

これまでの歴史と今後の展望 基盤整備（足回り）と学術側の資源（人）

林 祥介

神戸大学 理学研究科惑星学専攻／CPS(惑星科学研究センター)
日本気象学会学術委員会データ利用部会部会長
shosuke@gfd-dennou.org

提供

地球流体電脳倶楽部 <http://www.gfd-dennou.org/>

CPS惑星科学研究センター <http://www.cps-jp.org/>



2022年8月18日

地球デジタルツインの動向に関するワークショップ（オンライン）
第2回 日本気象学会 計算科学研究連絡会

項目

- 計算解析環境（ハードウェア）
- シミュレーション・解析可視化（ソフトウェア）
なんでもデータ（知見）
- データ
- 気象庁データと制度（ねじれたライセンス問題）
- 情報基盤
- まとめ？

計算解析環境 (ハードウェア)

- 80年代：大型計算機センター時代=成功した産学連携？
→ 計算環境の基盤的提供
 - 80年代のスパコン黄金時代にいたる汎用大型計算機の国産化と密接に結合
 - 60年代後半から拠点大学に共同利用の最高性能計算環境がほぼ自動的に展開
 - 世界を凌駕する潤沢な環境が、多くの利用者は自ら要求もしてないうちに導入
 - 格安(?)の使用料で学術利用
- 90年代以降陰り
 - ダウンサイジングの進行で分散、センターから人々が去っていった
 - HPC（高機能計算機）は汎用ではなくなり、科学技術計算に特化
開発経費が回収できない
 - ベクトル計算機の終焉
 - 全国展開は無理、フラグシップ機を一つだけ（ES, 京, 富岳）
 - 汎用計算機の生産のビジネスが終了→伝統的な国産計算機会社の撤退

計算解析環境 (ハードウェア)

- 今後のHPC環境の基盤的展開は？
 - 「地球シミュレータ」、「京」、「富岳」と来たが、**その先は？**
 - 大規模データの発生とネットワーク通信ネットワークによる分散からクラウド（再びセンター）化へ
 - 大規模データを扱う分野では、**データのアーカイブと（解析）計算環境の隣接融合が必須**
 - 当該分野のデータアーカイブと数値処理を可能とするセンター環境を分野ごとに用意する必要性が生じている
 - 従来の分野横断的共同利用計算機環境では役に立たない
 - JAMSTEC？理研？あるいは？
- 一方、手元計算機・データ（キャッシュ？）の必要は無くならない
 - 手軽な計算解析環境は紙と鉛筆の素直な延長
 - だけど、**オープンサイエンス・オープンデータ→検証可能な証拠を残すプロダクトや手順の長期アーカイブ**

シミュレーション・解析可視化 (ソフトウェア)

- 研究教育現場で、プログラムをライブラリとして整備し使いまわそう、という発想がこの界限でも広まったのは、やっぱり80年代
- それまではスクラッチアップするのが当たり前？
 - ゼロから始めるプログラミング生活は技術教育的でもある
- 80年代前半は
 - お絵描きソフトもなかった
 - 解析ツールもお手製
 - ネットもないからプログラムのやり取りも簡単ではない
(記憶媒体に格納して郵送)

シミュレーション・解析可視化 (ソフトウェア)

- 「地球流体電脳倶楽部」 1988
 - インターネット前夜、PC革命を受け、もうそういう時代じゃないだろう
 - **ライブラリ化して共有提供、使いまわす、**使われるソフトウェアがバグのないソフトウェア
 - **ことばとしてのソフトウェア (実験から理論へ)**
 - 読めるコード (維持継承と相互利用を意識)
 - 数式による理論の素直な延長
数式で行うのと同じ意識で計算コードをトレースし、改変する/できる
 - 検証できる計算=人の仕事がかめられる・認識の共有
 - DCL : まずはお絵描きソフトウェアの共有提供から
 - 塩谷雅人、酒井敏、林祥介+余田成男、山田道夫
 - 知見=様々な情報とその結合 (教科書、論文、資料、図、データ、…)
 - HTML前夜でもあった



