

# R & D 部門における効果的なデータ管理手法と 管理体制の作り方（下）

上島 豊 （株）キャトルアイ・サイエンス 代表取締役

《PROFILE》

略歴：  
1994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ガ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  
「光速の世界へご招待」  
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 ー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 ー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 効果的なデータ管理手法と管理体制 の作り方

前章では、「属人的な、もしくは形骸化したデータ管理」状況を生み出す原因を明らかにしました。本章では、その原因を取り除き、利用、活用されるデータ管理システムを実現する為の効果的なデータ管理手法と管理体制に関して、具体的な対処方法を提言します。

### 5.1 「実験及び解析を再現するのに十分な項目」の調査、分析

研究者自身が「実験及び解析を再現するのに十分な項目」を意識でき、それをグループなどで話し合い、合意形成を可能にする状態にできるのが一番好ましいのですが、研究者の特性上、難しい場合が多いのも事実です。そのような場合は、研究グループ内に支援者という名目でファシリテーターを配置し、「実験及び解析を再現するのに十分な項目」の調査、分析を行なわせる体制を作ることをお勧めします。「実験及び解析を再現するのに十分な項目」がわからない、再現性の難しい「実験及び解析」も中にはあるはずですが、ここでいう「実験及び

解析を再現するのに十分な項目」は、研究者自身が意識的に制御できていないものも含めて追及するという意味ではなく、意識的に制御しているものだけを意味します。もちろん、「実験及び解析」の理解が進むことで、意識的に制御できるものが増えてくるはずで、そういうものは随時追加できるような取り組みをすることが重要です。

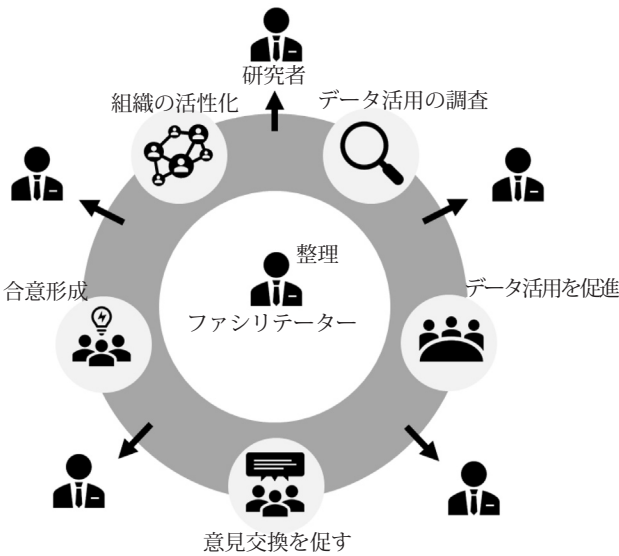


図4 ファシリテーターによる現状のデータ利用の実態や問題を把握、改善

## 5.2 「現状どのように過去データを探しているか？」 の調査、分析

研究者自身が「どのように過去データを探しているか」を意識でき、それをグループなどで話し合い、合意形成を可能にする状態にできることが一番好ましいですが、研究者の特性上、それが難しい場合が多いのも事実です。その場合は、研究グループ内に支援者という名目でファシリテーターを配置し、「どのように過去データを探しているか」の調査、分析を行なわせる体制を作ることをお勧めします。もちろん、研究の進め方が変わることで、「どのように過去データを探しているか」も変わってくるはずで、そういう変化に追従できる取り組みをすることが重要です。

## 5.3 「蓄積されたデータをどのように利用するか？」 の調査、分析

「蓄積されたデータをどのように利用するか？」は、データ管理システムを作る前に、入念に調査する必要がありますが、システム構築をする関係者が研究者自身に直接ヒアリングをしても多くの場合、適切な情報が入手できません。研究者は、自分が現在どのようにデータを利用しているのか？、データが蓄積されたらどう利用したいのか？を意識的に考えていることは稀で、多くの場合漠然としたイメージでしか意識できていません。また、システム構築の関係者は、システムを構築することに前のめりになるため、研究者の言葉を直接的に鵜呑みにしてしまうことが多く、本当にその利用方法で研究者の課題が解決するのかまでの考察ができません。実際、研究者が行なっていることを、システム構築関係者が理解、把握することが専門領域的にも難しいのも事実です。したがって、前節で述べた研究グループ内に支援者という名目で活躍できるファシリテーター的な人材を調査、分析の任につけることも検討する必要があります。さらに、本調査、分析はデータ管理システムを作る前だけでなく、システム運用後も継続し、システムの改善、改良に役立てる必要があります。R & D部門は他の業務部門と異なり、多種多様、経時変化していくものなので、そういう変化に追従できる取り組みをすることが重要です。

