

先進技術解説

計算工学の分野で注目を集めている先進技術をわかりやすく解説するページです。近年の研究では、膨大なデータやプロセスを通して結論を導くことがよくあります。今回は、解析/シミュレーションにおけるデータやプロセスを共有・管理するシステムASNAROです。共有・管理に必要となる先進技術について株式会社キャトルアイ・サイエンスの上島豊氏に解説していただきます。

解析/シミュレーション・データ処理 共有/管理システム — ASNARO —

上島 豊

1 はじめに

筆者が日本原子力研究開発機構の研究員だった当時、地球シミュレータ等HPCで大規模シミュレーションを数多く実行して物理現象を解明する研究を行っていた。その中で、物理問題ではなく、ツールであるコンピュータ環境について以下のような困難な問題に直面し、本来の研究目的である物理の論文をまとめる以前に多くの時間を奪われてしまう経験をしてきた。^[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
- ◆属人的作業のミス発生や効率化の限界
- ◆刻々と進展する結果及びコード、ツール共有の煩雑さ
- ◆チームでの議論を行うための正確かつ適切な資料準備の困難さ

そこで、これらを解決するためのシステムを考案し、後に「学際大規模情報基盤共同利用・共同研究拠点(JHPCN)の公募型共同研究」として、平成22、23、24年度の採択課題、「マルチパラメータサーベイ型シミュレーションを支えるシステム化技術に関する研究」^[2-3]で開発、検証を進めた。本稿では、CAE、解析に特化させて現在に至る「解析/シミュレーション・データ処理共有/管理システム—ASNARO—」について解説する。

筆者紹介



うえしま ゆたか

1997年 大阪大学大学院工学研究科博士課程修了。
同年 日本原子力研究所 博士研究員 入所。2000年 日本原子力研究所 研究員。2006年 (有)キャトルアイ・サイエンス 代表取締役。2008年 (株)キャトルアイ・サイエンス 代表取締役。2000年以降 ITBL、バイオグリッド、NAREGI、生体細胞機能シミュレーション、JHPCNプロジェクト等の様々なGRID-IT系プロジェクトに参加。

2 解析/シミュレーション環境の現状

筆者が経験したような前述の問題は、みなさんの周りでも生じているのではないだろうか。

近年、解析/シミュレーションを取り巻く環境は、高性能計算機やネットワークの普及、解析/シミュレーションや可視化手法の確立、多様化によって、結果に至るまでのプロセスや出力されるデータも多種多様となってきた。その結果、目的とする可視化/解析結果や最終的に導かれた結論が、類似の膨大な情報に埋もれてしまい、目的とする結果の導出過程の再現が困難となり、結論の論拠に関する不確実性を高めてしまう危険性さえ示唆されている。

一方で、実験・計測分野では再現性が重視され、結果とともに条件やプロセスを一元管理し、それを様々な視点で分析できるように情報の共有化が行われ、実験に関わった技術者だけでなく、実験の実施には関わっていない様々な分野の技術者もデータを分析し、研究を行っている。このような研究・開発環境のもとでは、実験結果の客観性、確実性をより強固にするだけでなく、特定の専門技術者らだけでは、見いだせなかった新事実の発見も稀ではなくなっている。

一般社会でも、e-コマースやSNS、ビッグデータの様々な情報が情報発信者だけでなく、多くの第三者により分析され、サマライズされ、再発信されることが当たり前になっており、多くの新しい事実発見や価値を生み続けている。

一方、解析/シミュレーションの分野では、最終的な論文や報告書は蓄積されるが、データ管理は個人の技術者の属人的判断に任されてしまうことが多く、また、属人的な管理は技術者の曖昧な記憶や簡単なメモによることも少なくない。このような状況では、解析/シミュレーションを担当した技術者でさえ、各データの迅速、正確な再現処理が困難となり、共有化どころの話ではない。