塩谷雅人 (2022. 2. 9没)
2019@バンドン工科大学

シミュレーション・解析可視化 (ソフトウェア)

地球流体電脳倶楽部は
片手間仕事、「日曜大エプログラミング」
であった

- 片手間では済まなくなってきた
 - 高度化大規模化複雑化
 - 学生さんの「お遊び」の自然な延長上というには距離が遠くなった
- 暇もなくなってきた
 - 研究成果主義vs遊んでる余裕
 - 学生さんが「お遊び」してる場合でない
 - まして研究者が足回り作ってる場合でない

シミュレーション・解析可視化 (ソフトウェア)

- 冷戦終結・地球環境問題の発生とともに気候シミュレーションモデル開発にも（天気予報を担っていた気象庁以外にも）大きな投資
 - 気候気象シミュレーション（地球システムシミュレーション）ができることは内外に対する国際政治にかかわる戦略的に重要な課題に変わった
 - 研究開発人員の増加？
 - ソフトウェア開発人員は増加したのか？
- 正しくライブラリ化が進んで（米国から流入して？）
 - 自分で作らなくてもこれを利用して考察を進める道ができてきた
 - 特に描画ツール
 - 解析はどうか？
 - シミュレーションはどうか？

シミュレーション・解析可視化 (ソフトウェア)

- じゃあ、シミュレーションモデル開発はどこで？
 - 我々の理解の総体が（大規模）シミュレーションモデルである、
という認識
 - 私が属する我々がシミュレーションモデルを開発維持できることが学術の展
開に必須
 - 認識の多様性を担保するモデルの多様性の担保
共通ライブラリ？CBLEAM（富田・西澤）
- 解析可視化ツールの開発はどこで？
 - 米国製の変遷に追従すればよい？
 - ツールが思考の型を決めてしまう（言葉に等しい）という認識
 - ツールを生み出していく能力を持っている必要は無いかな？
- データ同化
 - データ解析とシミュレーションの融合…そういった展開を支えるソフトウェ
ア開発力を持つ必要は無いかな？

シミュレーション・解析可視化 (ソフトウェア)

- ソフトウェアの維持管理継承
 - 研究成果と直結する開発はまだしも、過去のソフトウェア資源の維持や書き換えなどは？
 - 米 (GFDLとか) では e-sience 時代にソフトウェアを現代化更新書き換えするのに人員予算が付いた
 - 再現できること
 - 数値計算やデータ解析の検証、追実験、追体験
 - ドキュメンテーション
 - パラメタの根拠、モデルの根拠、実装説明
 - ...
 - 動く状態を維持すること
 - 共同開発を進め、それでいて、多様性を維持するためにはツールボックス化、モジュール化、階層化などなど
 - 独立して動かせるプロセスモデル
 - 動作確認

データ・データ通信

- 1980年ごろ、**デジタル化**は進みつつあったが、まだ
 - 数字がならんでる紙束はあった
 - 天気図は印刷されて冊子になっていた
- ちなみに1958 無線FAXで配信されるようになった
- FGGE1978（全球観測）
 - 全球天気予報に向けた数値シミュレーションモデルの大規模化
 - 静止気象衛星を含む衛星観測網と**大量データの生成**
- AMEDAS運用開始1979
- インターネット
 - TISN1989
 - IJ1992

データ・データ通信

特に、莫大なデータをどう整理しどうアーカイブして利用可能にし維持継承するか、日本の気象学コミュニティ全体としては、考えることをさぼっていたように思う

- 気象庁に頼っている(業務人員がいっぱいいる)
- 個別的に必要な人・グループ・組織はそれぞれに努力
 - 「基盤的」になっているか？
 - 研究用はさておき教育用や第三者利用は？
- **DIAS**: 情報科学の大規模データ研究の延長上に乗っかってたりする
 - 全部DIASへ？
 - 永続性？
 - 教育用や第三者は？(一般利用可能なデータも多々ではある)