## 5.4 「実験及び解析を再現するのに十分な項目」を 確実に記録させる体制

グループ内で「実験及び解析を再現するのに十分な項目」を記録する重要性について話し合いの場を持ち、自律的に確実に記録できるようになるのが一番良いのですが、そうならないことの方が多いはずですが、そのような場合は、研究グループ内に支援者という名目でファシリテーターを配置し、「実験及び解析を再現するのに十分な項目」が記録されているかどうかを確認し、足りなければ研究者に追記を依頼したり、ヒアリングをして自らが追記する体制を作ることをお勧めします。実際にこのような体制を採ることは、誤記録や無意識の記録漏れ、データ捏造の抑制や実験及び解析の進捗状況の把握にも有効です。

海外では、徹底した分業体制を採り、分業間でデータ引き渡しをするポイントにおいて責任の所在を明確化することで、各自の責任部分の項目を記録する確実性を担保しています。この方法は、分業部分毎の作業を明確化することで、業務の標準化と人材の配置転換の容易性を高くできる半面、各人の創意工夫の自由度や経験値を十分に生かしきれないというデメリットもあります。日本の多くの企業では、良くも悪くも責任の所在の明確化や徹底した分業をせず、各人の創意工夫の自由度や経験値を十分に生かした業務の取り組み方法で、効果的なR & Dを進めてきました。しかし、「各人の創意工夫の自由度や経験値を十分に生かす」ために、データ管理が属人的になってしまっており、昨今のデータが大量に発生するR & Dでは属人化のデメリットがメリットを上回ってきています。

ファシリテーターを配置しても改善できない、ファシリテーター自体を配置できない場合は、海外のような徹底した分業体制を取り、責任の所在の明確化と責任追及をする体制への移行も考える必要があります。

## 5.5 システム利用を推進する体制

単に「利用してください」とお願いをしたり、「どんな機能が必要ですか？」とヒアリングをしても、効果的にシステム（蓄積されたデータ）の利用が推進されることはありません。5.3節でも述べましたが、研究者は、自分が現在どのようにデータを利用しているのか？、

データが蓄積されたらどう利用したいのか？を意識的に考えていることは稀で、多くの場合、漠然としたイメージでしか意識できていません。そのことを前提にして、システム利用を推進しようとするのであれば、研究者のどのような課題がシステムで解決するかを考え、課題がどのように解決できるかを提示し、利用へ誘導していく必要があります。また、その過程で、現在のシステムで解決できない課題が分かり、それを解決するためにはどのような機能があるべきかを考えることで、システムの拡張方針が決まることもあります。

システムの利用推進や運用に関する体制には、もう一つ注意が必要な側面があります。運用者が研究者の行なっていることを理解、把握できなければならないことは、言わずもがなですが、理解、把握をしていたとしても、業務が完全に分離されている状況では、次第にそれぞれの立場に偏ってしまいます。その結果、研究者は重要度が高くない機能を要求し、運用者はそれを漫然と受け容れるか、完全に否定的な態度をとってしまう状態に陥りがちになります。これらの状況を打開するためには、人事交流による役割のローテーションが有効です。研究者を推進運用側に異動したり、兼務することで、相手の立場、観点を意識できるようになり、特定の役割に偏り過ぎた認識が抑制され、効果的なシステム改善が可能になります。

システムの利用を推進するためには、運用体制だけでなく、利用者の体制に関しても工夫が必要です。データ共有、管理ができていない状況では、当然ながら研究者は自分以外の人が生成したデータを使うことを前提として、研究を進めていません。しかし、データ共有、管理できる状態になり、いくら便利になったとしても、急に研究の進め方を変えられるわけではありません。もちろん、一部の好奇心が旺盛な研究者は、これを機に研究の進め方を変えるかもしれません。研究者は研究対象に対して好奇心は旺盛ですが、それ以外の部分に関しては、人並み以上に保守的な人が多いのも事実です。したがって、せっかく構築したデータ共有、管理システムを最大限効果的に活用するためには、従来の研究から離れて、他の研究者のデータのみで、研究を行なう専門部隊を立ち上げ、システムを最大限効果的に活用し、成果を出すことを彼らの使命にするのです。彼らは、自身で実験を行わないので、自分たちの成果をあげるために、有望そうな実験の提案やそのための説得資料作り、予算取りなども自律的に行なうようになるはずです。