論文や報告書の最終判断となるデータは、可視化、

グラフ、各種統計量であり、その再現処理は結論導出過程、つまり可視化、グラフ、各種統計量がすべて再現できることが重要である。そのために、解析/シミュレーションでは、膨大な処理手順、入出力ファイル間の紐づけ情報とそれらの設定パラメータを記録する必要があるが、そこまで細かな情報を完全に記録することは、非常に手間がかかるため、実際には、そこまでの記録は取られていないのが現実で、実は、「解析/シミュレーションは、実験と違い再現計算がいつでもできるから、結果は消してもよい」というのは、都市伝説だったのではないだろうか。

高速に大規模な計算を数多く実行できるようになった今こそ、大量の計算結果から有用な情報を選択、抽出するために様々な観点での分析が必要で、一人の技術者の観点だけでは、十分ではなくなってきており、実験やビックデータ分析で活用されているような多人数、多観点での視点導入が重要である。

これらの課題を鑑みると、解析/シミュレーションの各種制御パラメータや解析処理の手法選択、それらのパラメータなどの履歴をトレースできる支援環境を構築する必要性は明らかである。

さらに、技術者の作業履歴が完全にトレースバックできるようになると、それらを分析することで、熟練技術者の暗黙知を形式知化することができ、新人技術者教育にも利用することもできる。

この解決策の1つとして、図1に示すRCM (R&D Chain Management) システムを構築するための基盤ソフトウェアを開発し、解析/シミュレーションシステムへの応用について報告した。^[2-3]

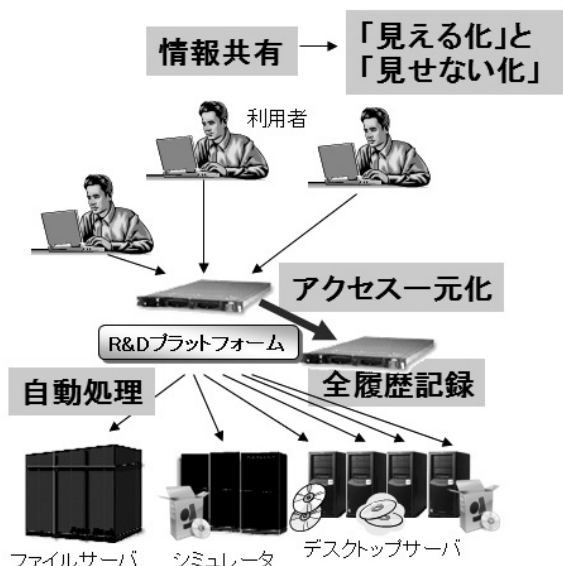


図1 RCMシステムソフトウェアの概念

本稿では、このRCMの技術をベースに、ユーザ・インターフェースを高度化し、解析/シミュレーションに必要な機能に特化させることで、ユーザビリティを向上させたASNAROを例に、解析/シミュレーションのプロセスを含めたデータベースを安全に運用・管理し、利用者同士が情報を簡単に共有できるシステムの具体例^[4]を紹介する。

3 解析/シミュレーション・データや処理プロセス共有の壁

一般的に解析/シミュレーション・データや処理プロセスを共有し、欲しい情報を探そうとすると、図2に示すような壁に突き当たることがある。

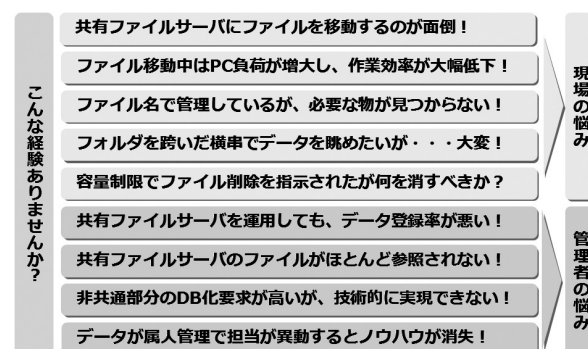


図2 データ/プロセス共有化の壁

これらを解決するために、ASNAROでは、様々な技術を用いている。

例えば、柔軟な絞り込み検索のためには、解析/シミュレーションの結果ファイルだけではなく、入力ファイルなどに記述されるモデルの情報、解析手法、条件・パラメータ、あるいは解析/シミュレーションプログラムとは独立したプリ・ポストプロセスに関するデータや情報等、多段に構成された処理の中で発生する多種、多岐にわたる値を分類用のキーワード(メタ情報)としてデータベース化されている必要がある。