データ・データ通信

岸保勘三郎らがひまわり初号機撮像データの保存を実現しようと走っていたちょうどそのころ、

米国では、衛星データの研究教育への利用のために、同様の理由で、**UCAR/Unidata**が建議された（1983）

- 日本では、結局、結実せず、ひまわり初号機初期の元データは失われた
 - 大蔵省が、天気予報終了後のデータはゴミだ、ゴミを保存する予算はない、と査定したとのこと（記憶による岸保伝）
 - 文部省には、気象庁のデータを保管する予算はない…学術サイドに置いておくには桁違いのビッグデータ
- 東大生産研高木幹雄研究室→喜連川優研究室
 - 衛星通信実験（研究としてはたしかそういう類の？）でひまわりデータを受信しアーカイブ
 - その後、脈々とつデータ通信やデータアーカイブ研究の題材となって**DIAS**に至る

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

Unidata is a diverse community of education and research institutions with the common goal of sharing geoscience data and the tools to access and visualize that data. For more than 30 years, Unidata has been providing data, software tools, and support to enhance Earth System education and research. Funded primarily by the NSF, Unidata is one of the UCAR's Community Programs (UCP).

- データと解析可視化ツールの共有と活用を共通の目的とする教育・研究機関の分散連合で、教育研究のためにデータとソフトウェアを提供する

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

The Unidata program was first **conceived in 1983**, when the NSF called together representatives from some **80 U. S. university meteorology and atmospheric science departments** to conceptualize ways of **making “real-time” weather observations and other data available to the university community.**

- 1983年、大学での研究教育における（衛星観測時代における）気象観測データの即時的利用に向け、構想
- 1980年代の諸問題の素直な延長上に、気象データから地球科学データにターゲットが広がった（地球環境問題対応）

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

As a result of that meeting, **the Unidata program** was established to **facilitate the transmission of weather data, imagery, and forecasts to the university community**; to support universities' local interactive computing capabilities; and to improve data access capabilities for researchers and educators.

- 大学を支援し、大学に対する、気象データ、衛星画像、予報データの提供を促進すべく設立
- 州立大学等での学部の気象学教育（学部卒が各州の天気予報現場に展開する人材）の必要
- データ本体は、それぞれのデータを取得する諸機関に分散展開

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

Since its founding, Unidata has not only succeeded in its original mission to ensure access to geoscience data, but has built a vibrant community of researchers and educators who have communicated their evolving needs so that the program can be of even greater service.

- データへのアクセスを確保提供するのみならず、データに対する様々な必要を持つ研究者・教育者のコミュニティーを涵養し、Unidataの活動にフィードバックした

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

As a result of Unidata's robust community governance structure, our long-term efforts to develop and support scientific software for data access, manipulation, analysis, and visualization have enjoyed both support and guidance from those we serve.

- **それぞれの分野で事情が異なるので、それぞれの分野で対策を講じること、そのための技術と経験を持つための組織**
- コミュニティーが運営するUnidataが確立
利用者たるコミュニティの助言の下、
データを利用し、解析し、可視化するソフトウェアの開発と支援が今日に至るまで実現

UCAR/Unidata

<https://www.unidata.ucar.edu/about/tour/>

技術進歩の展望を提供、研究教育現場への新技術の導入を補佐・促進

- netCDF + the Climate and Forecast metadata convention + the Common Data Model
- tools
 - the Integrated Data Viewer (IDV)
 - the National Weather Service' s Advanced Weather Interactive Processing System (AWIPS) package
 - the MetPy library
- データの分散統合提供
 - The Local Data Manager and the Internet Data Distribution system
 - Unidata cloud computing and remote data access technologies

DIAS

地データ統合・解析システム

DIAS (Data Integration and Analysis System)

<https://dias.jp.net/>

- 情報科学系の活動として発足、IPCCとともに展開、統合研究プラットフォームとしての環境提供
 - 1980年代：東大生産研高木幹雄研究室（ひまわり衛星データ）
木連川優研を中心として、数多くのプロジェクトの支援
 - 2006-2010：第I期 国家基幹技術海洋地球観測探索システム「データ統合・解析システム」
 - 2011-2015：第II期 国家基幹技術海洋地球観測探索システム「データ統合・解析システム」
 - 2016-2020：第III期 「地球環境情報プラットフォーム構築推進プログラム」
 - 2021-：「地球環境データ統合・解析プラットフォーム事業」
本年度よりJAMSTEC が運用管理機関となる

