

EASEにおけるデータ分析事例

奈良先端科学技術大学院大学 情報科学研究科
文部科学省EASEプロジェクト (www.empirical.jp)

松本 健一

エンピリカルソフトウェア工学における 3つの段階

エンピリカルアプローチの3段階

第1段階: 観察 (Empirical observations) の実施

- ・実験や調査により現象(事実)を確認する.
- ・現象を**表現**できるようになる.



第2段階: 法則 (Laws) の発見

- ・現象が起こるコンテキスト(状況)を理解する.
- ・現象を**予測**できるようになる.

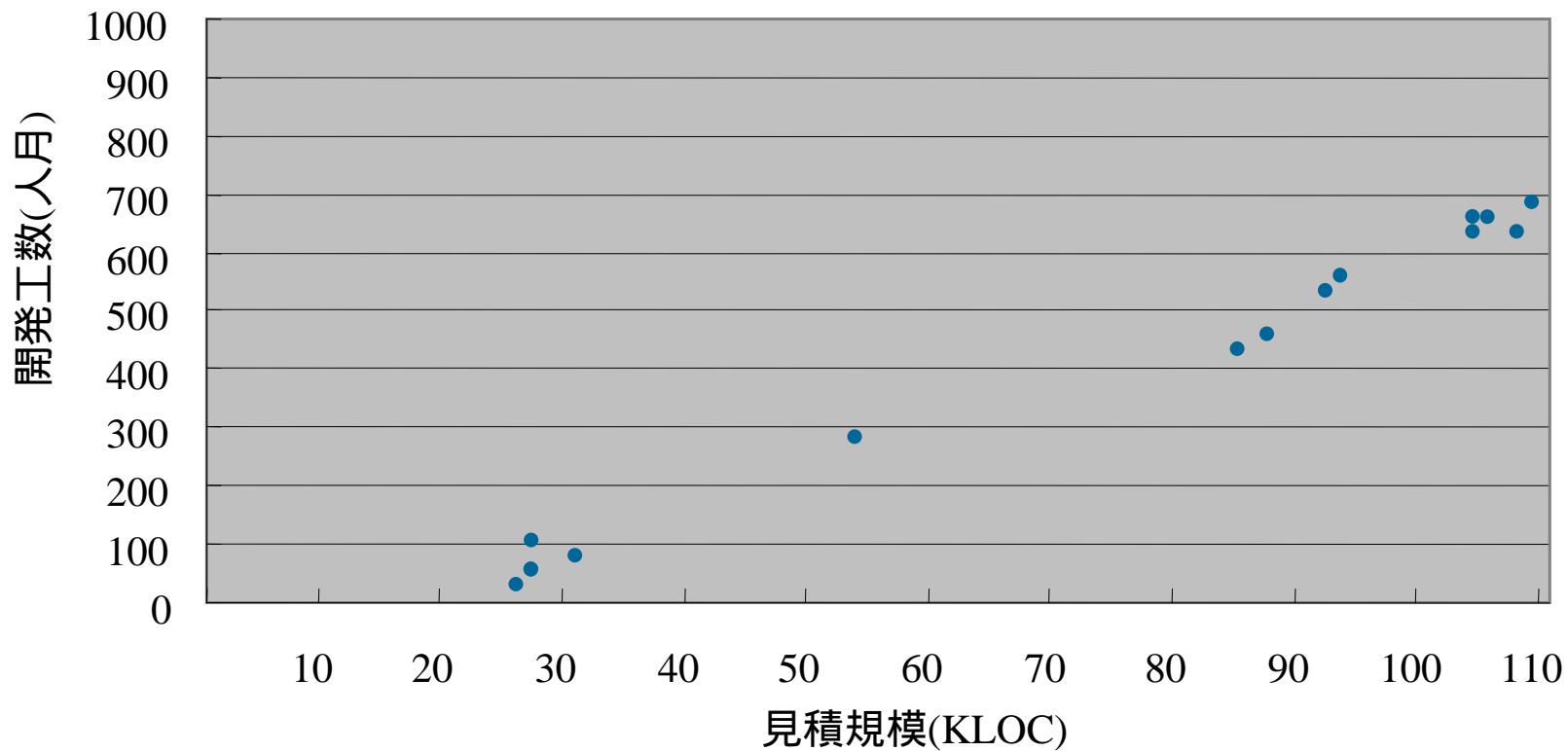


第3段階: 理論 (Theories) の確立

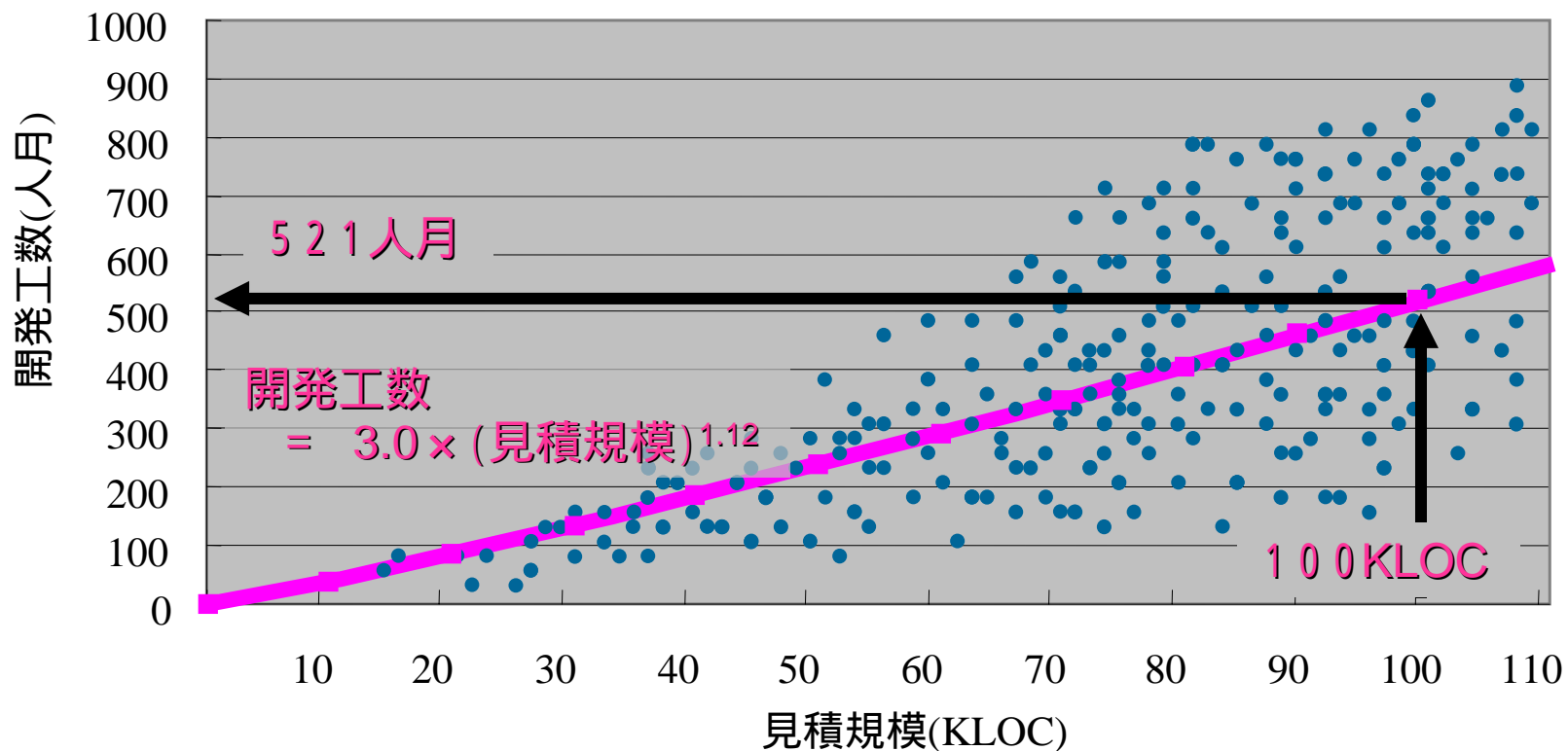
- ・因果関係を明らかにする.
- ・現象を**説明**できるようになる.

吉舖紀子訳, EASEプロジェクト監修, ソフトウェア工学・システム工学ハンドブック: エンピリカルアプローチによる法則とその理論, コンピュータ・エージ社(2005).

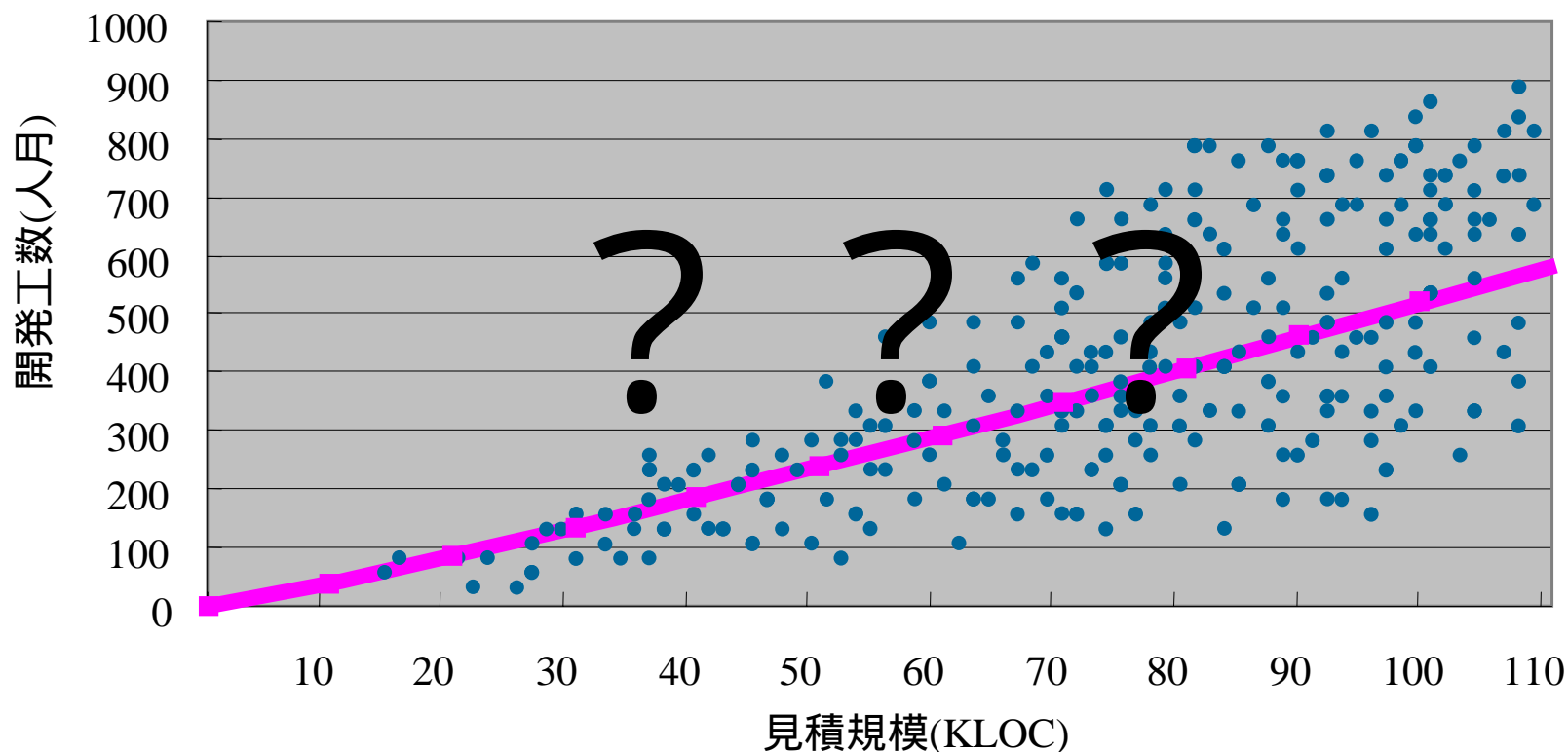
第1段階：観察の実施



第2段階：法則の発見



第3段階: 理論の確立???



