

◆連載◆

《連載 最終回》

R & D 部門における データ共有システム構築の事前準備の要所 (下)

上島 豊 (株) キャトルアイ・サイエンス 代表取締役



《PROFILE》

略歴：

1992年 3月 大阪大学工学部 原子力工学科 卒業
1997年 3月 大阪大学大学院工学研究科 電磁エネルギー工学専攻 博士課程修了
1997年 4月 日本原子力研究所 博士研究員
2000年 4月 日本原子力研究所 研究職員
2006年 3月 日本原子力研究開発機構 (旧日本原子力研究所) 退職
2006年 4月 キャトルアイ・サイエンス設立 代表取締役 就任

主な参加国家プロジェクト：

文部科学省 e-Japan プロジェクト「ITBL プロジェクト」、「バイオグリッドプロジェクト」
総務省 JGN プロジェクト「JGN を使った遠隔分散環境構築」
文部科学省リーディングプロジェクト「生体細胞機能シミュレーション」

主な受賞歴： 1999年 6月 日本原子力研究所 有功賞「高並列計算機を用いたギガ粒子シミュレーションコードの開発」
2003年 4月 第7回サイエンス展示・実験ショーアイデアコンテスト文部科学大臣賞「光速の世界へ招待」
2004年 12月 第1回理研ベンチマークコンテスト 無差別部門 優勝

主な著作： 培風館『PSE book—シミュレーション科学における問題解決のための環境 (基礎編)』ISBN：456301558X
培風館『PSE book—シミュレーション科学における問題解決のための環境 (応用編)』ISBN：4563015598
培風館『ベタフロップス コンピューティング』ISBN978-4-563-01571-8
臨川書店『視覚とマンガ表現』ISBN978-4-653-04012-5

5 属人的データ共有状況を脱するための 事前準備 (システム化対象特定編)

データ共有・利活用のシステム化及びシステム運用には必ず負担がかかる。したがって、闇雲にいろいろな作業をシステム化の対象と考えるのではなく、作業に品質(網羅性、均質性、再現性)が重要なもの及び作業の発生頻度と作業にかかるネット時間(研究者が実質的に作業をしている時間で、待ち時間や他の作業をしている時間を除く)の大きいものを序列化し、その上位の作業からシステム化対象にあてるべきである。実際、この序列は現業務の問題の深刻度とも対応するはずで、もし、現状深刻な問題を感じている業務が上記序列で上位にあがってきていない場合は、頻度や作業時間の算定が間違っているか?主観的な思い込みのバイアスがかかっているかのいずれかである。複数の人と意見交換することや再計測をすることで、その認識ずれは解消しておく必要がある。

システム化対象が決まったら、次はどのようにシステム化を行うか検討をしていく。この段階になって、初めてソフトウェア的観点を加味して考えることになり、システム開発関係者も参加して意見交換を交わすことになる。この時、注意すべき点はいくつかあるので、こ

こで触れておく。あくまでシステム化の主目的は、データ共有・利活用に対して、網羅性、均質性、再現性という品質を担保するためのものである。もちろん、作業の発生頻度×ネット時間が大きいものもシステム化の対象にはなるが、安易に作業負担をソフトウェアで軽減させよう考えてはいけない。通常「網羅性、均質性、再現性」を維持した手動作業より、システム化した場合の作業のネット時間が大きくなることは考えにくい。まずは「より使いやすく」、「より便利に」ということは考えずに、手順が確実に実行できるかどうかだけに焦点を絞ってシステム化を考えるべきである。つまり、システム化に便乗して「より楽に」ということを考えてはいけないということである。業務系システムでは、手順が明確になっており、手順の実行頻度も安定的に固定化されている。また、業務系システムの変更は簡単に実施できないので、1回のシステム導入や更新でできるだけ使用者が「より楽に」作業遂行できるように配慮する。しかしながら、R & Dでのシステムでは、R & Dの宿命上、多くの手順変更や新たな手順が発生し続ける。つまり、できるだけ迅速に手順追加や手順変更に対応できることが最優先課題で、「より楽に」、「より便利に」ということは次点以下の課題なのである。実際、「より楽に」、「より便利に」を考慮しだすと結局、運用、メンテ

ナンスが追い付かなくなり、本末転倒な状態になる。手順と同等のことが同等時間以下の作業で達成できれば、もれなく、網羅性、均質性、再現性という品質がついてくるので、これを大きな価値とする考えを持つべきである。実際、R & D 部門で、データ共有・利活用がシステムでうまくいっているところはすべて、このような方針が徹底された上で、開発・運用を行っている。

