

研究生活のためのインフォマティクス【第2回】

研究成果のトレーサビリティ—汝のディスクを闇で満たすなかれ

轟 眞市

物質・材料研究機構 光材料センター*

Shin-ichi TODOROKI

相次ぐ食の偽装・毒物混入問題を契機に、食の安全確保に関する意識は高まり、消費者が生産者や流通履歴を特定できる仕組みも次々と導入されている。我々研究者はこれを対岸の火事だとは言っていない。研究者の倫理を問われる事件が後を絶たない以上、いつ何時、火の粉が降り掛かってきても、やましいことは無いと胸を張れるだけの物証を日々築いていく必要がある。

あなたが出版した論文に使ったデータの生データを見せる、と言われた時に、直ちに示すことができるだろうか? 「ハードディスクのどこかにはあるはずなのに、見付からない」、では言い訳にしか聞こえない。本稿では筆者が講じている対処法を紹介する。特殊な方法に見えるかもしれないが、その裏に流れる哲学を感じていただけたらと思う。

データと成果を最短距離で結ぶ

実験を行ってデータを取る。データを寄せ集めてグラフや図を作る。使えるグラフや図を散りばめて論文を書く。これらの作業に用いられるソフトウェアは、十人十色、さまざまなものが挙げられるだろうが、その間をやりとりされるのはすべて電子的に保存された「ファイル」である(図1参照)。ここで、論文からデータに向かって逆に辿ることを考えると、多数のファイルを如何に辿りやすく運用するか、そのポリシーが大事であることがわかる。

辿りやすいとはどういう状態か? まず、生データから論文ファイルまでに介するファイルが少なければ少ない程良いところは明らかである。筆者は10年前

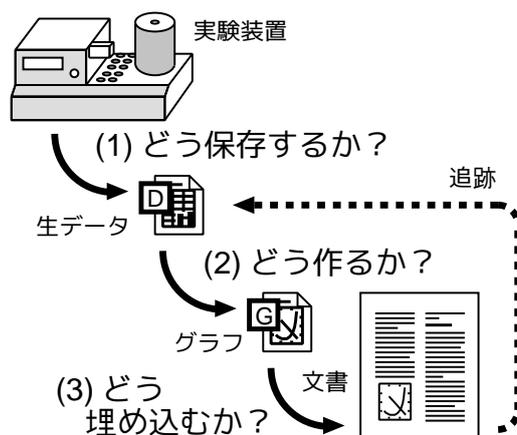


図1: 実験データの流れ。

の転職を機に、この「データと成果を最短距離で結ぶ」ことを優先してパソコンの執筆環境を整えてきた。その基本方針は、以下の3点である。(1) 生データは時間順に保存していく、(2) 生データを直接読み込ませてグラフを作る、(3) 文書に埋め込むグラフや図は同じ場所にコピーする。

破綻しない命名ルールで保存

保存するファイルの名前を、試料名と実験条件を書き連ねたものにするのは、一見便利な様だがいずれ破綻する。全ての実験条件を書き尽くすことはできないし、毎回長々と打ち込むのは億劫である。筆者は、測定日時を表す6桁の数字で保存し、実験条件との対応関係を電子化実験ノート [1] に記録している。保存先のフォルダ名は、測定月日を表す4桁の数字にしているので、ファイル名の重複はありえない(図2参照)。この様にして、全ての実験データをその内容に関わらず時間順に保存している。

*〒305-0044 茨城県つくば市並木 1-1
fax 029-854-9060
URL: http://www.geocities.jp/tokyo_1406/

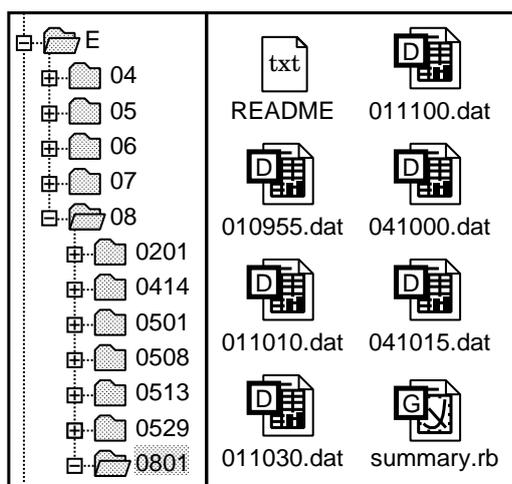


図 2: 実験データを保存するフォルダ構造の例。

データを保存する度に命名に悩む必要はないし、ファイル名を見れば測定日時が分かるので、そのデータに対応する実験ノートのパージをすぐ開くことができる[†]。測定が終了したら、ただちにメインで使っている PC にネットワーク経由でコピーする。連載第一回で述べた様に、自分の行動範囲にある PC は、全て同じファイルが見えるようにする仕組みを導入したので、将来事故でデータを紛失する心配は無い。

ファイルの数をむやみに増やさない

生データとグラフが常に一対一に対応するものばかりなら、後から追跡するのに苦労は無い。しかし例えば、生データがグラフ作成ソフトに直接読み込めない形式の場合、どうするか？読み込み可能な形式に変換した中間ファイルを作ってしまうと、保存するファイルの数が増え、後からの追跡が困難になる。中間ファイルの素性をきちんと記録しておかない限り、いずれ意味不明なファイルになってしまうからである。

そこで筆者は、どんな場合でも生データからグラフを描く、という原則を自らに課した。そのためのツールとして、Ruby/PGPLOT というグラフィックライブラリ (<http://pgplot.rubyforge.org/>) を使っている。つまり、プログラミング言語を介してグラフを描い

[†]この目的のための専用ユーティリティーを Ruby で自作した。文献 [1] の図 5 参照。

ているのである。最初のうちは時間を食うが、ノウハウが蓄積されていくにつれて、作業が短縮されていく。さらに、表現の自由度が高くなるので、シンプルでわかりやすい表現を追求できる。Ruby のオブジェクト指向プログラミングを利用できるので、多次元データを扱う場合でも見通し良く表現できる [2] し、動画作製に用いる多数の原画もこれで作成した (<http://www.youtube.com/tokyo1406> 参照)。

埋め込むなら同じ場所にコピーせよ

同じグラフを複数の場面で使うことはざらである。筆者は、文書に埋め込むべきグラフを描くプログラムを、その文書と同じフォルダにコピーすることになっている。そうする理由の第一は、グラフの表現がその文書に適したものになるように、プログラムを手直しするためである。

同じフォルダに保存されていれば、文書が指定されればその中に埋め込まれたグラフを描くプログラムもすぐに分かる。生データの場所を特定したい時は、プログラムの中身を読めば良い。既に述べた様に、ファイル名から生データを保存した日付が分かるので、対応する実験ノートの記述もただちに参照できる。

[参考文献]

- [1] 轟 眞市, 小西 智也, 井上 悟: “ブログを基にした実験ノート: 個人の研究活動を効率化する情報環境”, *Appl. Surface Sci.*, **252**, 7, pp. 2640-2645 (2006) の和訳。
- [2] S. Todoroki: “Object-oriented virtual sample library: a container of multi-dimensional data for acquisition, visualization and sharing”, *Meas. Sci. Technol.*, **16**, 1, pp. 285-291 (2005).

上記文献はセルフアーカイビング (自著ファイルをネットで公開すること) されています。電子ジャーナルの場合、著者版原稿ファイルの形で内容を読めます。タイトル等で検索を掛けてみてください。