

研究生活のためのインフォマティクス【第1回】

ポケットひとつの原則—ファイルは手ぶらで運ぶもの

轟 眞市

物質・材料研究機構 光材料センター*

Shin-ichi TODOROKI

USB メモリやノートパソコンを紛失してしまい、「記録されていた個人情報流出する恐れあり」と、責任者が謝罪する事件を目にするたび、あの仕組みを導入しておいて本当に良かった、と実感する。ネットから侵入してくるウイルス等の脅威もさることながら、情報を管理する側の人間に完璧を期待できない現実にも対処すべきである。

また、裁量労働制や在宅勤務の導入で、場所にとられない働き方が可能になってきた研究者は、どうすれば効率的なファイル管理ができるのであろうか。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。

わかっちゃいるけど、

なぜ、USB メモリやノートパソコンを持ち歩かなければならないのか？外回りでそれらが必要であるなら仕方がない。でももし、職場と自宅と同じファイルやソフトが必要、という理由だけなのであれば、他にもっとやり様がある。ネットワーク経由で運べば良いのだ。

例えば自分宛てのメールに必要なファイルを添付する。あるいは、オンラインストレージ[†]を利用する。でも、たくさんのファイルを参照する必要があるし、いちいち選んでいられない。急いでいる時はうっかり転送し忘れることもある。

そう、必要な作業をするのにいちいち手間がかか

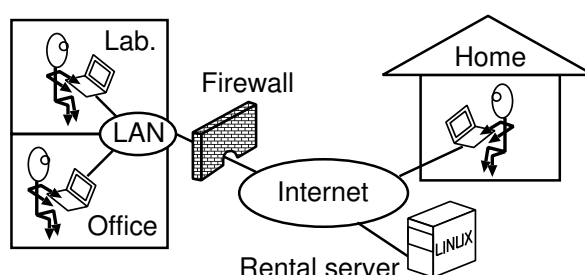


図1: ネットワーク構造

ると、それが億劫になって、結局望ましくない安易な方法(この場合はファイルやPCの持ち出し)に走ってしまうのが人間だ。でも、安易な選択をして自らに常に完璧な物品管理を課すよりも、パソコンに完璧な作業をやらせる方が遥かに気が楽である。手を抜かず常に努力を強いられる位なら、手を抜くための努力をする方が良い。

単純にすればコトは容易

解決策は単純なものが一番良い。職場や自宅のどのPCを立ち上げても、同じファイルが見えるようにすれば良い。席を離れるときにコマンドを打ち、別のパソコンを立ち上げたらまたコマンドを打つ。あとはパソコンが自動的に最新のファイルに上書きしてくれる。移動させるべきファイルを指定する必要はなく、勝手にパソコンが選んでくれる。その様なプログラムを自作した。

これを導入して以来、どこで仕事をしようとも、仮想的に同じパソコンを使っている状態となった。「情報を保存する場所は一ヶ所にまとめるべし」というポケットひとつの原則 [1] にもかなうやり方だ。

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

[†] サーバの空き領域を借り受け、ユーザがインターネット経由でそこにデータを保存できるようにするサービスのこと。

働のいい読者なら、「ちょっと待て」とお思いになるであろう。職場から退出する時にパソコンの電源は切るのだから、自宅のパソコンから職場のファイルを見ることはできない。そもそも職場にはファイアウォールが設置されているのだから、自宅からはアクセスできない。

その通り。そこで筆者は外部にサーバを借り受け、そこを介してファイル転送している(図1参照)。オンラインストレージと異なるのは、ファイル指定をする手間を省いている点だ。全てのOSはLinuxに統一しており[2]、転送にはrsyncというフリーソフトウェアを使っている。接続先のファイルと比較して、内容が異なるものだけを選んで転送してくれるので、仕事一日分のファイル転送量はたかが知れている。2001年2月にこの仕組みを導入して以来、ファイルの持ち運びにまつわる問題に悩まされることはほとんどなくなった。今まで7年半も手が抜けたのだから、その構築に努力した甲斐があったというものだ。

常時電源が入っているレンタルサーバーとファイルが同期されているということは、常にファイルがバックアップされていることを意味する。定期的にファイルを光ディスクにバックアップする必要がなくなった。また、今までに2度、ハードディスクが壊れて中身を失ったことがあったが、迅速にデータを復旧することができた。

