

プロセスと結果の「見える化」が切り拓く シミュレーションの新しい可能性！

個人研究としてのシミュレーション



シミュレーション活用のブレークスルー



(株)キャトルアイ・サイエンス 代表取締役 上島 豊 <http://www.i4s.co.jp/>

1

目次

- 1) 会社紹介
- 2) なぜ、シミュレーションの共有、利活用は、一向に進まないのか？
- 3) 実験系の共有、利活用が進んでいるのは、なぜか？
- 4) シミュレーションの共有、利活用は、本当に進まなくていいのか？
- 5) ASNAROで、シミュレーションの共有、利活用を革新する！
- 6) まとめ

2

研究所で培ったR&D業務のシステム化ノウハウを結集し、 R&D統合プラットフォームを製品化しました。

社名の由来

感性と知性を
情報技術で
革新します



・会社名	株式会社 キャトルアイ・サイエンス
・代表取締役	上島 豊
・本社	京都府京田辺市山手南2丁目1-3 ハチセンビル 3号館 3階
・設立	2006年1月
・資本金	22,500,000円

会社設立経緯

上島が研究員だった当時の課題(日本原子力研究開発機構でHPCによるシミュレーションを行う)

- 1) HPC性能向上によるデータ数・容量増大とそれらの処理・管理作業の増大
- 2) 属人的作業のミス発生や効率化の限界
- 3) 刻々と進展する結果及びコード、ツール共有の煩雑さ

これの問題を解決する仕組みを研究員の時代に考案・開発

2000年	大規模データマネジメント研究会を立ち上げ
2003年	製薬会社や各種研究機関よりシステム相談を受け始める
2006年	キャトルアイ・サイエンスを設立 『研究開発を革新する』を原動力とし、R&D統合プラットフォームを製品化

3

販売開始から7年を経えており、R&Dプラットフォームとして、 お客様から厚い信頼を得ております。

主要取引先(順不同、敬称略)

東京大学	計算科学振興財団
大阪大学	鉄道総合技術研究所
北海道大学	パナソニック株式会社
宇宙航空研究開発機構	シャープ株式会社
日本原子力研究開発機構	株式会社神戸製鋼所
防災科学技術研究所	川崎重工業株式会社
高エネルギー加速器研究機構	新日鐵住金株式会社
情報・システム研究機構	中部地区自動車関連会社 等
情報通信研究機構	
国立天文台	
理化学研究所	



4

2)なぜ、シミュレーションの共有、利活用は、 一向に進まないのか？

5

なぜ、共有できない？

シミュレーションは、「安価」、「やり直しがきく」、 「古いコード・アプリの結果は不要」は、都市伝説！

「シミュレーションは、安価」というのは、計算機、ソフトウェア代、電気代また、関わっている人間全体の給与等コストおよび実験のデータ生成コストを客観的に把握していない発言。比較の為に、実験と同じデータを解析で作ることを考えると、全てのデータに関して、解析が「安価」というわけではない。例えば、実験で破壊検査的なものは、解析が比較的「安価」だが、非破壊検査的でパラメータでデータを振って傾向を見るものは、解析の方が「高価」である。

「シミュレーションは、やり直しがきく」というのは、データ管理がしっかりしていない以上、机上の空論。例えば、複数結果の3次元データの比較グラフで同じグラフが書けるかどうかを考えてみれば、多くの壁が横たわっていることがわかんと思う。

「シミュレーションは、古いコード・アプリの結果は不要」というのは、データ管理がしっかりできておらず比較が困難であることを隠す為の言い訳。新しいコードの良さと悪さを理解するためには、常に比較するという観点が必要不可欠。

「仕組み作りが困難＆面倒」
「共有、利活用のメリットが無い」
「最低限の業務は、個人レベルで閉じて行える」が本当の理由！

6

データ・処理の紐付けが不完全で、実施者に聞かないとその結果がどんなものか分からない！

Who, When, What, Where, Howを確実に！明確に！

解析結果から作ったグラフを再現するのに必要な情報がないと、第三者はその解析結果を正しく、把握できない。

(Overviewの場合はいいのだが、若干でも考察をしようとするグラフレベルで再現できる程度の情報が必要になるはず)

分からない部分を実際の解析実施者に聞くという行為は、「相手に迷惑がかかる」、「異なる立場の人に自分の意図をうまく伝えられない」などで、精神的、時間的に高い壁がある。