実際、「より楽に」、「より便利に」は、客観的でない（主観）欲求であることがほとんどで、客観的負担対効果を蔑ろにする考えであり、システム開発、利用に行き詰まりを生む元凶である。「より楽に」、「より便利に」という要求があった場合、あくまで、手で共有・利活用をした時と比べて、どれだけ省力化できるかどうかという客観的指標で考えるようにする。そして、この客観的省力化量がシステム化の優先序列であり、省力化量の小さいものは、客観的負担対効果は小さく、取り扱うべきでないという判断をすべきである。もちろん、「より楽に」、「より便利に」の部分があるのであれば、それは採用されてしかりである。

システム化に便乗して、今まで行っていないデータ探査、データ処理をこの機にシステムに実装してしまいたいという要望が挙がるがよくある。しかし、これは絶対に行ってはならないことの一つである。以下、なぜこのようなことは行ってはいけないのかを説明しておく。

5.1 第三者が手動処理で確実に実施できる手順書がない

システムは、オルゴールの外箱であり、楽譜があって、それをドラムに刻み込まなければシステムは完成しないことは上述している。手順書がないということは、楽譜がないということである。楽譜なしに音楽をドラムに刻み込むのは不可能であり、また、即興で楽譜を書いたとしても天才でない限り、まともな音楽になっていることは奇跡である。実際、すでに手動実施していることを手順書に書き下すことでさえ、悪戦苦闘している状況を考えると、如何に現実的でないか理解できると思う。当然、そのような楽譜や音楽は、オルゴールに組み込むべきではない。

5.2 その作業が他の作業よりシステム化すべきか客観的に判断できない

即興で実際に実施可能な手順書が書けたとしても、「その手順通り実施した結果が価値の高いものか？」や「その実施頻度がどの程度か？」は、判断がつかないということである。つまり、そもそもシステム化対象を決める序列の「作業に品質（網羅性、均質性、再現性）の重要性」、「作業の発生頻度×ネット時間」の正確な値が算出できず、他のものと客観的に比較ができないのである。

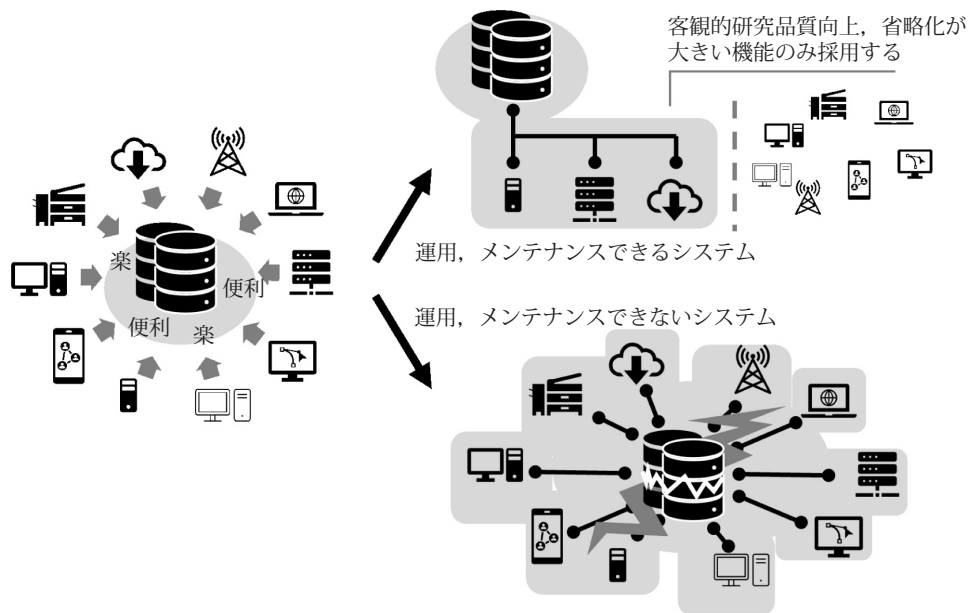


図 12 「より楽に」、「より便利に」が目標であってはならない

「今まで行っていないデータ探査、データ処理をこの機にシステムに実装してしまいたいという要望」の中で、ツール自体もまだ存在していない場合もある。当然このような作業のシステム化は、上記理由より行うべきではないが、「ツールが存在していない」状態で、システム化を行ってはいけないもう一つの重要な理由を以下で説明しておく。