人間を含む数多くの要素が複雑に絡み合っ構成されるソフトウェア開発では、「理論の確立」まで成し遂げることは、少なくとも現状では非常に難しい。

エンピリカルアプローチの3段階(改)

第1段階: 観察 (Empirical observations) の実施

- ・実験や調査により現象(事実)を確認する.
- ・現象を**表現**できるようになる.



第2段階: 法則 (Laws) の発見

- ・現象が起こるコンテキスト(状況)を理解する.
- ・現象を**予測**できるようになる.



第3段階: 開発 (Development) の支援

- ・法則を利用して現象を予測, 制御する.
- ・開発を**効率よく行い, 改善**できるようになる.

「法則の利用」が可能にする開発支援

- プロジェクトの工数, 日程, 品質を高精度で見積もる.
- 進行中のプロジェクトをコントロールする.
- 進行中のトラブルプロジェクトを再計画する.
- 組織内の全プロジェクトに対して, リソース割り当ての全体計画を作成する.
- 開発プロセスがどの程度改善されたかをモニターする.

L. H. Putnam, W. Myers(著), 山浦恒央(訳), 初めて学ぶソフトウェアメトリクス ~ プロジェクト見積もりのためのデータの導き方~, 日経BP社, 2005.

開発支援をめざしたデータ分析例

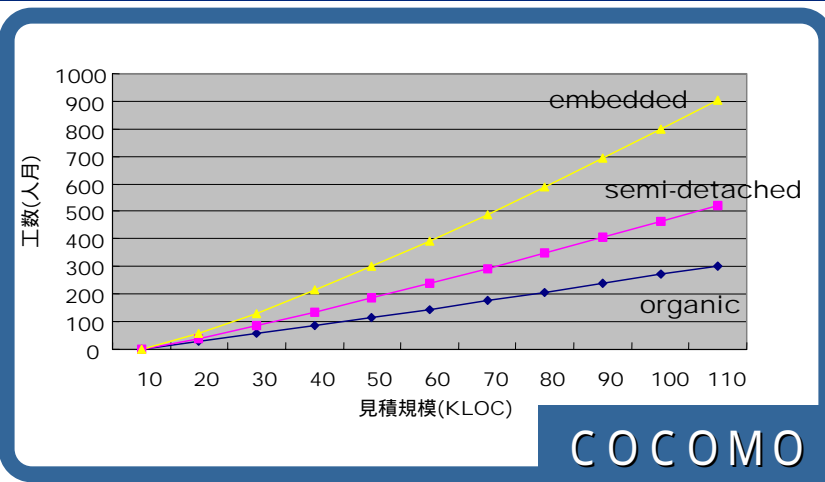
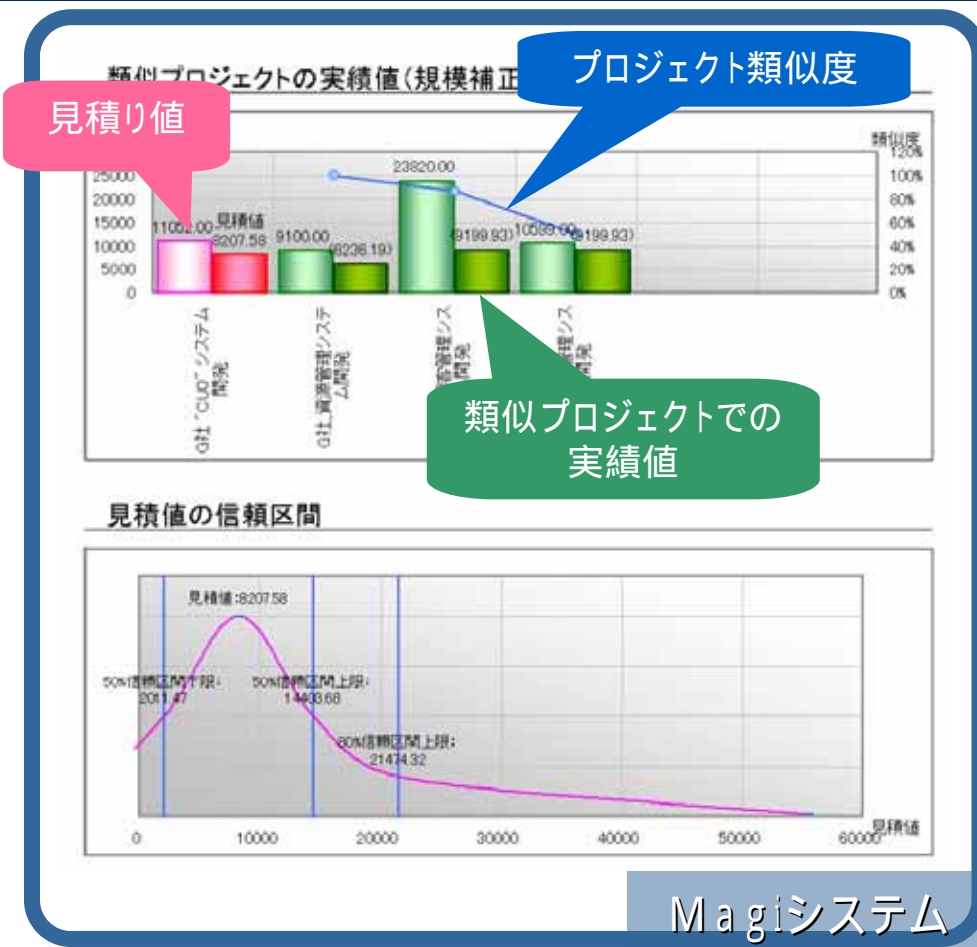
ワンクリック見積りツール Magi