7

フリー（特に大学開発）のアプリ、ツールは、DLはできるが、インストールが上手いれない！

サポート、情報の貧弱さ

インストーラや設定UIの完成度



動作保証環境レンジの狭さ、あいまいさ

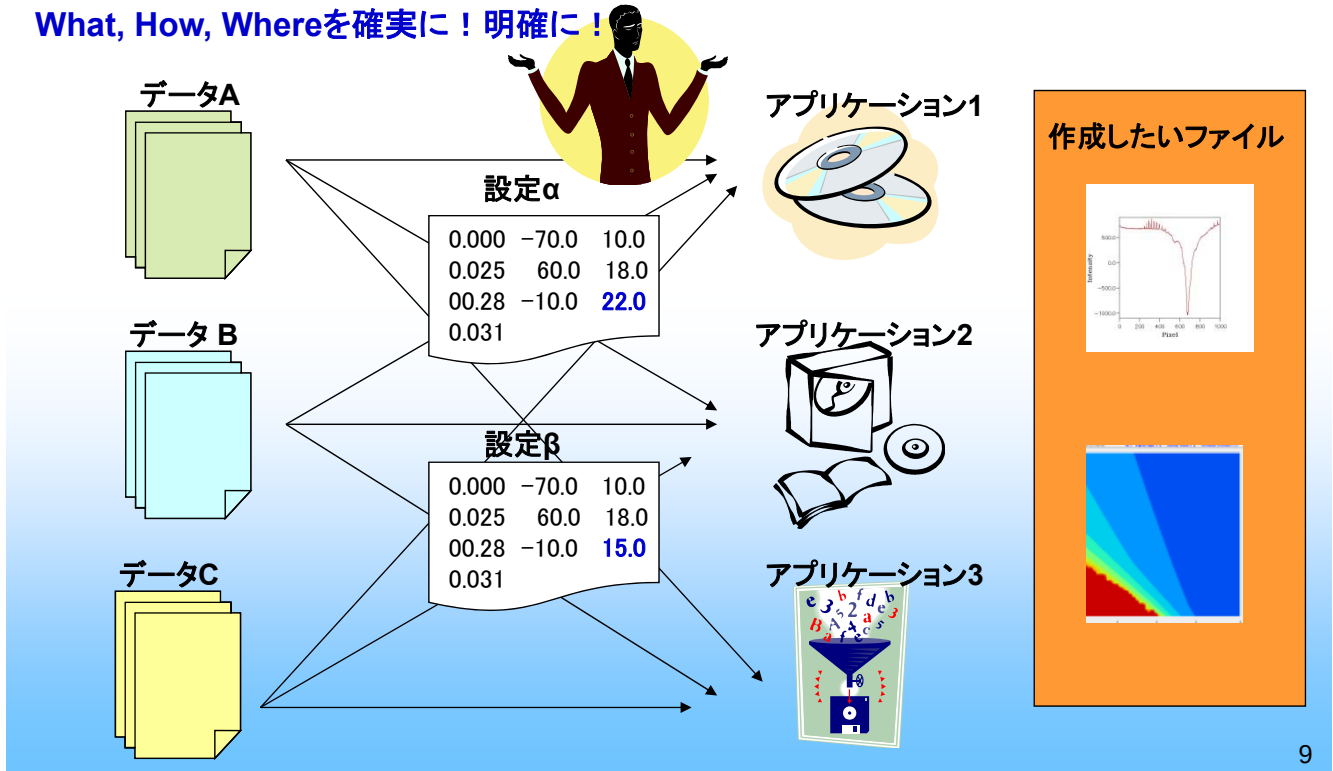
ソフトウェア、インストーラのバグ

8

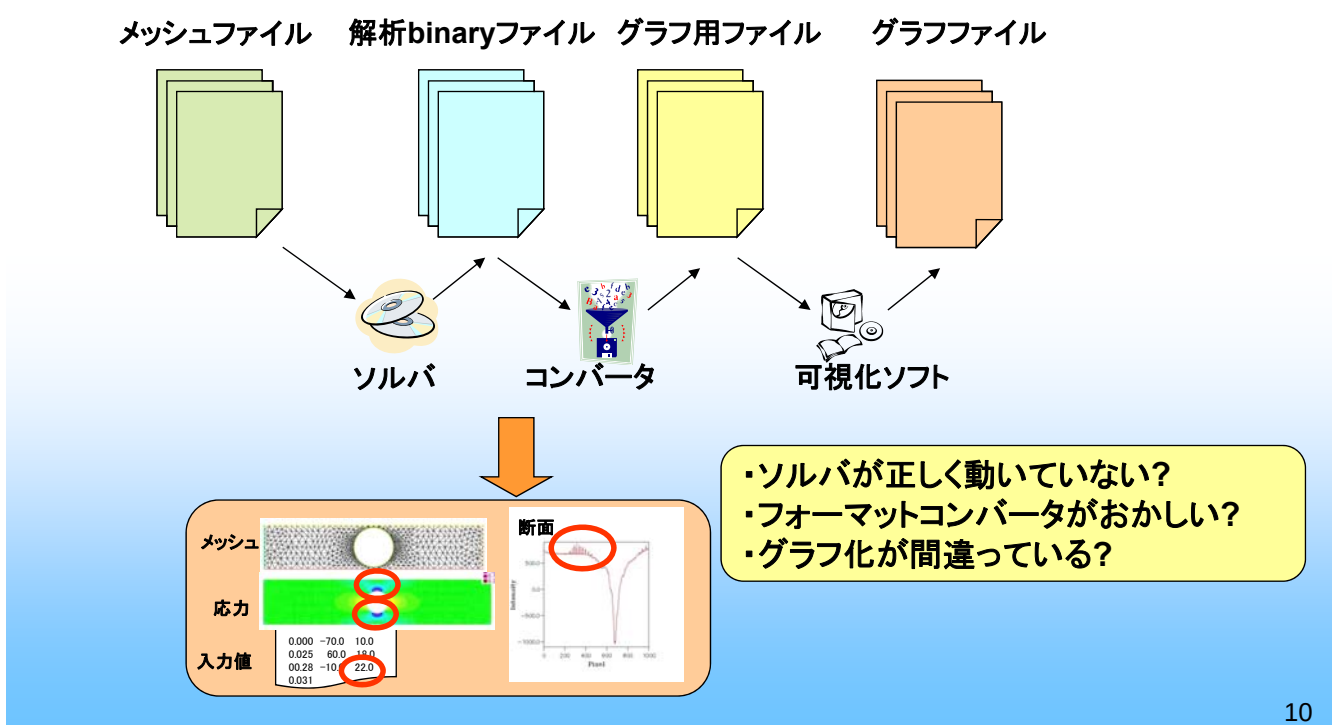


アプリ、ツールがあっても、どのデータに対して、 どのツールを、どのように使えば、良いか分からない！

What, How, Whereを確実に！明確に！



アプリ、ツールが使えても、正しく使えているのか不安でも、 それを確かめるのが、困難！



シミュレーションには、研究開発的なものと CADが設計基準を満たすかを確認するものがある！

a)研究開発的なシミュレーション

研究開発は、どうしても仮説、推論ベースで結論を導き出していくので、研究者の信念が災いし、ドグマに陥る可能性(結論に無理やり合うようにデータを集めてしまうことも悪気なく起こってしまう)を常に有している。詳細なデータなしに結論を裏付ける過程を検証するのは非常に**困難が伴う**。また、**発見プロセスのノウハウ継承も困難**がつきまとう。

b)設計基準を満たしているかの確認シミュレーション

CADデータごとの1点もののシミュレーションであるが、設計基準を満たしているか確認するためには、毎回一定したデータ処理が必要である。**手動処理だと、評価の一定性、客観性、再現性が毀損**する恐れがあるととも**に第三者の確認処理、ノウハウ継承・刷新が困難**である。

**良く考えると、これらすべては、実験も同じことで、
シミュレーションだけが抱える問題ではない！**

3)実験系の共有、利活用が進んでいるは、 なぜか？

実験は、「最低限の業務でも個人レベルで閉じて行えない」 ので「共有、利活用、仕組み作り」が必須！

「実験は、やり直しがきかない」というのは、精度問題である。実質的には、シミュレーションより、確実な再実験が可能なが多い。同じ実験をすることより、他の実験をして、新しい情報を得る方が良いという意識から、再実験は避けるという判断になっている。