また、このようなデータベース化作業を手動で行うことにしていると、「後で時間がある時にまとめて」、「時間が無いから、とりあえずの仮登録」といったことが起こり、「登録の確実性」、「登録内容の正確性」が確保されなくなる。実際に手動作業で実施しているところでは、時間経過とともにデータベース化率の低下が顕著に表れ、データベースの利活用が進んでいないという声が聞かれる。つまり、「多段で行われる処理の関連性が明確に紐づけられた情報と生成されるファイルの中身、関連性を同時にデータベースに登録、管理すること」は非常に難しく、実質的に運用、利活用するためには、自動で多段処理を制御しながらデータベース化することが必要である。

ASNAROでは、次のような機能を実装し、これらの問題を解決している。

- ◆スパコン、クラスタ、ワークステーション、クライアントPCなど様々な計算機資源へのジョブ投入を可能にし、ジョブ実行後、自動的に共有ファイルサーバにファイルをステージングするジョブ投入機能
- ◆異なる計算機間において、ジョブ間のファイル、情報引き渡しを制御しながら、複数のジョブを逐次実行していくことを可能にするワークフロー機能
- ◆ジョブ投入機能、ワークフロー機能と密接に連携し、ジョブ実行と同時に入出力ファイル、入出力値などのメタ情報をデータベース設計なしに自動的にデータベース化することを可能にするスキーマレス非構造データベース機能^[5-6]

4 ASNAROによるデータ処理、共有・管理

4.1 ジョブ投入機能

計算サーバで解析/シミュレーションの実行をさせたり、可視化サーバで結果を画像化したりする処理を担うのが、図3に示すジョブ投入機能である。ASNAROのジョブランチャーのジョブ投入機能では、クライアントPCからスパコン、クラスタ、ワークステーションなどの計算サーバへのバッチジョブ投入だけでなく、X-Window、VNC-Viewerと連動した対話的処理やクライアントPCのアプリケーションでのバッチ・対話処理もジョブ投入が可能である。

また、ASNAROのジョブ投入機能では、結果データを共有ファイルサーバに自動的にステージングする機能も有しており、データ共有のために毎回手動で行っていた共有ファイルサーバへのファイル転送操作は不要である。さらに、ステージイン・アウト時には、ASNAROサーバの制御によって、ファイルサーバ側でファイルの圧縮・解凍処理を行うため、利用者が圧縮・解凍処理を意識することなく、共有サーバの容量節減が可能である。

ここで重要なことは、ASNAROのジョブ投入機能は、単にプログラムを実行し、共有ファイルサーバにステージングするだけでなく、図に示すように、入出力ファイルや重要な入出力値を自動的にメタ情報として、データベース化するところである。そのメタ情報を使うことで、共有ファイルサーバに蓄積される結果データは、統一した形で自動分類されるので、誰もが利用しやすいように整理された形で、解析/シミュレーションの結果の共有が実現されるのである。

ちなみに、1つのジョブを投入する機能をパッケージ化したものをASNAROではコンポーネントと呼んでおり、ジョブを投入するシェルスクリプト、プログラムの入力ファイル、ファイルのどの部分を変数化するかを指定するルールファイル、どの計算機で動作させるかを定義するだけで、つまり、特殊な専用言語を覚える必要なく、作成が可能である。また、このコン