- ワンクリックで

- 対象データに含まれる**未知の値を見積る** .
- 協調フィルタリング技術を応用しているため, データ欠損に強く, 類似プロジェクトの情報を見積り値と併せて提示することができる .
- Trial Editionを公開中 .
<http://se.naist.jp/cgi-bin/magi/licensing/te/getkey.cgi>
IPA版EPMにも同梱されている .
<http://sec.ipa.go.jp/tool/epm.php>
- Professional Editionでは, 「どの程度見積りに使えるか」という観点から, 対象**データの品質を診断**することができる .

工数等の見積りに利用できるデータ・分析・法則

ワンクリック見積りツール Magi

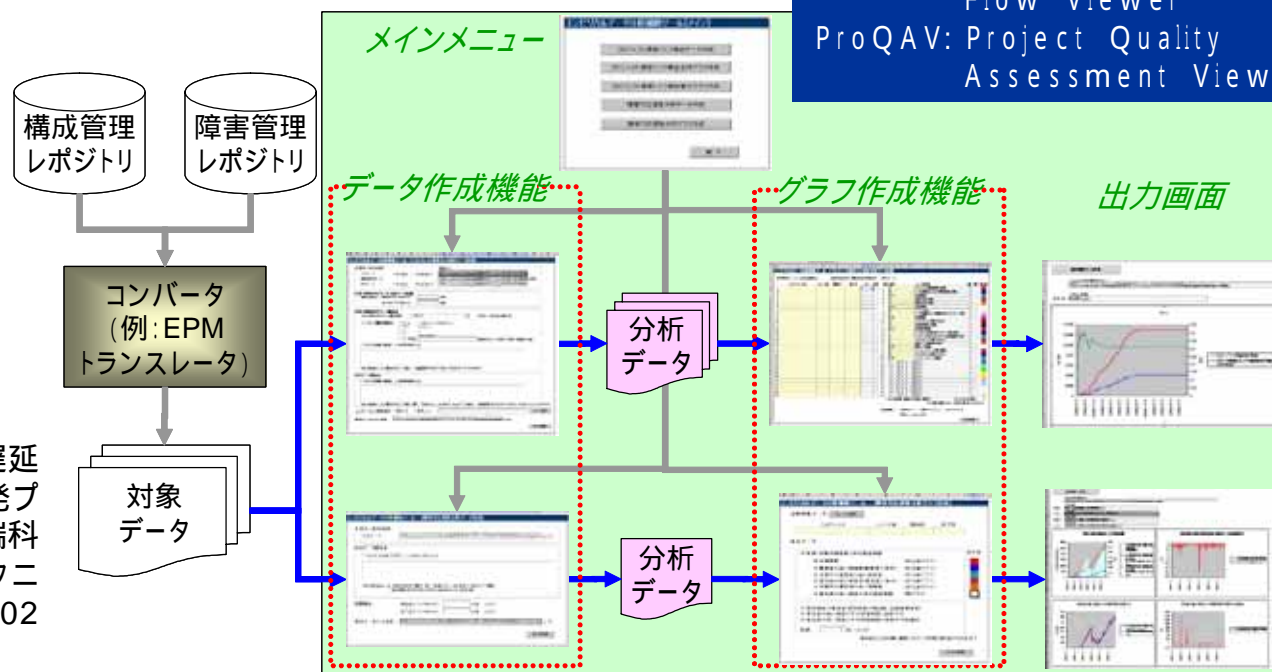


B.W. Boehm: "Software Engineering Economics, Prentice-Hall (1981).

ProPFV, ProQAV

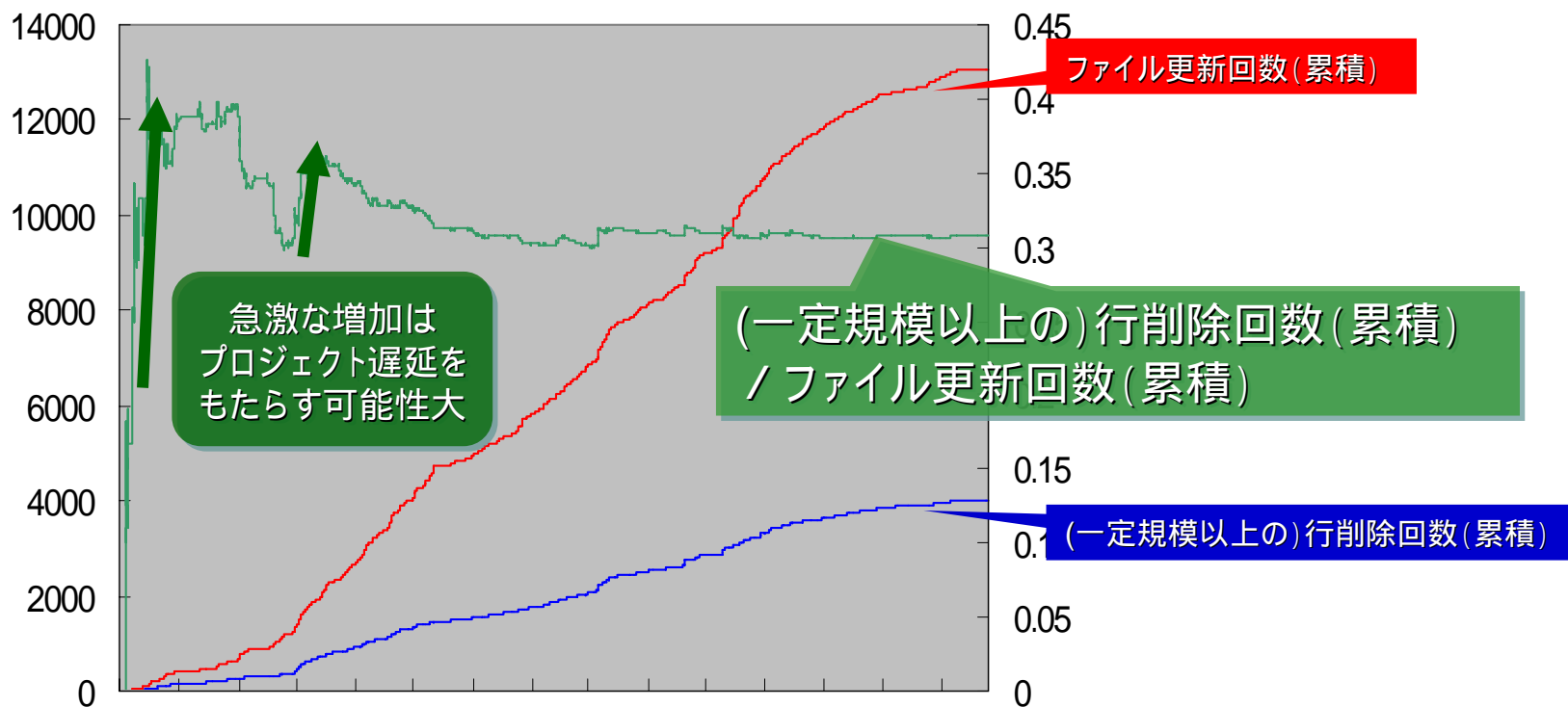
- 構成管理データや障害管理データなどのソフトウェア開発データから、プロジェクトの状況、特に「問題発見」「問題解決」に役立つ情報を抽出 / 可視化する。