DIAS

- 情報科学的研究プロダクトであった
 - 様々な分野のデータが多様な目的で統合的に活用できる、手法としての一般性を追求
 - 地球環境問題への対応として必要不可欠な方向ではある
- 今は「気候変動適応戦略イニシアティブ」
 - JAMSTECの運用
 - 広範な分野のアーカイブ、広範な分野からの利用
 - 利用アプリケーションの研究開発がメイン
 - そのような開発の場となる
- 各分野での、技術と経験をためる場が不足（DIASの範疇ではないかも）
 - 各分野から参集してもらう開発体制
 - 参集可能な各分野側の人員が足りないのでは？→それぞれの分野で確保しないとそれぞれの分野に展開し、それぞれの分野の個別事情に対応し、データ利用の布教に貢献する人員が足りない

DIAS

- 時限的運用が続いており永続性が保証されていない？
 - 永続性が担保されないのは学術データの「図書館」として不十分
 - 10年に伸びたけど
 - まあなくなることはないだろう
- すなおな延長上に世間のイメージする？
Digital Twin of the Earth?

気象庁データ

ずっと懸案

- 日本ではやはり気象庁発のデータが気象学の礎であり、したがって、研究教育の基盤に位置づけられているはずなのだが
 - これは気象庁の都合とは関係なく、学問の性格上自動的に
- **が、しかし、基盤たるにふさわしい自由・安価(?)な利用が実現できていない**
 - 灰色に実現されていて批判されている

気象庁データ入手問題

気象業務支援センター問題(2002.06 ~)

- 本来の趣旨は業務支援＝天気予報(解説)会社へのデータ提供
- 一方で情報公開法が制定され、官公庁作成のデータは自由に配布可能となった
- 研究・教育セクターもそこに放り込まれて、**データを買う(公開コストを負担する)**こととなった
 - インターネット開闢(TISN、89.1)以来10年ほど、気象庁との議論
 - GPV データ当初は民間事業者一口分を学術利用として購入、利用あるいは貢献する研究室約30?ほどで分割して負担。現在は、「学割」という位置づけ
- **研究教育**は「民間気象業務」か?で気象庁企画課と学会がやりあったが押し切られた
- 公開取得されたデータの**二次配布は自由**(情報公開法にのっとる公的データ)
- 「特殊」データ(民間業者の必要としないデータ)は気象庁との共同**研究**という形で(公開コストを気象庁が負担する＝職員が働くことで)入手できた
- 「特殊」データを純然たる**教育**目的あるいは試行目的で入手するパスがない

気象庁データ入手問題

気象研究コンソーシアム問題

- 気象庁との共同研究は、公開コストを気象庁が担う(気象庁職員が働く)
- 気象研究コンソーシアム
そこで**気象学会との包括的共同研究契約**という形を導入
 - 気象学会以外の研究者は、従来どおり、個別的に共同研究契約を結ぶ
 - 気象学会会員でも民間企業所属の研究者を代表とする研究には、包括的共同研究契約の趣旨に照らし、運営委員会はこれを適用しない
- **データの二次配布は、原則、禁止している**
 - 本来、公開コストが問題であっただけで、禁止する理由はないと思うが、現在の規約では禁止
 - 二次配布したければ、別途相談(あるいは情報公開請求)?
→第三者が教育や試行に使えない

気象庁データの今後？

そうこうしているうち、

交通政策審議会気象分科会, 2020,

気象業務における産学官連携の推進(提言).

https://www.jma.go.jp/jma/press/2012/23b/bunkakai_teig

結局、学術利用にも民間と同様の応分の負担が要求される？

- 気象庁には気象庁の都合(予算根拠)がある
 - その昔、ひまわりのデータを捨てることになったのと同様
- 当たり前であるが予算が潤沢であればおそらく問題ない
 - 今度こそ、気象庁データの無料公開が実現され、産官学の諸活動のための情報基盤として提供・確立されよう
 - Unidataと同様、解析可視化ツールや、さらには、データ同化の時代であるからモデルも含まれる日もこよう

