンすると多大な社会的損失
  - ソフトウェアの生産性
    - 開発期間の短縮要請
    - 人海戦術による限界
- ソフトウェア開発に関するデータの収集・計測・分析が必須の課題

2004/2/25

DSW04

3

### 問題点

- ソフトウェア計測に関する諸技術は多数提案されている
- 計測に関する十分な経験がなければ、一貫性のある定量的なデータを収集し計測するのが困難(経験的なノウハウやアドホックな手法、ツールが未だに多く使われている原因)
  - 改善目標の設定
  - 必要なソフトウェア・メトリクスの準備
  - 次の改善へ向けた分析結果の有効利用
- GQMなどの計測技法を利用することは可能ではあるが...
  - あらかじめ改善目標を明示的に設定することが困難な場合も多い
  - 計測にかかるコストが大きい

2004/2/25

DSW04

4

### 定量的データ収集に基づくプロセス改善のためのアプローチ

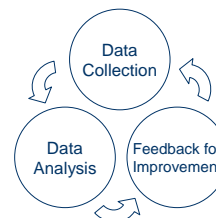
- 一貫性のある定量的開発データの収集と計測
  - 個人の経験に依存しない定量的な計測
  - 特別な作業負荷やコストを必要としない計測
- 多様なプロジェクトおよび組織のプロセス改善目的に対応可能な、定量的データ収集基盤の提供
  - エンピリカルソフトウェア工学環境  
ESEE: Empirical Software Engineering Environment
  - 産学連携による実際の開発現場のデータの分析
  - オープンソース開発プロジェクトのデータの分析

2004/2/25

DSW04

5

### プロセス改善のための基本モデル



- 大規模データ収集
- 集約的データ分析
- プロセス改善のためのフィードバック

厳格な手続きを踏むのではなく、プロジェクトおよび組織の目的に沿った柔軟なプロセス改善実施のためのモデル

2004/2/25

DSW04

6

## 方針: データ収集

- 改善目標の設定優先ではなく、大規模データ収集を優先
  - 理想的には改善のための目標設定が重要であるが、あらかじめ明確な改善目標を設定すること自体が困難な場合が少なくないため
- プロダクトデータを中心に収集
  - データ収集のための管理者・開発者の余分な負担を最小限に抑えるため、プロセスデータはプロダクトデータから可能な限り抽出する
- 人為的操作のない加工されていないデータを収集
  - ドキュメント等、主観的なデータが混入しやすいものを排除するため
- リアルタイムでのデータ収集
  - 現状を即座に把握し開発プロセスの管理・改善に連動させるため
- 広く利用されている開発支援システムからデータを収集
  - 開発形態に依存しない多様なプロジェクトも支援の対象とするため

2004/2/25

DSW04

7

## 方針: データ分析

- 単一のプロジェクトを対象とする  
プロセス・プロダクトメトリクス
- 複数のプロジェクト間のメトリクス
- プロジェクトの分類・進化予測
- コンポーネント・専門知識の再利用
- さらなる複雑・高度な分析へ

容易なものから  
段階的に実現



2004/2/25

DSW04

8

## 方針: フィードバック

- 目的に応じたフィードバック
  - 様々な仕掛けが必要

当面は、収集データおよび

- 単一のプロジェクトを対象とするプロセス・プロダクトメトリクス
- 複数のプロジェクト間のメトリクス

の分析結果を視覚的に表示する

ブラウザ (Empirical Project Monitor: EPM) の実装と評価

2004/2/25

DSW04

9

## エンピリカルソフトウェア工学環境

### ESEE: Empirical Software Engineering Environment

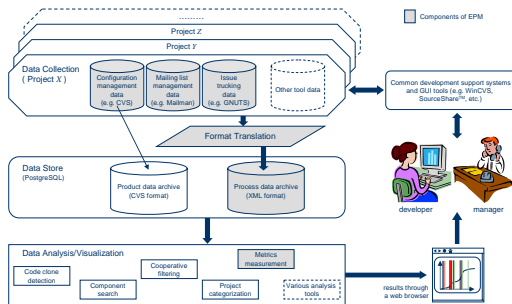
- プロセス改善のすべのステップを支援する巨大な支援環境ではなく、分析の目的・方法に沿って交換可能なプラグイン形式の柔軟なシステム
- ESEEの本質的機能
  - 数千規模のプロジェクトの開発データ収集
  - プロセス改善および組織の利益に直結する分析結果(フィードバック)の提供
- ESEEの基本構成
  - データ収集
  - フォーマット変換
  - データ蓄積
  - データ分析・可視化