ProPFV: Project Process Flow Viewer
ProQAV: Project Quality Assessment Viewer



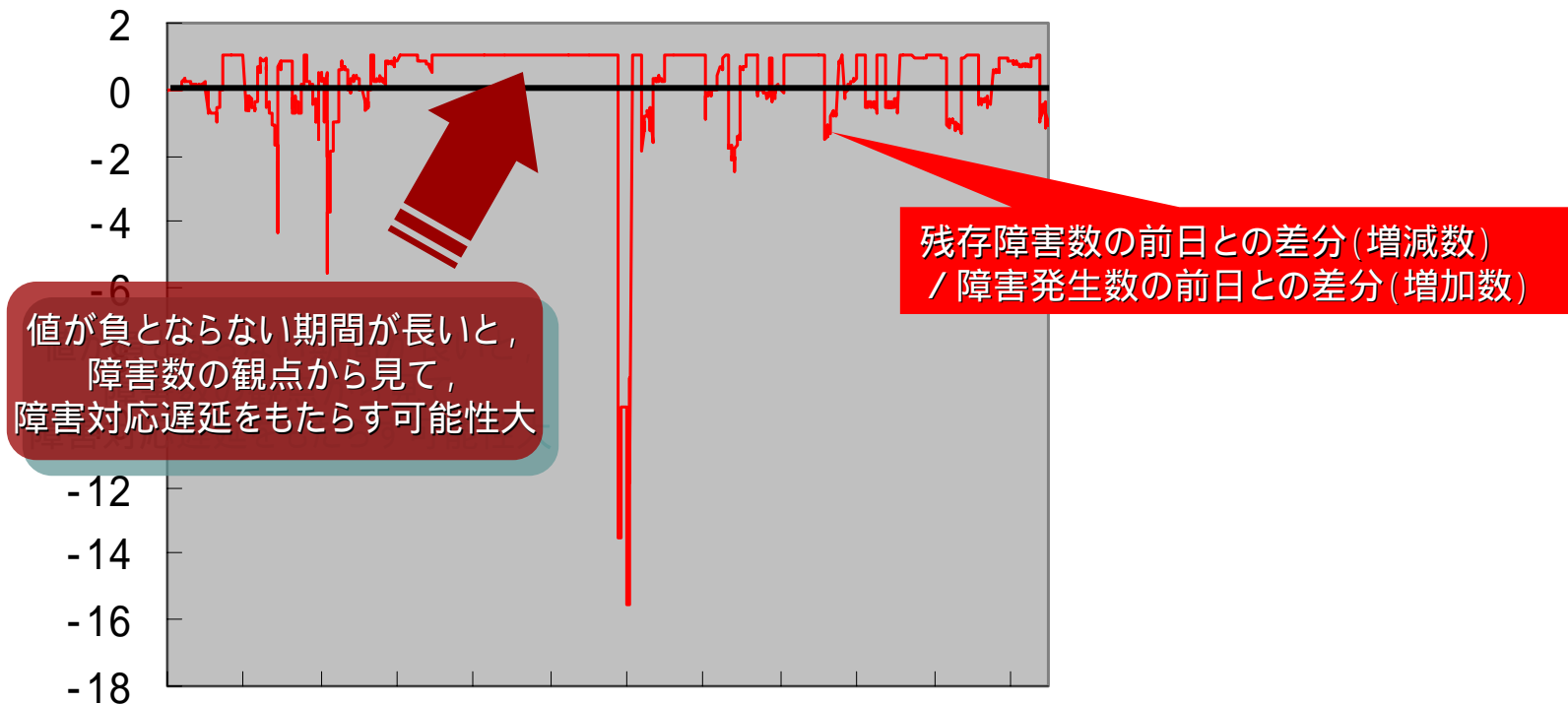
玉田, 松村, 森崎, 松本, “プロジェクト遅延リスク検出を目的とするソフトウェア開発プロセス可視化ツール ProStar”, 奈良先端科学技術大学院大学情報科学研究科テクニカルレポート, NAIST-IS-TR2007002 (Feb. 2007).