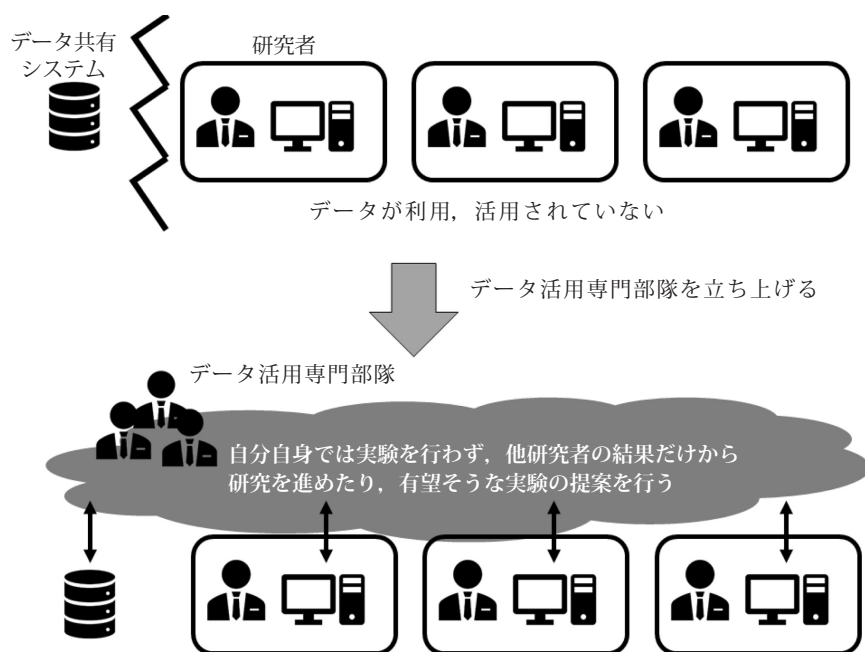


図5 データ活用専門部隊によるデータ活用と既存研究者の活性化

## 5.6 R & Dの変化に追従できる項目及びフォーマットのデータ記録シートとシステムの採用

ここまでの節の課題がクリアできると、データ管理のためのシステムを構築するための山の9割がたは越えており、後は適切なシステムプラットフォームを採用すれば問題はありません。つまり、R & Dの変化に追従できる項目及びフォーマットを簡単にメンテナンスできて、それらをベースにデータ登録とデータ利用ができるシステムプラットフォームを採用するということです。しかし、それを実際探すとすると結構大変です。なぜなら、一般的なシステムは、固定的な項目及びフォーマットを扱い、データの利用方法も画一化されたもので、R & Dのフレキシビリティを受容できるものはほとんどありません。逆に言うと、普通は、固定的な項目及びフォーマットで利用者が同じ作業をするものに関して、効率化、確実化のためにシステム化は実施されるのです。R & Dのデータ管理システムは、既存のシステムプラットフォームの考え方からは、そもそも対象外ということになります。以下で、R & Dプラットフォームとして備えるべき特性を、より具体的に列挙しておきます。

### (1) R & Dの変化に追従できる項目及びフォーマットを簡単にメンテナンスできること

⇒構造化されたデータを格納できながらフレキシブルである必要があります。従来の構造化データベース（リレーショナルデータベース）や単純な非構造データベースでは対応できません。実際に多くのR & D部門への導入実績のあるR & D専用のシステムを採用すべきです。また、R & D部門への導入実績といっても、生産に近いところや業務的に定型的なところのみにしか導入していないこともありますので、R & Dのどのような部署への導入かまで、しっかり確認したほうが良いと思います。

### (2) データ記録はシステムから切り離されたファイル（エクセル等）で記録できること

⇒実験の最中にシステムにログインすることは現実的ではなく、記録段階では、システムと切り離されていても使える必要があります。しかし、項目名及びフォーマットは、用語ブレを抑制するためにシステムに登録されたものを参照する必要があります。また、利用者に使い慣れた

インタフェースであることも重要です。

### (3) 検索結果のデータを他のサーバ、PCの既存アプリケーションに処理を受け渡せること

⇒業務系のデータは、多くの場合、検索結果は表形式で表示できればこと済みます。また、グラフなどを書く場合でも定型的なので、システムに機能を埋め込むことも可能です。一方、R & Dでは、検索結果は表形式で表示できればこと済むことはほとんどなく、多くの場合、データごとに専用の市販アプリケーションや自作の簡単なアプリケーションでデータ処理を行います。したがって、R & Dのシステムでは、検索結果のデータを他のサーバ、PCの既存アプリケーションに処理を受け渡せる機能を有していることが必須です。