⇒現状は、解析よりやり直しは確実！

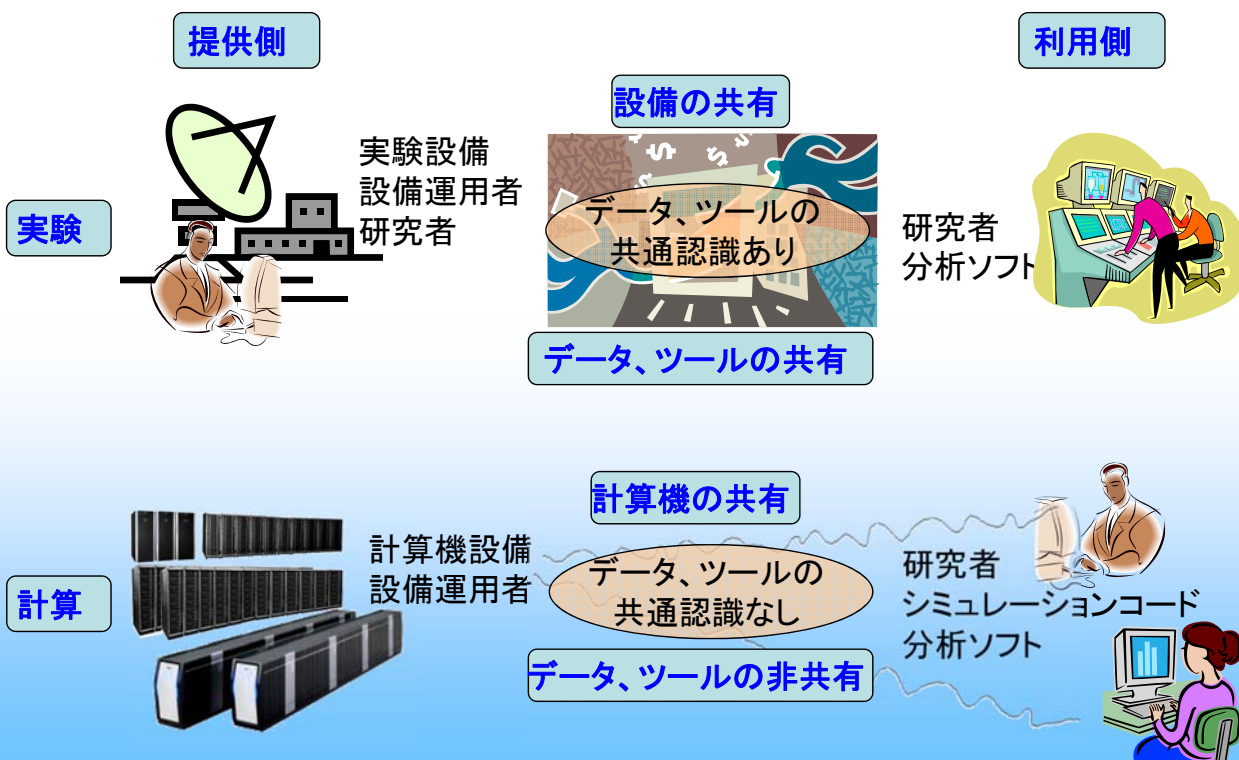
「実験は、古い実験装置の結果は不要」なものとして、扱わない。理由は、新しい実験装置の良さと悪さを理解するためには、過去の結果も常に比較するという観点が必要だからである。

「実験は、高価」というのは、一部の実験に対してである。シミュレーションと同じデータを作ることを考えると、全てのデータに関して、「高価」というわけではない。例えば、実験で破壊検査的なものは、シミュレーションが比較的「安価」だが、非破壊検査的でパラメーターでデータを振って傾向を見るものは、実験が「安価」である。

「最低限の業務でも個人レベルで閉じて行えない」
「共有、利活用が必須」
「実験自体と比較すると仕組み作りは、困難というほどでない」

13

実験では、データ、ツールの共通認識が存在するが、 シミュレーションでは、そこが欠落している。



14



4) シミュレーションの共有、利活用は、本当に進まなくていいのか？

15

共有しなくていいのか？



すばる望遠鏡建設費は、地球シミュレータと同程度、 運用費は1/3で、データ、ツールは共有化済！



参照:wikipedia:国立天文台ハワイ観測所すばる望遠鏡

建設・開発費

約400億円

維持費

約20億円

一日単価

約500万円



参照:wikipedia:地球シミュレータ

約400億円

約50億円

約1300万円



参照:wikipedia:京(スーパーコンピュータ)

約1100億円

約80億円

約2000万円

16



データ保存、移動が大変なのは実験も計算機も同じ！ 再現に必要な情報を確保するのが最優先！

実験	データ量	ファイル数
すばるSC	: 約 14GB/日	約 600ファイル/日
すばるHSC	: 約140GB/日	約 6000ファイル/日
ATLAS	: 約 10TB/日	約 5000ファイル/日
J-PARC	: 約 3TB/日	約 3000ファイル/日

計算機	データ量	ファイル数
JAEA-AlphaSC	: 約 1.5TB/日	約 270ファイル/日
地球シミュレータ	: 約 10TB/日	約 640ファイル/日
京計算機	: 約 1PB/日	約80000ファイル/日