2004/2/25

DSW04

10

## Framework for ESEE



2004/2/25

DSW04

11

## 開発データ自動収集・分析支援システム

### Empirical Project Monitor (EPM)

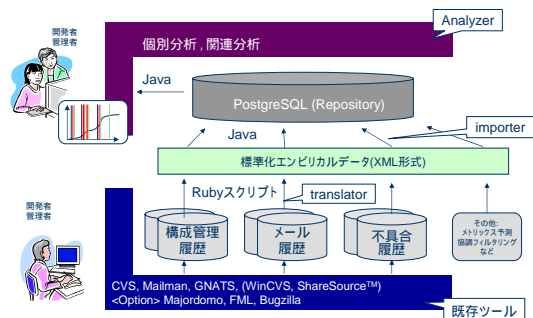
- エンピリカル環境の一部として開発中
- プロジェクトの制御に有益な開発データを収集、分析
- データ収集源
  - 構成管理ツールCVS
  - メールリスト管理ツールMailman, Majordomo, FML
  - 障害管理ツールGNATS, Bugzilla

2004/2/25

DSW04

12

## EPMのアーキテクチャ



2004/2/25

DSW04

13

## EPMの特徴

- オープンソース開発に使われているツール利用  
容易に導入可能
- データ収集の負担  
簡単な規約で適用可能
  - プロダクトデータ(例えば1日1回CVSに保存)
  - 連絡はメールで、障害は管理ツールで記録を残す
- 他のツールから標準エンピリカルSEデータ形式への  
変換容易  
他の開発環境との連携可能

2004/2/25

DSW04

14

## EPMの適用対象

- 大規模プロジェクト
  - 情報の共有がリアルタイムで可能
  - 管理の負荷が低減
  - 人為的なデータ操作が入りにくい
- 小規模プロジェクト
  - 今まで手間等の問題で管理しにくかったプロジェクトにも適用しやすい
  - いろいろなプロセス(XPなど)や組織をまたがる分散開発にも適用可能

2004/2/25

DSW04

15

## EPMの分析ツール

- 個別分析
  - ソースコード規模
  - 障害解決時間
  - 累積・未解決障害件数/平均障害滞留時間, ほか
- 関連分析
  - 更新/参照数
  - 更新・障害報告/メール投稿数
  - 更新と障害件数, ほか