ProPFVによるプロジェクト遅延リスク検出



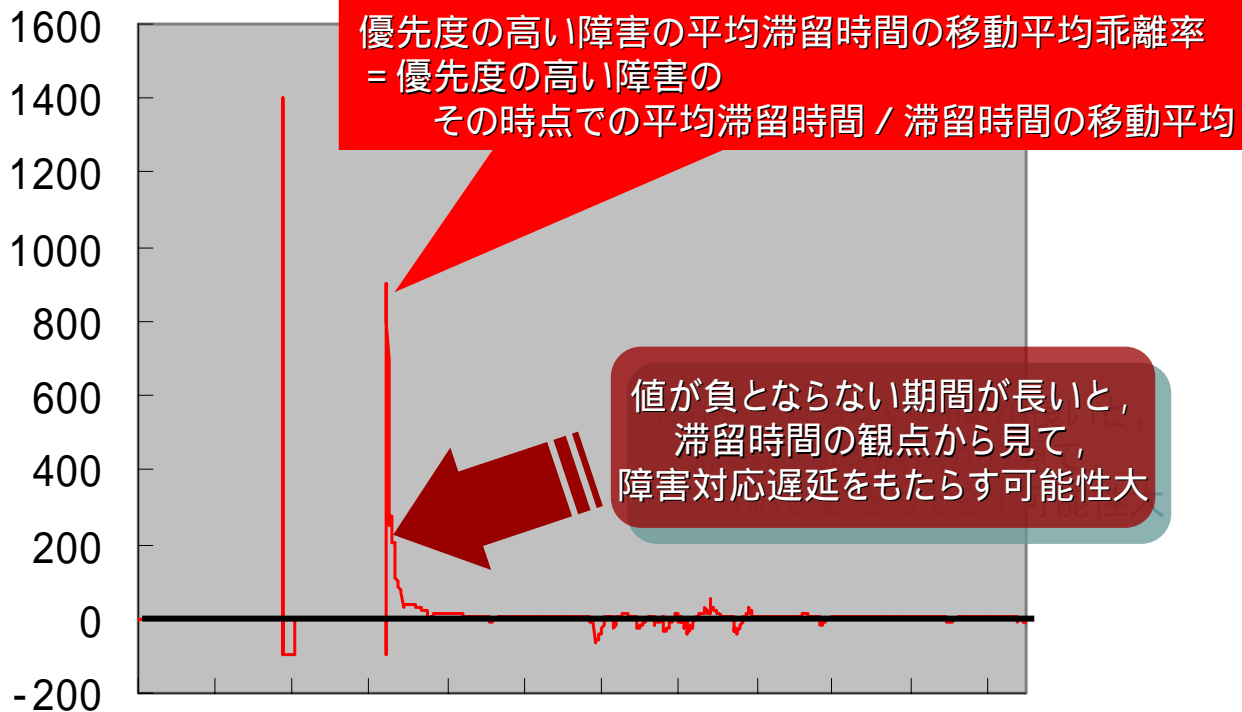
玉田, 松村, 森崎, 松本, “プロジェクト遅延リスク検出を目的とするソフトウェア開発プロセス可視化ツール ProStar”, 奈良先端科学技術大学院大学情報科学研究科テクニカルレポート, NAIST-IS-TR2007002 (Feb. 2007).

ProQAVによる障害対応遅延検出(1)



玉田, 松村, 森崎, 松本, “プロジェクト遅延リスク検出を目的とするソフトウェア開発プロセス可視化ツール ProStar”, 奈良先端科学技術大学院大学情報科学研究科テクニカルレポート, NAIST-IS-TR2007002 (Feb. 2007).

ProQAVによる障害対応遅延検出(2)



玉田, 松村, 森崎, 松本, “プロジェクト遅延リスク検出を目的とするソフトウェア開発プロセス可視化ツール ProStar”, 奈良先端科学技術大学院大学情報科学研究科テクニカルレポート, NAIST-IS-TR2007002 (Feb. 2007).

進行中のトラブルプロジェクトの再計画に利用できるデータ・分析・法則 相関ルール分析ツールNEEDLE

- 事象間の強い関係(相関ルール)を発見することができる。
 - A B
 ある事象Aが発生するならばある事象Bも高い確率で発生する。
- データマイニングプロダクト部門(ユニシス)との共同開発。
- Windows®で実行できるコマンドラインツール群
 - 数値の離散化等の前処理プログラム
 - ルール抽出プログラム
 - ルール選択プログラム

	A	B	C	D	E	F	G
1	バグID	検出できなかった	混入工程	再現頻度	検出工程	種別	除去工程
10	9	レビュー指摘もれ	製造/単体テスト	再現頻度大	製造/単体テスト	コード不具合	製造/単体テスト