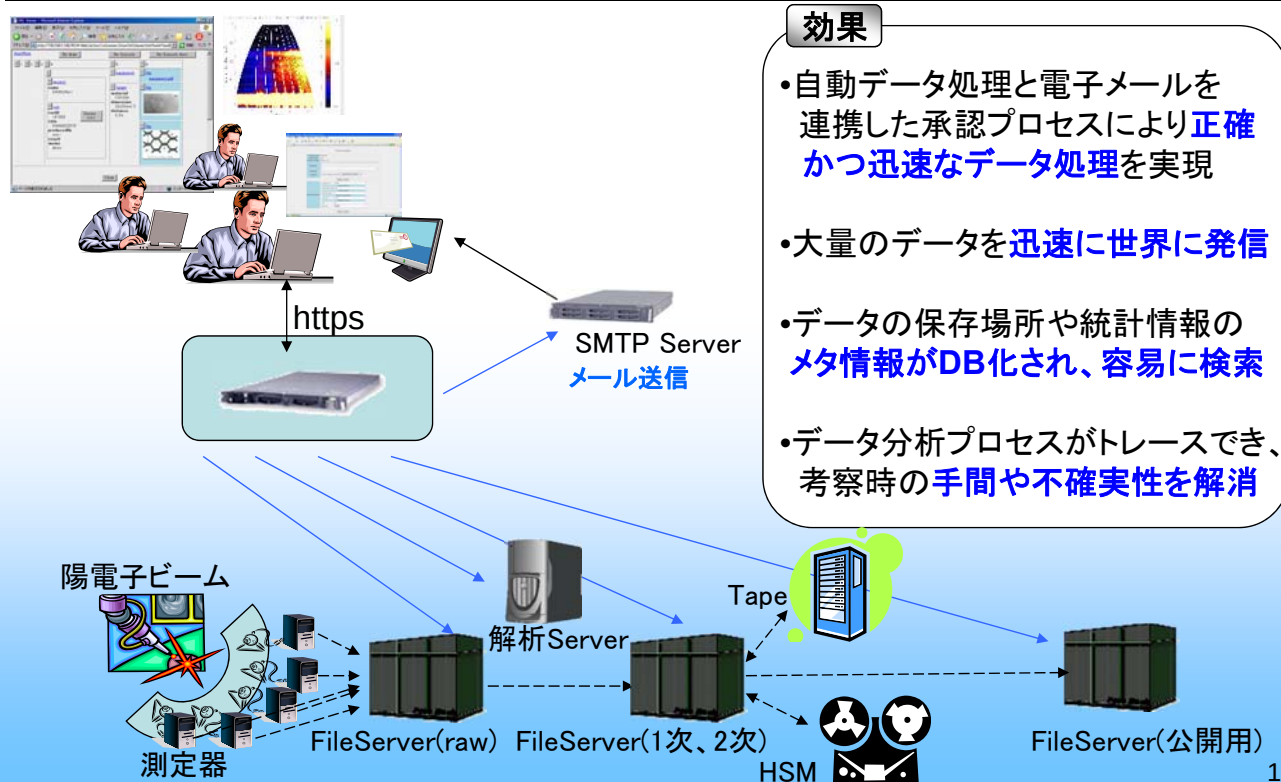
ファイルの移動: 1Gbpsで100MB/s
100Mbpsで 10MB/s(PCにダウンロードするならこの程度が上限)
→100GBで3時間！

複数の研究者が同時に
→ネットワークの帯域を枯渇させ、さらに低速化！

↓
ファイルの存在するところで、解析をするしかない。=>共有化は、行い易いはず！



大規模実験は、共同実験、データ公開、公開データを使った分析等、データ共有化・利活用が進んでいる。





着手の容易さ、データの空間解像度は、実験と比べて、シミュレーションにアドバンテージがある。

実験は、計測器の設置の関係上、PIVなどを除くと、原則的にデータの空間解像度は、大きくできない。一方、シミュレーションは、空間格子レベルでの高い空間解像度があり、実験では、計測できないような、より詳細な現象を察知し、原因を推測することができる。
⇒つまり、高い対話的データ分析が可能であり、ある意味それが必要ともいえる。

実験は、ピンポイントで再実験をしようとしても一人で勝手に実施できるものではなく、実験スケジュール調整や実験準備などがあり、再実験の着手には高いハードルがある。一方、シミュレーションは、ピンポイントの再シミュレーションであれば、少なくとも個人の判断で実施できることが多く、着手ハードルは低い。

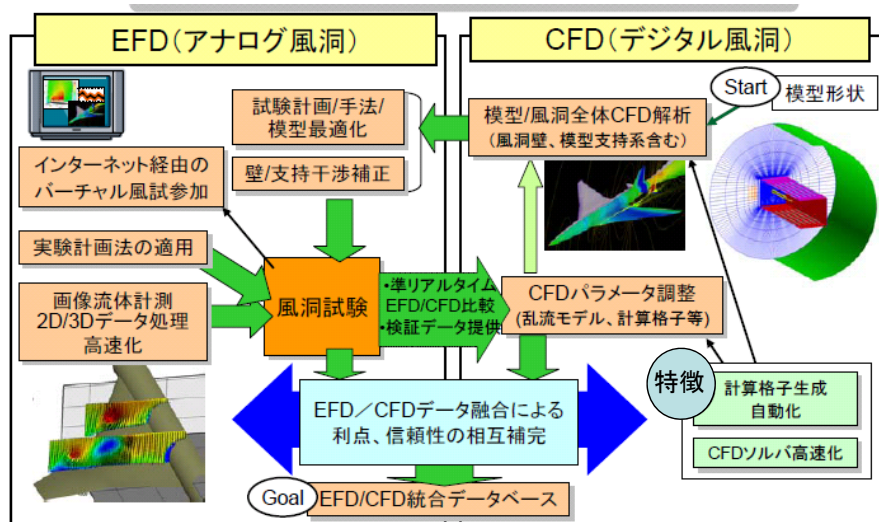
「データ処理の高い自由度とそれに伴う高い対話性」、
「着手ハードルの低さ」が、シミュレーションの利点であり、
それが再現性を低減させている原因でもある。