2004/2/25

DSW04

16

## EPMのインターフェース

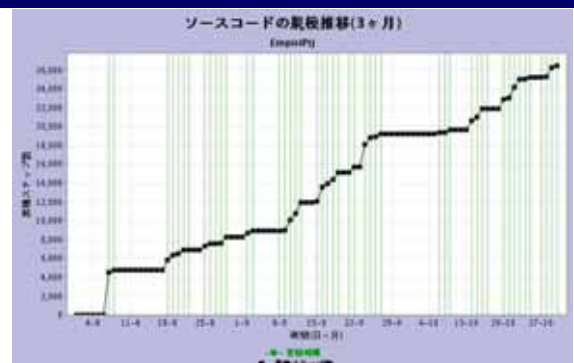


2004/2/25

DSW04

17

## ソースコードの規模推移とチェックインとの関係

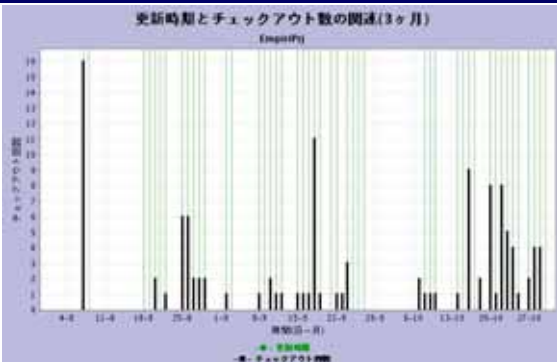


2004/2/25

DSW04

18

### チェックインとチェックアウト数との関係



2004/2/25

DSW04

19

### チェックインと累積障害件数との関係

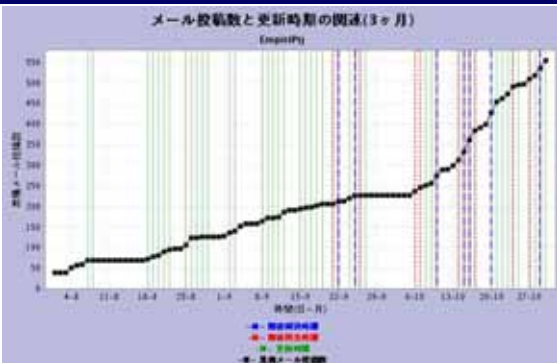


2004/2/25

DSW04

20

### 累積メール投稿数とチェックイン/障害発生/障害解決時期との関係

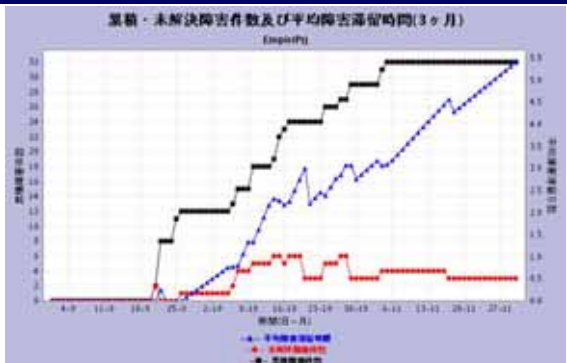


2004/2/25

DSW04

21

### 累積障害件数/未解決障害件数/平均障害滞留時間との関係



2004/2/25

DSW04

22

### CVS詳細情報

2004/2/25

DSW04

23

### メール詳細情報

2004/2/25

DSW04

24

## 適用例

- 対象プロジェクト: EPM 開発グループ
- 利用言語: Ruby, Java
- 開発者: 4 名
- 分析対象ツール: CVS, mailman, GNATS
- 開発期間: 約3 ヶ月
- 導入準備: 約1 週間 (EPM および関連するオープンソースの管理技術の習得)
  
- ツール導入の効果 (感想)
  - グラフの特徴的な部分から、なぜこの現象が生じたかを確認することでプロジェクトの把握が容易
  - 可視化によってリアルタイムでの状況把握が容易
  - プロジェクトの問題の検出
  - メール一覧から、どのような議論が行われたかが容易になった

2004/2/25

DSW04

25

## まとめ

- エンピリカルソフトウェア工学環境 (ESEE) の提案
  
- ESEEの部分的実装であるEmpirical Project Monitor (EPM)の紹介
  - 導入の障壁が低い
  - 様々な環境に適用可能
  
  - 計測をおこなっていない場合
    - 既存の開発環境と連携して手軽にプロジェクトのモニター
    - プロジェクトの異常検出・過去や予想との比較可能
    - プロジェクトの蓄積が容易
  
  - すでに計測をおこなっている場合
    - 全社的に統一的なデータをリアルタイムで描えられるようになる
      - 個人の目標との比較・部署やプロジェクトの間の比較・他社と比べて、海外と比べてどうか? など
    - エンピリカルデータの統合により、情報の共有や再利用が (個人の経験や助に頼らずに) 自動的に行えるようになる
    - 似たようなプロジェクト、似たようなソフトウェア開発での知見の発掘、重複作業の排除

2004/2/25

DSW04

26

## 今後の予定

- 現在
  - EPMのデモ版
- 今年度中
  - EPMの配布版 リリース
- 来年度
  - EPMの現場での適用, フィードバック
  - 分析ツール追加

2004/2/25

DSW04

27

文部科学省 リーディングプロジェクト「e-society基盤ソフトウェアの総合開発」  
データ収集に基づくソフトウェア開発支援システム  
EASEプロジェクト  
<http://www.empirical.jp/>