ユビキタス My PC

自分の行動範囲にあるどのPCを使っても同じファイルが見えるということは、重さの無いファイルキャビネットを持ち歩いている様なものだ。今まで紙媒体で保存していたものをハードディスク上への保存に切替えることの利益は測り知れない。電子ファイルの形で収集した論文等はどこでも読み出せるし、実験ノートも電子化が可能である[3]。忘れて困ることは何でも入力しておく習慣がついた。

古来、良いアイデアが浮かぶ場所として三つの場所が挙げられている。馬上・枕上・廁上、すなわち移動中、就寝中、排泄中である。思い付いたアイデアを、忘れないうちに最寄りのPCに入力しておけば、あとでどこからでも検索することができる。自分専用のアイデアデータベースとなる。

整理するなら効率的に

昨今のハードディスクの容量は肥大化し、溢れる心配はほぼ無用になった。しかし、せっかく保存しておいても、必要な時に必要なものを見付けられなければ、ゴミを大事に抱えている様なものだ。後で見付け易くなるように整理することも大事だが、全てを整理しては、時間が幾らあっても足りないことは目に見えている。選択と集中が必要である。

自分の研究業績に関するファイルは後で参照することが多い。また、研究業績リストは研究者であれば常に更新しておくべきものである。ならば、これを電子化しておき、関連するファイルへのリンクを埋め込んでおけば、必要なファイルが見付け易くなる[4]。このリストに載らないファイルなら利用頻度も少ないから、別途全文検索ソフトウェアで探す方が効率的である。

机やキャビネットを使い易く整理することは一仕事だが、ハードディスクの整理と検索は、パソコンの力を借りることができる分、うまいやり方があるのである。

[参考文献]

- [1] 野口 悠紀雄：“「超」整理法”，中公新書 1159 (1993)
- [2] 轟 眞市：“技術者の工具箱 (2) Linux との巡り合わせは葉隠の如し”，マテリアルインテグレーション, **21**, 8, pp.65-66 (2008).
- [3] 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。
- [4] 轟 眞市：“研究業績リストの電子化 研究者のための執筆・発表支援システム”，セラミックス, **42**, 7, pp. 520-524 (2007).

筆者の文献はすべてインターネットからダウンロードできます。タイトルで検索を掛けてみてください。

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

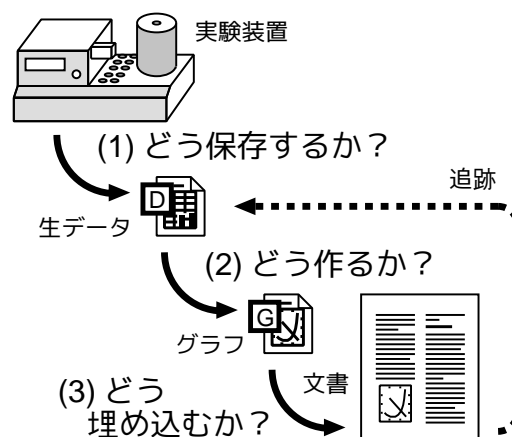


図1: 実験データの流れ。

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/

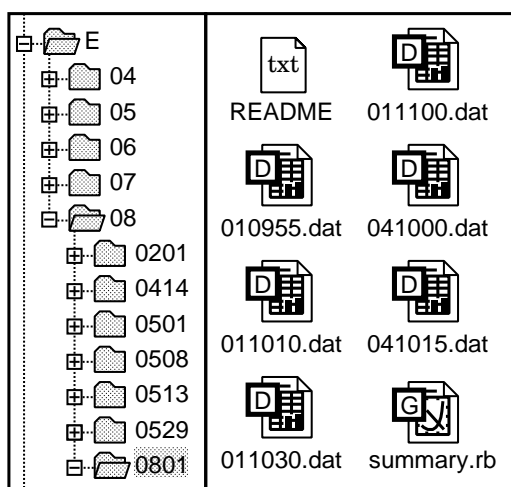


図 2: 実験データを保存するフォルダ構造の例。

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

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。

[†]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。

研究生活のためのインフォマティクス【第3回】

セルフアーカイビングのすすめ かわいい著作には旅をさせよ

轟 眞市

物質・材料研究機構 光材料センター*

Shin-ichi TODOROKI

調べ物するならまず検索、という世の中になって久しい。自分の良く知らないことを調べるなら、まずは日本語で検索するのが普通であろう。研究稼業の標準語は英語であるにしても、手っ取り早く知識を得るのに母国語の手軽さは捨てられない。