シミュレーションも蓄積していけば、実験と同様に 共有化 メリットがあり、コードの強み・弱みも定量的に把握！

JAXA DAHWINシステムでの適応事例

実験の弱点(壁・支持棒効果等)と計算の弱点(モデル適用範囲、精度等)を相互に補完！
実験変更決定、コード改良指針、優先度も明確に！



精緻かつ多量の蓄積データに統計解析を適用することで、仮説の定量的検証や新たな発見が可能となる。

統計解析、学習システム等を適用する上で重要なこと

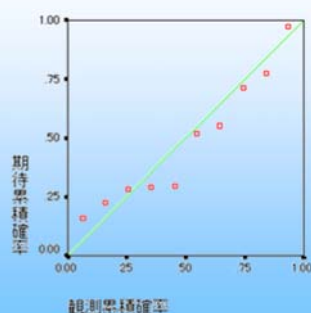
1) データの生成過程の完全な情報記録

=> データ母集団の品質問題

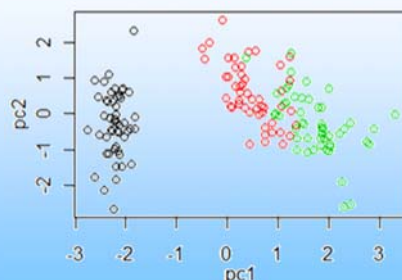
品質とは、情報に間違いが無いという意味と
恣意性が働いていないという意味

2) データ母集団の量

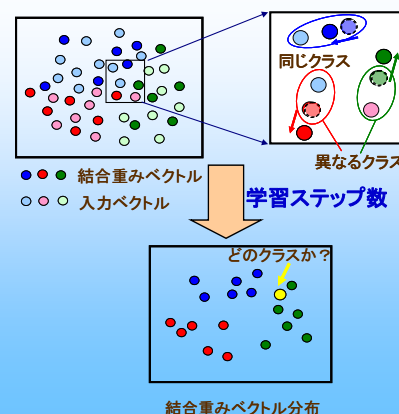
重回帰分析



主成分分析



学習ベクトル量子化法



21

「共有、利活用のメリットが無い」、「最低限の業務は、個人レベルで閉じて行える」では、実験を超えられない！

a) 研究開発的なシミュレーション

「研究開発的なシミュレーション」においては、**結論を裏付けるためや発見プロセスのノウハウ継承のため**には、**詳細な解析プロセスの記録に基づく多様な視点からの検証**が必要である。

b) 設計基準を満たしているかの確認シミュレーション

「設計基準を満たしているかの確認シミュレーション」においては、できる限り手動処理を排除した自動化により、**評価の客観性、再現性を確保しつつ、解析ノウハウ、ルールの見直しが実業務データに基づき、逐次行われることが必要**である。