5.3 ツールとシステムの境界が曖昧になる

R & D 部門では項目追加、変更だけでなく、データ処理に関しても追加、変更が頻出する。これが業務系と異なり、システム化を難しくしている原因である。このような R & D 部門でもシステム化を可能にするためには、ツールとシステムの分離が重要で、システムからツールを呼び出す形のシステムでなければならない。そのようにすることで、システムなしにツールを使い、手順を確立することが可能になる。しかし、ツールとシステムを同時に開発すると、手順の確立ができなくなってしまふ。また、同時開発はツールとシステムの境界を曖昧にしてしまい、その結果、ツール部のみの改良やシステムを改良したときにツール部に問題が生じないかの確認、検証が難しくなる。つまり、ツールやシステムのバージョンアップ時にシステム自体の長期の停止が必要になるなど、メンテナンス性が悪くなるのである。

5.4 ツール部、システム部及び手順の問題の切り分けができない

ツールとシステムを同時に開発すると、何か問題があったときに、手順に問題があったのか？、ツール仕様の問題があったのか？、システムに問題があったのか？の切り分けが非常に難しくなる。そもそも、システムを使わない状態での正解が示されないとその判別自体が原理的にはできないということになる。システムの問題であることを排除するためには、手動でツールのみを使った手順の確立が必須である。もちろん、ツールの問題であることを排除するためには、ツールを使わない手順の確立が重要である。したがって、ツールを自作する場合には、他の複数の商用ツールなどでも同じことができる手順を確立しておくべきである。また、システム開発会社にすべてを丸投げしてしまう場合は、丸投げ体質

が染みつき、もし、ツール部やシステム部に問題があっても、気づかず間違った分析をしたままになる可能性が高くなる。実際、正確な仕様書と手順書がないとシステム会社も何が正しいのかを把握していないことになるので、わからないところに丸投げをするという非常に危機的な状態に陥る。

「それでは、将来展望など、現時点でできないことを実現するためにはどうすればいいのか？」という質問がよくある。すでに答えは上述しているが、以下でまとめておく。

1) データ処理を行うツールを開発する

データ処理を行うツールが存在しないのであれば、まず、手順での実施を可能にするためのツール作成を行う。ツールの動作を確認する為のサンプルデータとその結果も整えておくこと。実際、ツールに BUG がないことを保証するのは、相当大変ですが、システムに組み込む前に十分確認しておかないと、皆がツールには問題がないとして使いだすので、十分注意が必要である。もし、ツールをリリースした1年後などに問題が見つかった場合は、その連絡および関係部署への影響範囲の確認などが発生し、結構大きな問題になる。

2) 第三者が手動処理で確実に実施できる手順書の作成

ツールを使って、手動で実施可能な手順を確立し、手順書を作成する。また、この過程で、データの項目名や項目値の単位がばらばらだとか、数値の項目値に数値以外の値が入っている(> 1, ~ 10, // とか、目視では意味が分かるが、システムには意味判別ができない記述) もの、適切な数値に修正をしたり、新たな項目(下限値, 上限値, 信頼幅など)を作り、その値に変更するなどの作業をする。この作業は一般的にデータクレンジングと呼ぶ。



図 13 手動で行えていないことをシステムに組み込むまでの正しい手順

6 おわりに

本記事では、R & D 部門におけるデータ共有システム構築の事前準備として、まず、属人的データ共有状況を脱するための具体的な手順を説明した。次に、R & D 部門におけるデータベース、システムは、決して魔法の箱ではなく、システムもしくはデータベースがない状態で実施できることに網羅性、均質性、再現性を付与し、実施できることであることを説明した。最後に、どの部分をシステム化すべきかを特定する方法に関して、説明を行った。読んでいただいておりますが、R & D のデータ共有を行なうということは、何かを買って、解決するという問題ではなく、企業風土及び R & D 従事者の意識変革が必要である。R & D 従事者は、業務内容が極めて専門的で、会社内でも対話する機会が少ない人たちだと思う。研究者は、とっつきにくいところはあるが、地頭は良く、非常にまじめなので、これを機に話をしてみれば、仲良くなれるかもしれない。そういう人とのつながりを作ることが、R & D 部門におけるデータ共有の第一歩である。