翻って考えるに、自分の研究成果を知ってもらうのにも言葉の壁は存在する。英語で原著論文を書き上げても、読んでくれるのは専門分野に近い研究者だけだ。専門家集団の枠を越えて自分の仕事を知ってもらうには、まず言語障壁の無い日本人をターゲットにするのが第一歩である。そのためには、和文雑誌等に記事を寄稿する機会を得たら、それを最大限活用すべきだ。出版社の権利を侵さない範囲で、原稿ファイルをインターネット上で公開すれば、読者は検索エンジンが連れてきてくれる。

このセルフアーカイビングを手助けするサービスが、最近手軽に利用できる様になった。大学などの研究機関に所属している人であれば、機関リポジトリを利用すれば良い[†]。そうでない人なら、YouTubeの文書版と形容される scribd.com が便利だ。本稿では、筆者が数年前から取り組んでいるセルフアーカイビングの効用を紹介する。

ホームページへのアクセスとの関係

筆者が自分のホームページを開設したのは1995年末のことだ。その後、2000年頃からセルフアーカイビングを始めた記録が残っている。図1に、ホームページへの累積アクセス数を、累積論文数(和英別)

とともにプロットしてみた。一見、相関するのは当り前の様に思えるが、細かに見てみると、興味深い特徴が浮かびあがってきた。

2007年7月、とある学会の会報に、研究発表に関するノウハウをまとめた記事[1]を発表した。発行と同時にホームページ上でも公開し、この記事に興味を持ちそうな人の集う掲示板に宣伝を載せてみた。その結果、その月のアクセス数は、普段に比べ約700も多くなったのである(図1の矢印(1)参照)。その雑誌の公称発行部数は6000部であるから、1割強の読者を追加で呼び寄せたことになる。彼らが紙媒体でその記事を目にすることはまず考えられないので、電子公開した甲斐があったのである。

セルフアーカイビングされた記事を見て、動いてくれた人も現れた。図1の2007年2月から始まる黒丸のプロットは連載記事[2]の発行を示している。この話を最初に持ち掛けてくれた編集者は、過去の筆者の記事(おそらく矢印(2)と点線で繋いだ2つの記事)を読んだことがきっかけ、と話してくれた。これが縁となって、連載物を執筆する機会が出版社を変えて現在まで続いている。各記事には過去の和文記事をURL付きで引用したので、自然とホームページへのアクセスが増えていった。

図1の 印を結ぶ線で示される、査読無し和文記事の累積数の増加傾向は、実線で示されるホームページへの累積アクセス数と、2006年後半以降において良い相関を示している。一方、査読付き英文原著論文(筆頭著者分)の累積数を示す 印の線との相関は見られない。原著論文を読んで引用文献(当然英語!)を引くのに、筆者のホームページを経由することを試みる人はほとんど居ないであろうから、当然のことである。

*〒305-0044 茨城県つくば市並木1-1

fax 029-854-9060

URL: http://www.geocities.jp/tokyo_1406/[†]我が研究機関においても今秋、独 Max-Planck Digital Library と共同で試験運用が始まっており、大いに期待している。