「共有、利活用のメリットが無い」

=> 上記のように多くのメリットがあるはず！

「最低限の業務は、個人レベルで閉じて行える」

=> もっと上を目指さないと実験再現のための解析どまり！

「仕組み作りが困難＆面倒」

=> 弊社ソフトウェアASNAROで解決！

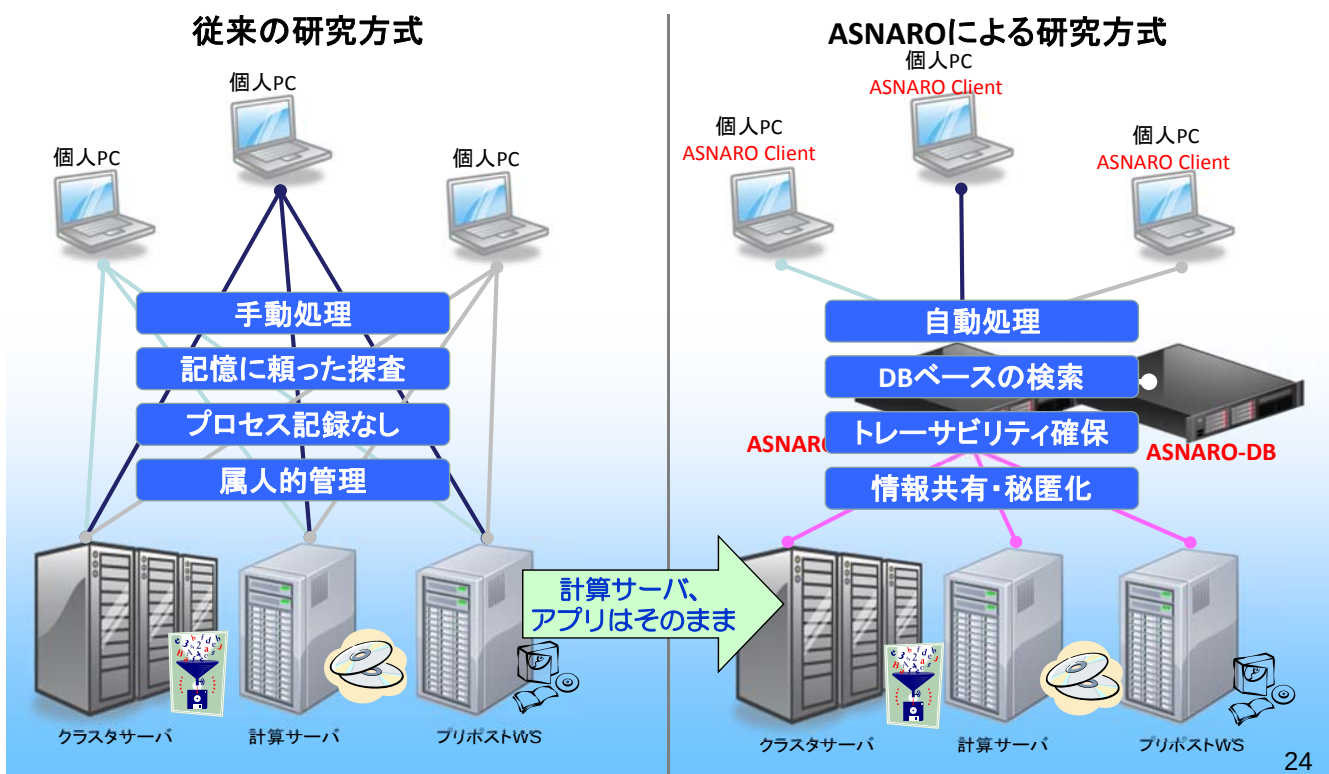
22

5) ASNAROで、シミュレーションの共有、利活用を革新する！

23

ASNAROで革新

ASNAROは、開発フリーでシミュレーションコード、結果の共有、利活用を革新する！

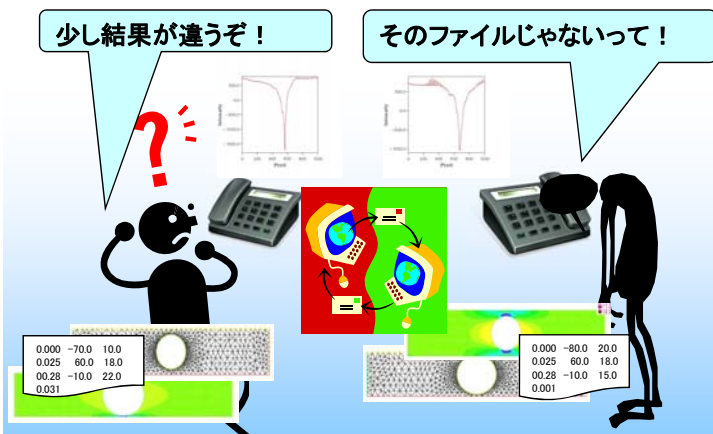


24

ASNAROを使うことで、解析担当者の手を煩わせることなく、解析を再現できるレベルの詳細情報が簡単に参照可能！

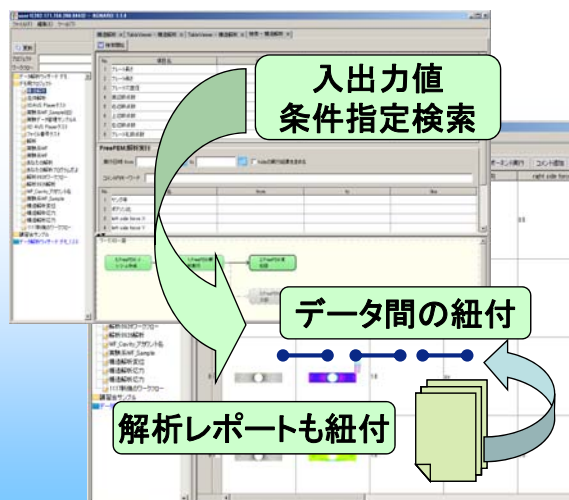
従来方式

電話、メール、対面で説明はしているのに...