ポーネントは、共有設定ができ、複数人で共有が可能になっている。

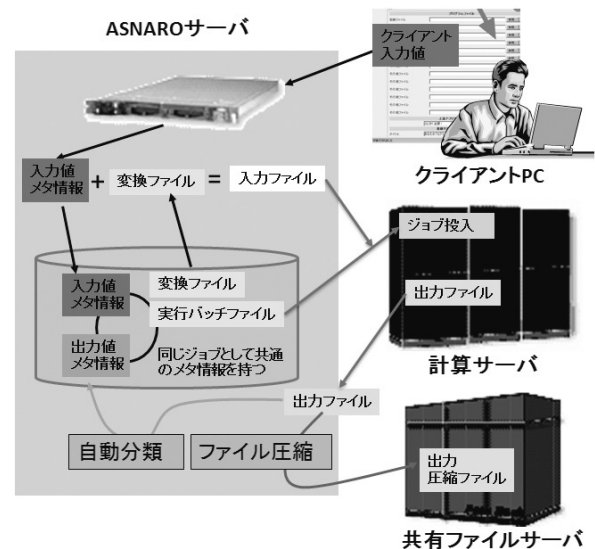


図3 ジョブ投入機能

4.2 ワークフロー機能

4.1では、1つの処理についてのASNAROのジョブランチャーのジョブ投入機能を解説したが、業務は通常それだけにとどまらず、例えばメッシュ生成～解析/シミュレーション～統計解析、あるいは可視化処理等の複数段の処理の流れがある。それらは、連続的に処理が行われる必要があり、上流の出力結果は、下流の入力データとなるため、実行前に確実なファイルの引き渡しなども必要である。

また、ジョブ投入機能を使いこれらの処理を1つ1つ実行していくと、ファイルの引き渡しなどの面倒な手作業が発生するだけでなく、処理間のつながりの情報が作業者の記憶やメモにしか残らなくなる。つまり、最後のグラフ処理をした結果の図を見る時に、「どのメッシュデータの解析だったのか?」、「どのような境界条件で計算した結果なのか?」をはっきりさせることは、作業員以外では、ほとんど困難で、作業員自身でも作業から時間が経ってしまえば、絶望的な状況になってしまう。

ASNAROのジョブランチャーは、異なる計算機間において、ジョブ間のファイル、情報引き渡しを制御しながら、複数のジョブを逐次実行していくことを可能にするシングルサインオン・ワークフロー機能を有し、これら問題を解消している。図4は、これら機能を有するASNAROのジョブランチャーの利用画面である。上部の表様式の部分は、各ジョブを投入するときのパラメータを設定する部分で、その下部がジョブの実行順序をあらわすワークフロー図である。ジョブ投入の基本単位であるコンポーネントは、緑色の四角で表示されており、それらをつなぐことで、ワークフローを構成している。

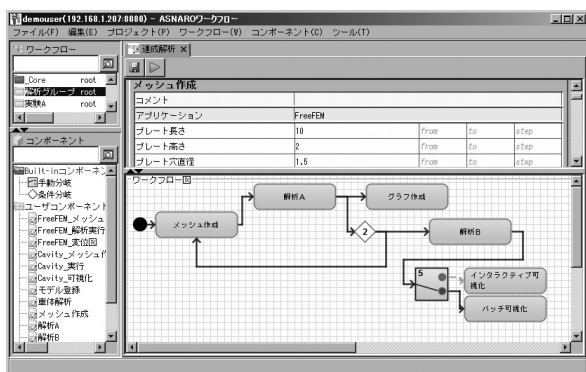


図4 ワークフロー作成画面

当然、ワークフローでは処理と処理の間で条件判断やループ、手動での処理の切り替えや並行処理が流れの中で要求されるが、ASNAROではそれらの機能を有したビルトインコンポーネントが予め用意されている。

また、解析/シミュレーションや可視化処理が常に同じマシンで行われるわけではなく、さらに各マシンのユーザアカウントは様々なことも多いはずである。ASNAROでは、ワークフロー実行時に異なるアカウントであっても、各マシンへのログイン処理を自動で処理することで、サーバごとのログイン操作の手間を省くことができるシングルサインオン機能を有している。アカウント認証方式は、通常のID/Password方式のほか、秘密鍵-公開鍵方式にも対応している。

4.3 スキーマレス非構造データベース機能

解析/シミュレーションや研究業務のすべてをデータベース化して共有することを考えた場合、その情報は可変、多種、多岐にわたる値、キーワード(メタ情報)があり、それも時々刻々とその種類が増えていくことに、追従できる必要がある。

一般的に、データベースといえは、リレーショナルデータベースが広く使われている。これはデータを表に似た構造で管理し、複数の表(データ群)の相対関係を予め定義し、その構造に合わせてデータを格納していくモデルである。しかし、データベース化を開始する前に、スキーマと呼ばれる表の各カラムに格納する値と表間の同一カラムの紐づけ情報を定義しておく必要があり、取り扱う情報が固定化されている受発注管理情報や人事情報、物品管理情報などの事務系業務には適しているが、解析/シミュレーションのように扱うモノや解析手法により、記録する情報(メタ情報)が様々である業務には、適していない。