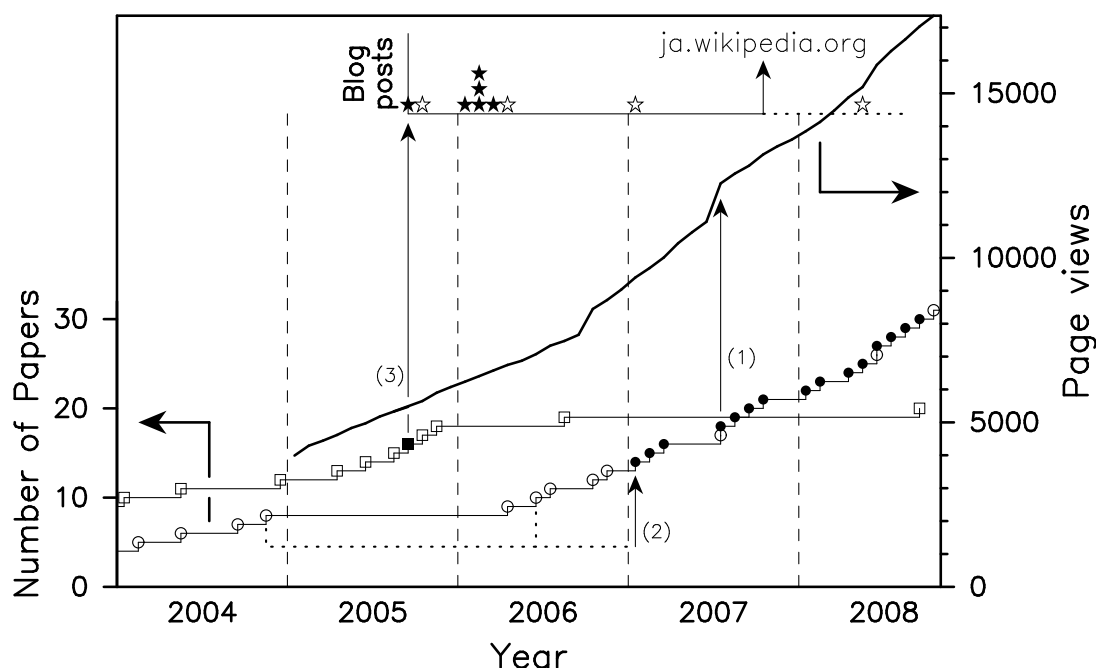


図 1: 筆者のホームページへの累積アクセス数 (太実線、右軸)、筆者の英文原著論文 (筆頭著者分、査読付き) の累積数 (印、左軸)、および筆者の和文記事 (査読無し) の累積数 (印、左軸; は連載記事)。上部の印は、文献 [3] (印) を紹介してくれたブログの記事の数を示す。(1)~(3)の縦矢印は、データ間の相関を示す (本文参照)。

翻訳の効用

2004年12月、材料インフォマティクスの国際ワークショップで実験ノートの電子化について発表し、そのプロシーディングス (査読済、英文) [3] が2005年9月にオンライン公開となった (図1の印)。日本人にも広く読んで欲しかったので、早速和訳して公開したところ、3人の日本人ブロガーが紹介してくれるに至った (図1の印)。これらのブログへのリンクを取りまとめて、ホームページに掲載しておいたのが功を奏したのか、2007年10月にはフリー百科事典『ウィキペディア』の「実験ノート」項目にも取り上げられることとなった。

さらに、これを見た言語学先攻のポスドクの方が、自らのブログで紹介してくれた (2008年5月)。論文のテーマが材料科学に閉じていない内容であったからこそ、専門分野の壁を越え易かったのだろうが、和訳を公表していなければ、文系の方々にも読まれていることを知る機会は無かったであろう。

セルフアーカイビングはまず母国語の文書から取り組んだ方が、反響を知り易いと思う。反響を知れば、励みになる。筆者はこれに味を占めて、和文記事

の英訳も公開してみた [4]。15ヶ月後ではあったが、スペイン人ブロガーの目に留まったのだから、セルフアーカイビングは止められない。

[参考文献]

- [1] 轟 眞市：“研究業績リストの電子化 研究者のための執筆・発表支援システム”，セラミックス, **42**, 7, pp. 520–524 (2007).
- [2] 轟 眞市：“セレンディピティの磨き方 ファイバヒューズ研究に至った縁と偶然 (連載全3回)”，工業材料, **55**, 2~4月号 (2007).
- [3] S. Todoroki, T. Konishi and S. Inoue: “Blog-based research notebook: personal informatics workbench for high-throughput experimentation”, *Appl. Surface Sci.*, **252**, 7, pp. 2640–2645 (2006).
- [4] S. Todoroki: “Two serendipitous episodes — How I embarked on fiber fuse research”, (2007).

上記文献は翻訳を含めすべてセルフアーカイビングされています。タイトルで検索してみてください。