気象庁データの今後？

- しかし、予算が潤沢であるはずはないので、例によって、公開コストに対する応分の負担を求められることになりかねない
- その際、またもや、学(研究・教育)が産と同じ扱いをされるとこれまで以上に困った事態に陥りうる
 - オープンサイエンス・オープンデータの時代であるから、学術論文に用いたデータは、検証可能性を担保するべく、参照可能にしておく必要があり、できれば参照可能なDOIが付され、公開され、自由にアクセスでき、論文知見が検証できるのが望ましい
 - 研究や教育が産官学でますますシームレスになっていくと想像されるなか、従来の、二次配布禁止・二次利用困難な共同研究枠組みの活用によるデータ利用は早晚行き詰るだろう
- 課金制度を導入すると、制度の設計に注意しないと、知見情報の流通を阻害する
 - 再配布は禁止しない、が必須

気象庁データの今後？

学術(研究・教育)領域としては、

現在の「気象研究コンソーシアム」を拡充・拡大し、気象庁データに対し、当該分野研究教育の情報基盤たるにふさわしい流通自由度が確保されることを望みたい

- もともと税金で取得し整形したデータであるのだからだれでも自由に無料で参照できる、である、たとえば
 - 気象学会会員であるという条件は外す
 - 二次配布の制限は外す
- 大学一般教育は無論、初等中等教育に至るまでの利用、あるいは、大学での企業連携研究、気象予報士個人の活動での活用などなど自由に

気象庁データの今後？

- 気象庁データは気象学の情報基盤を構成する大きな要素なのであるから、学術(研究・教育)的な意味での基盤化には**学術コミュニティとしても努力**することが必要だろう
 - 産官学といっても予算的には3つのセクターがあるわけではなくて、公的セクター(官学)と**企業セクター(産)の二つがあるだけ**で、国交省・気象庁と文科省とで費用をどう負担しても国民や産から見れば同じ税金
 - とりあえず文科省経由での予算を流す努力はしないと
- 出版社に対応する何かが必要だし、それを作り維持発展させるには当然**費用**がいる
 - 本や雑誌は買ってきたのだ(だんだん難しくなってきたけど)

日本学術会議への提案(榎本)

気象学会学術委員会データ利用部会

Unidata が1980年代に着手したこと(気象データのアーカイブと広く他分野を含む研究教育コミュニティへのシームレスな提供と必要なソフトウェア等の研究開発整備提供ならびに技術指導・人材育成などなど)を、遅ればせながら、しかし、今日的な技術環境と、日本の事情に合わせて着手する、その人的資源的予算要求ぐらいはしないと...

情報基盤

ハコモノ（普通に言う情報基盤のイメージ？）

- 紙と鉛筆
- スマホ？
- パソコン
- 計算機（HPCまで）
- ネットワーク
- データサーバ
- 付随するソフトウェア（OSやユーティリティー）

情報基盤

知見（情報の蓄積・理解の体系＝基盤的知見）
我々が気象のことを考えるときに利用する情報
理解＝言語化された知識、すなわち、「**言葉**」

- 理解の総体（我々が知っていることとその体系）
 - 辞書・教科書・論文・総説・ネット上に置かれた知識断片
- データ
 - 数値、画像、動画、WEBページに乗っかっている様々
- ソフトウェア
 - モデル、解析可視化ツール、データ同化ツール
- ことば
 - **母語（日本語）**（母語で考えられるって凄いことなのよ）

情報基盤

情報基盤というからには

- 思考を行うための道具あるいは部品なので
- **誰もが、自由に、無料で**(あるいは、学生でも困らないぐらいのできるだけ安価に)利用できるように維持管理提供されるべき
 - 道路とか電気水道ガスとか空気とか
 - 言葉とか数式とか

情報基盤

使い方: 箱ものの運用や知見の蓄積と利用の(第三者が使えるようにする)ためのノウハウ

- ノートをどう取るか(紙と鉛筆...他人も使えるように)
- 辞書をどう作るか
- 教科書(理解の体系)をどう表現するか
- 計算機やネットワークやデータサーバをどう構成し、構築し、維持管理運用するか
- ソフトウェアをどう構成し、構築し、改造し、...
- データやソフトウェアをどう整理して参照・継承可能にするか
- ...