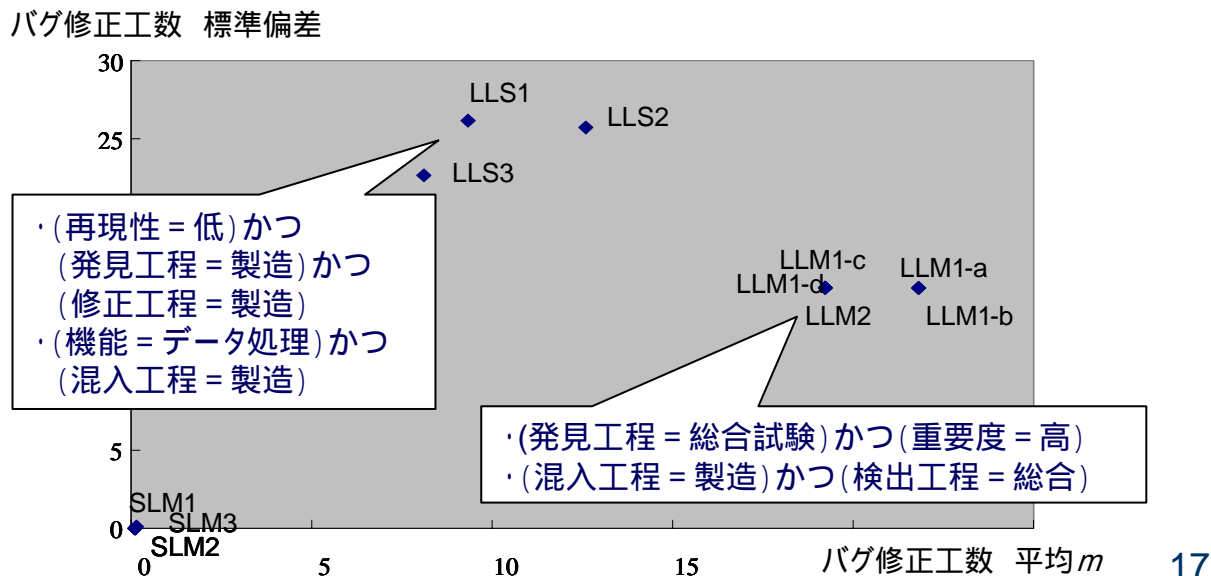
	A	B	C	D	E
1	ル	ルール	支持度	信頼度	リフト値
13		(検出工程 = 製造/単体テスト) & (修正工程 = 製造/単体テスト) →	0.112745	0.496403	3.164568
14		(修正工数 = 3.[0.75人時以上1.00人時未満])			

<http://empirical.jp/needle/>

進行中のトラブルプロジェクトの再計画に利用できるデータ・分析・法則 相関ルール分析ツールNEEDLE

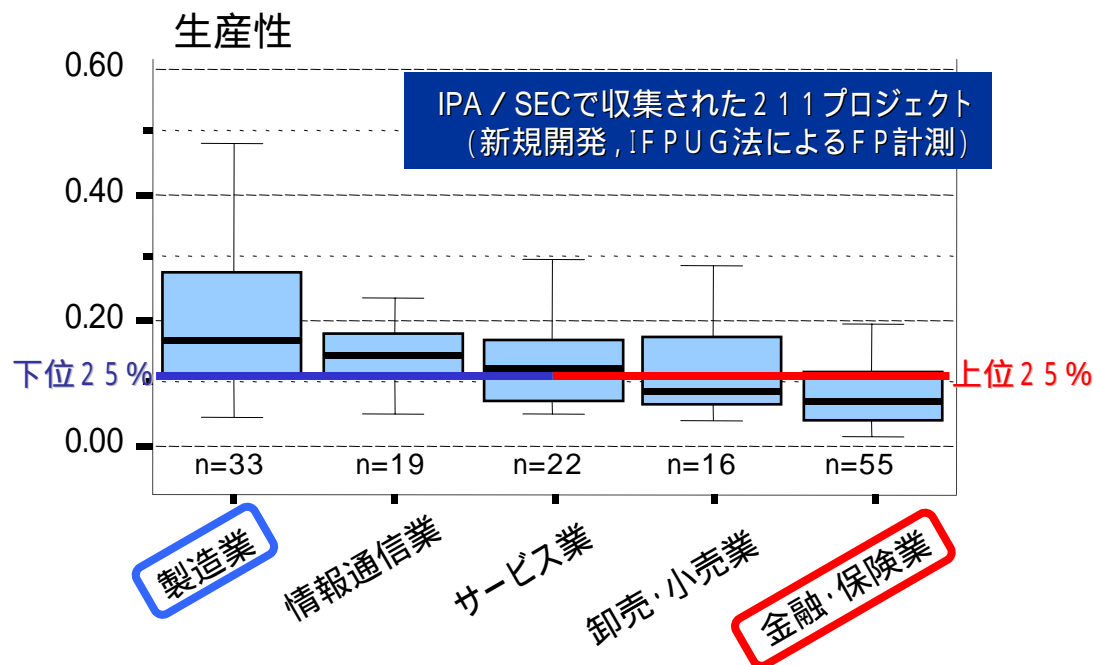
- バグ修正工数の平均やばらつきが大きくなる事象(条件) Aを特定する。
 - A (バグ修正工数 平均 $m > a$, 標準偏差 $> b$) (a, b : 定数)
- 条件Aが成立する場合, 修正するバグに優先順位をつけ, 場合によっては, 再レビューを実施する, リリースを遅らせる, . . .

森崎, 松村, 玉田, 大杉, 門田, 松本,
“相関ルール分析を用いた障害対応データの特徴分析”, 奈良先端科学技術大学院大学情報科学研究科テクニカルレポート, NAIST-IS-TR2007003 (Feb. 2007).