### (4) データの探査方法や利用方法をシステム停止無しに容易に追加することが可能なこと

⇒R & Dでは、研究内容が変わるごとに、データの探査方法や利用方法が変わってしまい、それに合わせたシステムの拡張が必要になります。そのシステム拡張ごとにシステム全体の停止をしていると既存の利用者の業務に支障が出てしまいます。したがって、データの探査方法や利用方法をシステム停止無しに、容易に追加することが可能である必要があります。

### (5) 個別データ毎にデータベースレベルでアクセス制御ができること

⇒R & Dでは、扱うデータの機密性が高く、同じ研究グループとさえい、データを共有できないこともあります。一方、R & Dではシステム拡張が頻繁に行なわれるので、その中で一定レベルの不具合も発生することは、避けられません。そのような状況においても、アクセス制御に問題があってははいけません。したがって、アクセス制御はシステムの拡張部分によって制御されるのではなく、システムプラットフォームとして、データベースレベルでデータベースの最小単位レベルで制御されている必要があります。



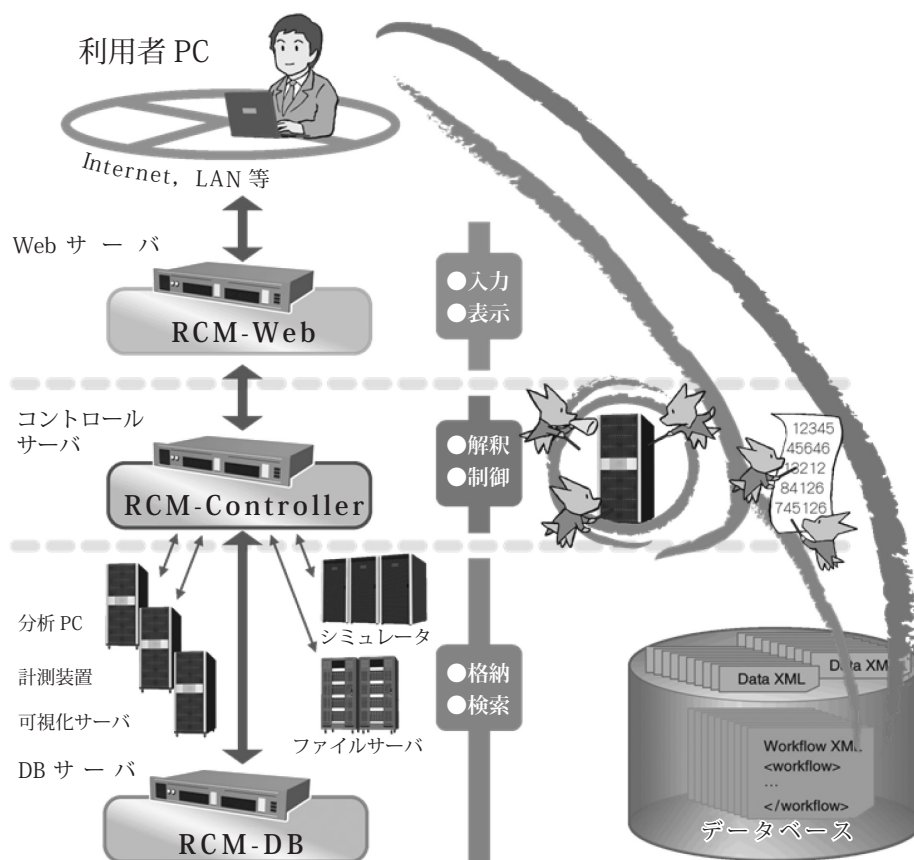


図6 R &amp; Dの変化に追従するために必要なシステム

## 6 おわりに

本記事では、R & D 部門におけるデータ管理の実情が属人的、形骸化したデータ管理となっている実態を紹介し、その問題点及び原因を検討しました。それらを基に効果的なデータ管理手法と管理体制の作り方に関して提言をしました。読んでいただいておりますが、R & D のデータ管理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す。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います。研究者は、とっつきにくいところはあるかもしれませんが、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ません。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管理の第一歩です。是非、皆様頑張ってみてください。

## 参考文献

- 1) 川田重夫, 田子精男, 梅谷征雄, 南多善, 上島豊, 他 PSE book  
ー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),  
川田重夫, 田子精男, 梅谷征雄, 南 多善 共編, 培風館, (2005),  
p69-82
- 2) 谷啓二, 奥田洋司, 福井義成, 上島豊 ペタフロップスコンピューティング, 矢川元基 監修, 培風館, (2007), p183-202
- 3) 牧野圭一, 上島豊, 視覚とマンガ表現, 臨川書店, (2007),  
p1-5,221-229