情報基盤

しくみ・制度

- 出版社・本屋さん・図書館が情報基盤の要素であり、これを支えるノウハウを蓄積維持
 - 地図帳・図鑑からデジタルツインへ？
 - 図書館学
- 学会の大きな仕事の一つが出版(論文誌)
 - 知見の蓄積、理解の蓄積・改良
- データやソフトウェアはどう構築・整理・蓄積？
 - 情報科学
 - 我々のデータは巨大だ、ソフトウェアは複雑だ
 - ライセンス問題: 自由な閲覧と利用の担保
- ネット上の知見はどう継承？
 - ネット上の知見は消滅しやすい
 - 研究グループやプロジェクトのWEBページは揮発的

情報基盤（図書館と「データ」）

維持継承利用

- 紙の時代は、辞書や教科書や論文誌や総説や、...、といった実体に情報を圧縮、出版社という仕組みが確立し、図書館にこれを収めればよかった（という流れを数世紀かけて確立した）
- なんでもデータになっちゃった時代のやり方はまだ確立していない
 - それぞれの分野で事情が異なるので、それぞれの分野で対策を講じること、そのための技術と経験を持つことが必要
 - データ量が少くても、永続性は？WEBページなど？
 - まして気象学等データ量が莫大なのでどうやってもコストがかかる
- 「e-なんちゃら」と言われて久しいが（25年ぐらい前？）、残念ながら、我が国での予算は研究予算が主で支援経費はほとんどないので、結局、情報科学丸投げ
 - それぞれの分野でそれぞれの都合に応じて対策を講じる人員を配することが困難、研究の片手間でなんとかするしかなかった

情報なんちゃら(DX?)に対する我が国の問題点 「情報科学工学者・技術者魔法使い論」

- それぞれの分野で事情が異なるので、それぞれの分野で対策を講じること、そのための技術と経験を持つための場、を涵養するたえの資本投資が弱い
- 「情報科学」「情報関連企業」に頼りがち、投資対象も「情報科学」「情報関連企業」となりがちで自分の問題になってない
 - DIASが例であるが、現場のマインドも予算当局も企業も
 - 官公庁でも、自分たちのシステムを内製化する方向にはない
外注するのが普通、外注するための内部の人材育成すら考えてない
 - 民間活動も同じく、情報関連企業に丸投げの傾向
 - 研究現場でもこちらの方向に人員を割く余裕がない、別途ポストが生成できない、人を育てない
- いきなり統一環境(どの分野も使える)にジャンプしてもそこに繋がられない
 - J-STAGE とかなかなかいいんだけど
 - 情報科学や情報関連企業への投資となる
 - 各分野、現場、様々な産業固有の問題を理解した人材が育たない

まとめ（今後）

digital twin とかに際しての

- 当該分野のあらゆる知見はデータ（なんでもデータ）
 - 文字通りのデータ、様々なソフトウェア、ノウハウ、理解、…
 - 特に、当該分野が担うべき、
基礎的な知見は氏素性のしっかりした固い情報に磨く
 - それらを統合した「モデル」
- ネット時代の図書館
そのような知見を維持し再現し継承し参照利用できること
 - 動的データ：本と図書館のあらまほしき進化、当該分野でのあらまほしき形
 - 学術=教育・研究目的、当該分野の責務
→他分野・学生・一般ユーザへもシームレスに提供
できるだけ簡単なライセンスとか
- 当該分野において、それを可能とする計算環境、記憶環境、データ通信環境の確保
 - 自分たちで要求主張していかないと自動的に確保されないだろう
- そういったことを担っていく当該分野学術領域に人員確保と処遇確保
 - 今後へ向けての活動経費の確保