ASNARO方式

解析再現に必要な全情報が検索一発！

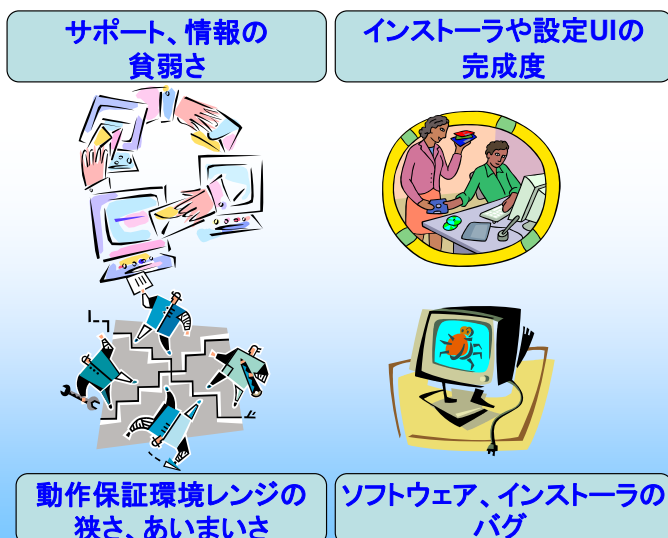


25

誰かがインストールしたものを共用することで、アプリ、ツールの整備ハードルは大きく低減！

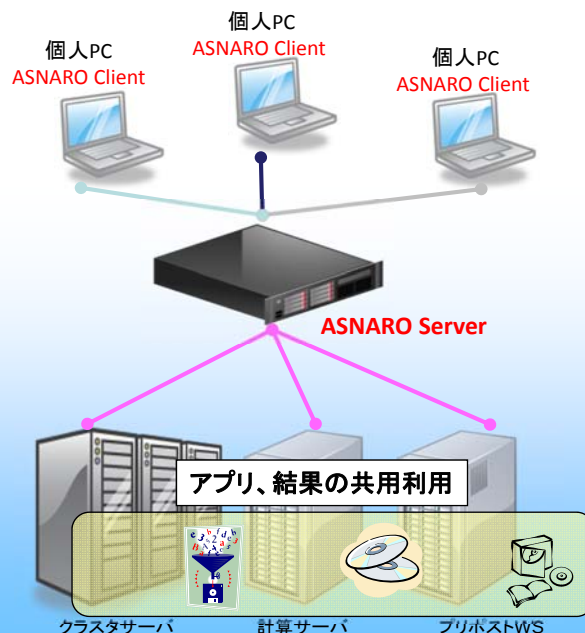
従来方式

インストール自体が大変！



ASNARO方式

誰かがインストールしたらそれを共用利用！



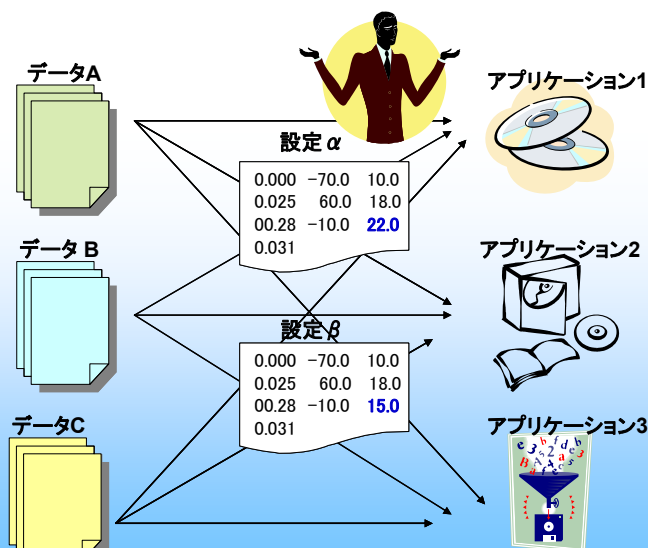
26



データとツールの紐付け機構により、どのデータには、どのツールを、どのように使えば良いかが一目瞭然！

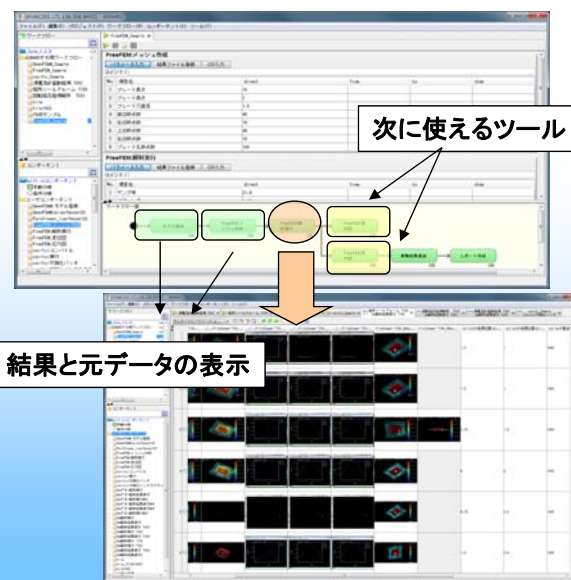
従来方式

どのツールが使えるのか分からない！



ASNARO方式

DBとWorkFlowでデータとツールが紐付け！



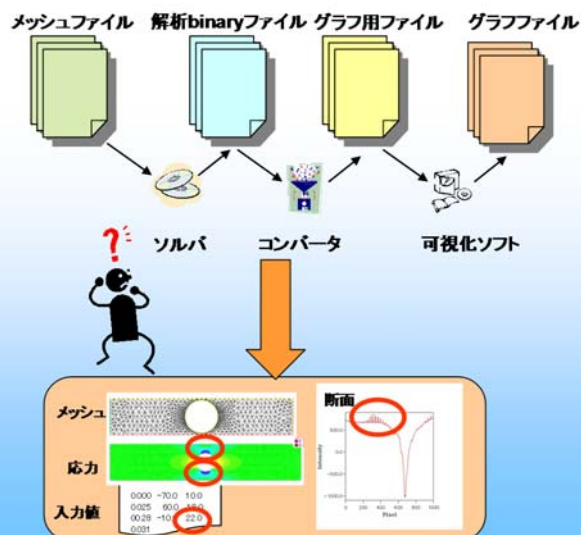
27



データ生成のプロセス履歴が検索でき、表形式で並べられるので、類似結果から正しさを判断できる！

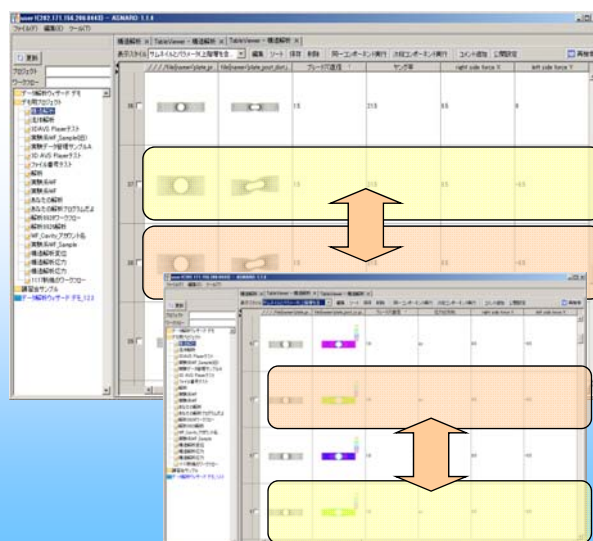
従来方式

結果が正しいか判断ができない！



ASNARO方式

解析の過去類似結果から正しいか判断ができる！

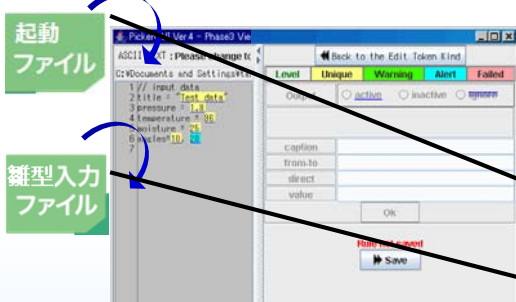


28

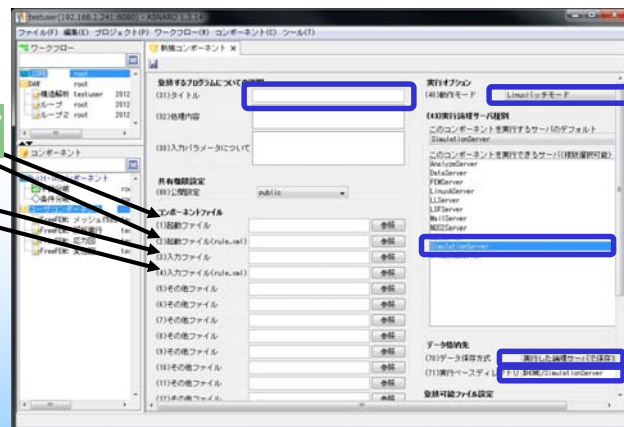


Batchシェルと雛型入力ファイルが準備できたら、 3stepで、簡単に入出力値のDB化ができる。

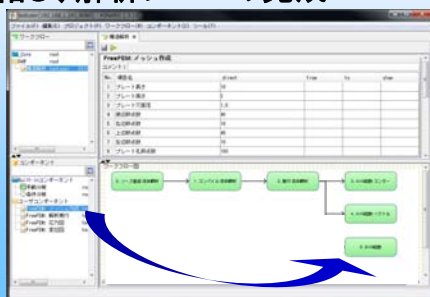
- ①Pickerツールで起動、雛型入力ファイル
の変数化・文字列置換位置を指定する。