実際にデータベース化の後にメタ情報の種類が増える度、データベースの運用を止めてスキーマの変更が必要であり、スキーマ変更は、その変換作業だけでも数日、変更前後でのデータ整合性確認まで含めると週間間にわたる作業が発生することもある。つまり、解析/シミュレーションのようにメタ情報の種類が頻繁に増える業務を対象にした場合には、頻繁な運用の停止

が必要になり、現実的なシステムにならない。これが今まで解析/シミュレーション分野にデータベースが持ち込まれなかった大きな原因である。

ASNAROでは、この課題を解決するために、スキーマを必要としないスキーマレス非構造データベースを内蔵しており、設計が不要で利用者がメタ情報の定義を自由に設定できるシステムとなっている。

スキーマレス非構造データベースでは、情報やファイルを分類するためのメタ情報を付与することで検索が可能となる。ASNAROはパターンマッチングエンジンを内蔵しており、テキスト中のトークンを特定し、文字列を抽出したり、置き換えたりする機能を有する。そして、解析/シミュレーションのパラメータファイルや結果ファイルから分類、検索に有用な、あるいは後に結果の比較で利用しそうなパラメータをメタ情報として自動でデータベース化する機能を有しており、計算結果に1つ1つメタ情報を手動で追加する必要はない。計算が終わればメタ情報が付いた結果が共有ファイルサーバにどんどん蓄積されるのである。これで、この可視化結果がどのような計算を経て出力したものであるかを見失うことは、なくなるはずである。

また、ASNARO内部では、データベース要素がXML形式で格納されており、XMLの1タグが1つのメタ情報に対応し、1タグレベルで読み出し、書き込みのアクセス制御権を詳細に設定できる。アクセス制御権では、全員または個別アカウントという指定以外に、ユーザを複数まとめたグループを指定することもできる。例えば、設計グループは解析/シミュレーションのメタ情報にアクセスできるが、議事録は見えないという設定など、細かな情報共有設定が可能である。

4.4 その他、解析プロセス、結果の共有、利活用に重要な機能

(1) 検索結果表形式表示機能

解析/シミュレーションの検索結果は、どのように表示されると見やすいだろうか？皆さんは、階層構造フォルダの中に結果ファイルだけが並んでいるViewに慣れているかもしれない。また、「検索」といえば、googleの検索結果表示画面のようなViewもよく使っているかもしれない。しかし、解析/シミュレーションの結果を確認、比較をする場合、それらのViewで必要な情報を直ぐに把握、比較が可能だろうか？

同じようなファイルがたくさんあり、それも紐づけられた関連ファイル、パラメータがたくさんある解析/シミュレーションの結果は、従来の検索表示Viewでは、結局、解析メモを確認しながら、フォルダやファイルをたくさん開かなければ、結局、結果の把握、比較ができないことは、明白である。

ASNAROでは、そのような問題を解決するために、検索結果を紐づけられたその上流コンポーネントの結果のメタ情報、ファイルともに表形式に表示するViewer (ASNARO TableView) を有している。つまり、グラフの図を検索で絞り込み表示をした時に、その上

流処理であるメッシュや解析のパラメーターも検索結果の同じ行内に表示され、簡単に確認できるのである。

また、このViewerは、カラムごとのソートやフィルタ機能を有しているため、結果データの計算条件を絞り込んだり、他の結果と比較することも簡単である。計算結果がグラフやイメージの場合は、図5のようにサムネイル表示になるので、直感的な比較にも役立つはずである。また、結果に関する簡単なコメントなどもメタ情報で追加できるため、結果の情報整理や意見交換などの役割も果たせるのである。