組織内プロジェクトへのリソース割り当てに利用できるデータ・分析・法則

業種間での生産性比較

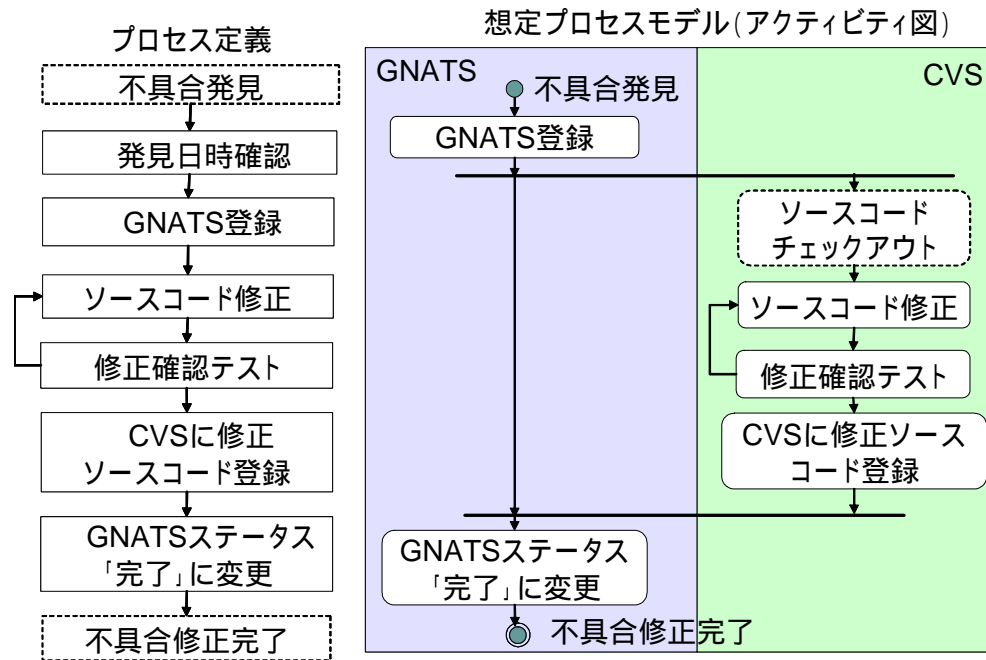


海外の研究でも同様の傾向を示す .
K. Maxwell, and P. Forselius, "Benchmarking Software-Development Productivity," IEEE Software, Vol.17, No.1, pp.80-88, 2000.
R. Premraj, M. Shepperd, B. Kitchenham, and P. Forselius, "An Empirical Analysis of Software Productivity over Time," In Proc. of 11th IEEE International Software Metrics Symposium (METRICS'05), Como, Italy, pp.37, Sep., 2005.
C. Lokan, T. Wright, P. Hill, and M. Stringer, "Organizational Benchmarking Using the ISBSG Data Repository," IEEE Software, Vol.18, No.5, pp.26-32, 2001.

角田, 門田, 宿久, 菊地, 松本, "外部委託率に着目したソフトウェアプロジェクトの生産性分析", 電子情報通信学会技術報告, ソフトウェアサイエンス研究会, SS2006-11, pp.19-24 (2006年4月).

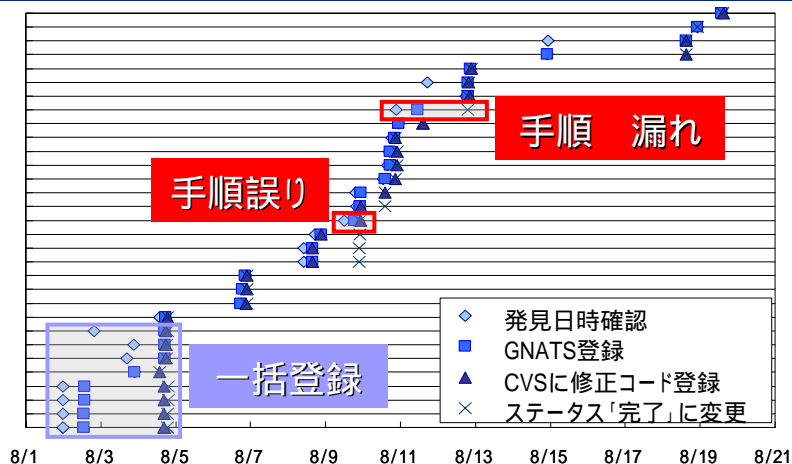
マイクロプロセス分析(デバッグプロセス)

- 作業記録からプロセスに関する情報(作業の順序, 時間, 作業量)を直接得る.
- 基準となるプロセスモデルを与え, 作業記録からプロセスインスタンスを得る.
- プロセスインスタンスに基づいて, プロセス品質を評価する.

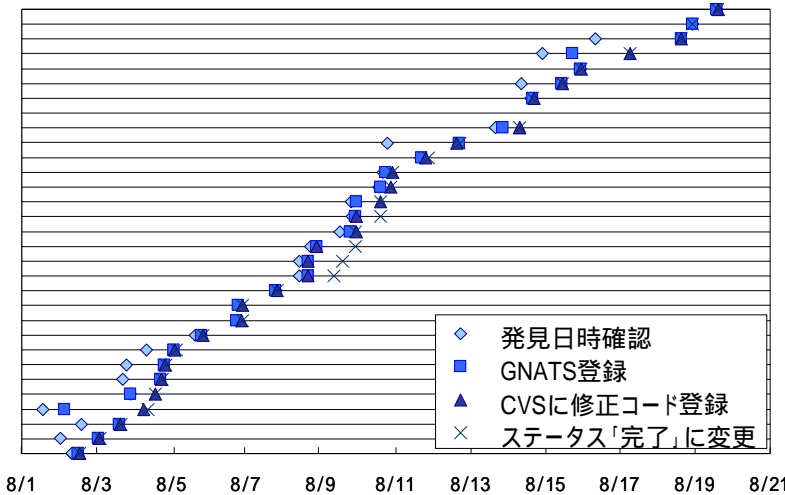


森崎, 松村, 大蔵, 伏田, 川口, 飯田, “エンピリカルデータを対象としたマイクロプロセス分析”, 情報処理学会研究報告, 2006, 125, pp.9-15 (2006年11月).

マイクロプロセス分析 (デバッグプロセス)



「デバッグ作業の見直し」や「構成 / 障害管理ルール of 徹底」により、「手順誤り」、「手順漏れ」、「一括登録」がなくなり、デバッグがプロセスモデルどおりに実施されるようになった (デバッグプロセスが改善された)。



まとめ

- ソフトウェア開発支援をめざしたデータ分析の事例をいくつか紹介した。
 - 「理論の確立」ではなく「開発の支援」をめざしている。
 - 開発した手法のツール化を進めている。
 - 一部は公開し、利用説明会を開催している。
 - 企業との共同研究を通じて、手法やツールの高度化を行う。
 - 共同研究先の企業との守秘義務契約により、現時点では、紹介できない事例もある。

EASEツール展開