- ②コンポーネント登録画面で①ファイルを選択し、
コンポーネント名、動作モード等を設定する。



- ③ワークフロー編集画面で必要なコンポーネント
を連結し、解析フローの完成

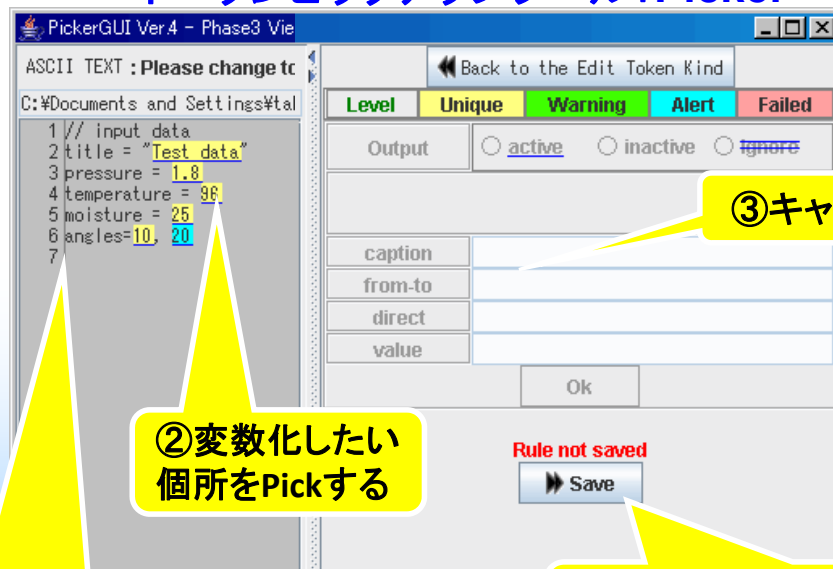


29



雛型入力ファイルと起動ファイルを準備し、 GUI化、DB化の為のルールファイルを作成する。

トークンピックアップツール:Picker



- ①入力ファイルをLoadする

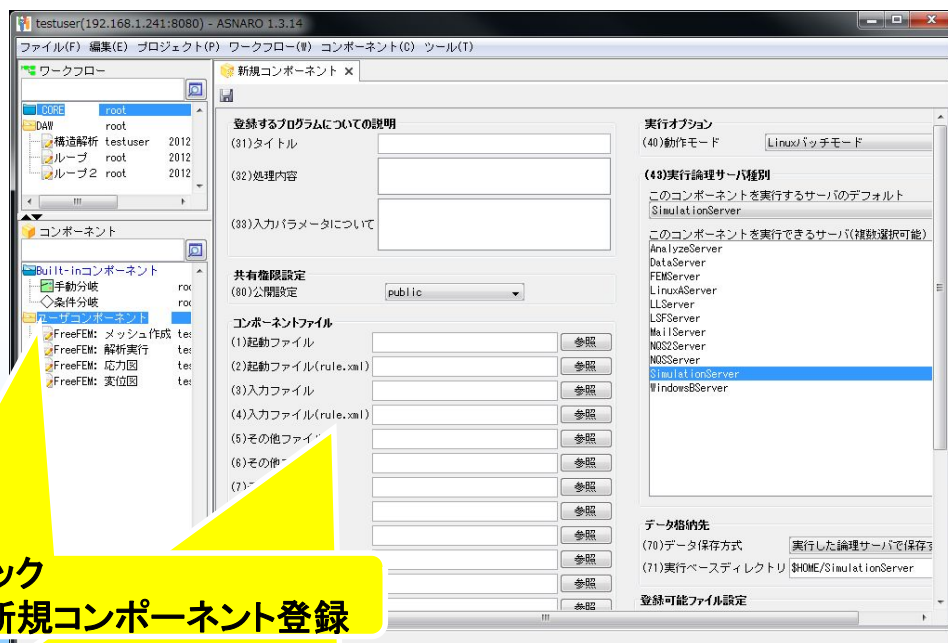
- ②変数化したい
個所をPickする

- ③キャプションを指定する

- ④Saveをクリックすると
入力ルールファイルが作成できる

30

入力、起動、パターンマッチングファイルを選んで、
コンポーネント名を入れるだけで登録は完了です！

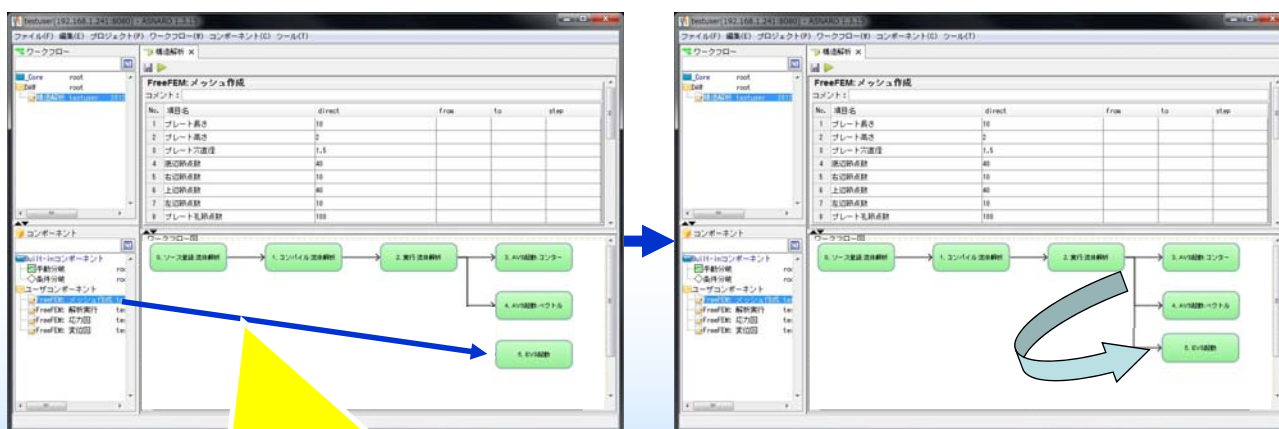


①右クリック
=>新規コンポーネント登録

②新規コンポーネント登録は、入力ファイルと起動ファイル、
それらのパターンマッチングファイルをファイルセレクトで選択するだけ！

31

登録済コンポーネントをドラック&ドロップするだけで、
新しい解析フローが簡単に作成できる。



①登録済みのコンポーネントをドラック&ドロップするだけで、新しい解析
フローが簡単に作成！類似解析フローの複製からの編集作成も可能！

32



6)まとめ

33

まとめ



データ、ツールの共有、利活用を進めると シミュレーションの可能性は、まだまだ広がる！

シミュレーションデータ、ツールの共有、利活用により、新たな価値が創造される。

- データ、ツール利用の拡大で、**多様な観点により、新たな発見**が生まれる。
- ノウハウの共有・継承により、**R&Dの安定的効率化**が図られる。
- ツール、データ生成者外の第三者確認ができ、**結論に信頼性が担保**できる。

しかし、シミュレーションは、実験と異なり、その**手軽さ・単独実施性**および共有・利活用**環境を整える困難さ**が共有、利活用の妨げになっていた。

シミュレーションの共有、利活用を進めるためには、実験と比べ、格段に簡単に共有・利用環境を整えられることが必要である。

ASNAROは、それら問題を解決し、R&D環境を革新します。

- ツール、コード共有機構によるインストールハードルの低減
- データとツールの紐付け機構による処理手順、パターン選択ハードルの低減
- 完全なトレーサビリティ記録によるデータ生成プロセスを把握するハードルの低減

34

「ASNARO」は、R&Dプロセスを革新し、 技術者の創造性を支援します！

キャトルアイ・サイエンスは、
感性と知性を
情報技術で
革新します！

ASNARO