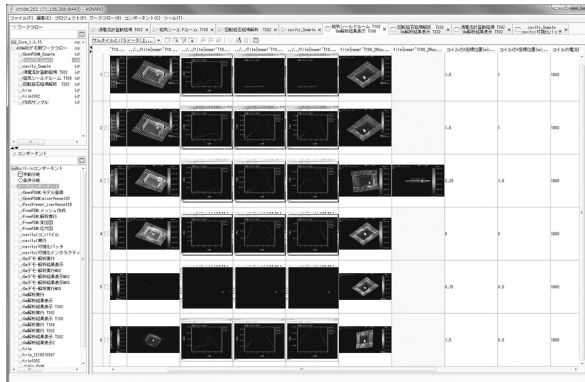


図5 検索結果のサムネイル&パラメータ表示

(2) 仮想フォルダ表示機能

一般的な共有ファイルシステムでは、単一の階層構造の上でフォルダ名やファイル名に規則性を持たせることで、データを整理し、必要な時に探し出せるように工夫がされているが、登録ファイル数やファイルサーバの利用者が増えるにつれ、フォルダ階層は深くなりすぎたり、命名規則の一貫性が担保されなくなるとデータの探査性が低下したり、登録者以外の人には探せない・活用できないデータ管理状態に陥ることがある。

ASNAROは、既存の固定的なフォルダ階層構造に加え、ファイルに付与されたメタ情報を使い、条件に合致したファイルをフォルダ内に表示する仮想フォルダ階層表示機能を有しており、ユーザーが任意の仮想的なフォルダ階層、及び階層順序を作り、既存のフォルダ階層構造とは異なる階層表示で、それに合致するファイルを表示することができる。

図6のように、仮想フォルダ階層表示機能は自由に複数のフォルダ階層構造を設定することができるため、様々な観点からファイルを分類(横串分類)することが可能である。

このような表示機能を活用することでデータの探査性が向上し、「保存したはずのファイルが見当たらない」、「見つけれない」、「このデータの計算プロセスが不明」といった現場が抱える問題を解決し、過去のデータの二次利用を促進することができるのである。

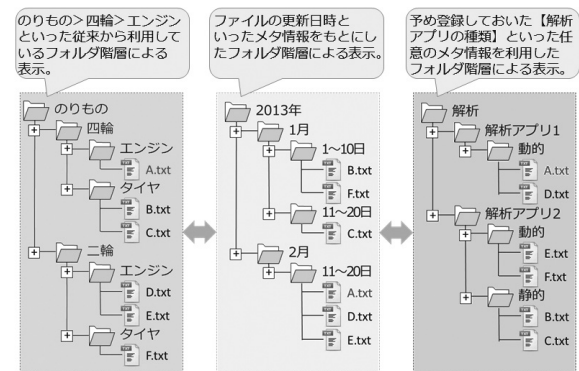


図6 仮想フォルダの概念

(3) 一括メタ情報付与機能

研究、業務として実施する解析/シミュレーションでは、出力データ以外に、その解析/シミュレーションに付随する計画書や考察の会議録、報告書などを紐づけて共有・管理することも重要である。

そのためには、これらの関連データが簡単に効率よくアップロード・ダウンロード及び簡易表示でき、解析/シミュレーション・データと同じメタ情報を意図的、かつ簡単に付与し、それを用いた高速な検索可能な機能を実装することを考えなければならない。

しかし、現実的には、メタ情報の登録作業を手動で行うことは面倒であり、誰も進んで登録しようとは思わず、また、登録時の入力操作ミスも避けられない。

ASNAROでは、予めフォルダにメタ情報を定義づけることができ、図7に示すようにフォルダにファイルをドラッグ&ドロップするだけで、複数のファイルに複数のメタ情報を簡単に追加できる機能を有しており、その問題を解決している。

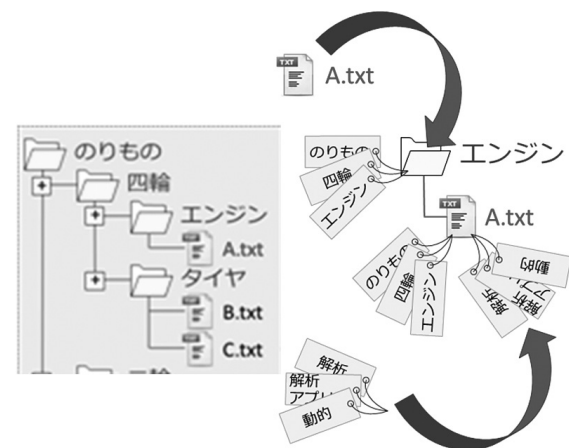


図7 メタ情報の追加機能

また、ワークフロー実行前に結果ファイルに付与するメタ情報を手入力することは可能だが、仮想フォルダに対応している複数メタ情報をワークフロー実行

画面にドラッグ&ドロップすることで一括付与する機能をASNAROは有しており、作業効率を大きく向上させるとともに手入力による誤入力や属人的なキーワードの使用を抑制できるように工夫されている。

5 おわりに

本稿では、解析/シミュレーション・データの共有化を実験・計測分野や一般社会で行われているレベル以上に引き上げ、解析/シミュレーションのあり方を革新することの重要性を説き、そこに至る問題点、必要な機能要素^[7]について、弊社製品であるASNAROを例にとり、説明した。

しかしながら、解析/シミュレーションプロセス、結果の共有化、利活用は、その重要性の認識がまだまだ不十分であり、取り組みの機運も高くなっていない。本紹介が、それらの重要性を正しく認識し、解析/シミュレーション研究、業務のきっかけになればと願っている。

■参考文献

[1] 上島豊、超並列計算機を使った超大規模光量子シミュ

レーションの現状と課題、シミュレーション Vol.70, pp.261-270, (2000)

[2] 上島豊、奥田洋司、他、ベタフロップス コンピューティング、pp.183-202, 培風館、(2007)

[3] マルチパラメータサーベイ型シミュレーションを支えるシステム化技術に関する研究、iic-HPC 北海道大学情報基盤センター大型計算機システムニュース VOL.22, pp.4-9, (2011)
http://www.hucc.hokudai.ac.jp/iic-HPC/2011_22.pdf

[4] 村上慎吾、鈴木真悟、上島豊、野村泰伸、倉智嘉久、薬物誘発性不整脈の発生危険度予測システムの開発、生体医工学 48, pp.106-114, (2010)

[5] K. Nishihara, Y. Ueshima, et al, "e-Science in high energy density science research" Fusion Engineering and Design Volume 83, Issues 2-3, pp.525-529, (2008)

[6] 上島豊、西原功修、XMLデータベースを用いた実験・シミュレーション統合Webシステム、航空宇宙数値シミュレーション技術シンポジウム講演集、40th-2008, pp.155-158, (2008)

[7] 上島豊、前田茂樹、犬飼貴史、江口和宏、石原典雄、PaaS-CAE基盤技術に関する研究開発、計算工学講演会論文集 Vol.16, F-5-4, (2011)



(P10より続く)

1982年に、前述したHallquist博士を訪ねた後に、もう一箇所Ann Arborにあるミシガン大学に菊池昇准教授(当時)を訪ねた。有限要素法の話で3時間ほど話をしたが、とても興味深かったのは、「有限要素法の品質管理が出来るかも知れない」という話だった。

この話が忘れられず1984年に再度菊池先生を訪ね、「あの時に伺った有限要素法の品質管理はその後進展しましたか？」と尋ねた。すると当時博士課程の学生だった鳥垣さん(現、日産自動車)が、中心部に亀裂がある四角い平板の両端を引っ張った時の亀裂先端の応力が、元のメッシュ形状に較べ、メッシュトポロジーを一切変えずに3倍になったというグラフを見せてくれた。計算は弾性を仮定しているので、亀裂先端部の応力は無限大になるのだが、限られた数の要素分割でも、節点の座標位置を変えるだけでこれ程の違いが出ることに驚きを覚えた。誤差解析を基に、評価した誤差の測度に従い、節点の座標値を移動させ、近隣との相対誤差を最小にする方法だとの説明を受けた。他に、同じ問題で、評価した誤差の大きな要素を細分割する方法も見せていただいた。これは物理量の変化が急なところはどんどん細かくするので、当然精度は上がるだろうけれども、自分は前者の方法に深い感銘を受けた。

そこで、「このプログラムを日本で販売しても良いですか？」と菊池先生に打診してみた。

(P31へ続く)

