

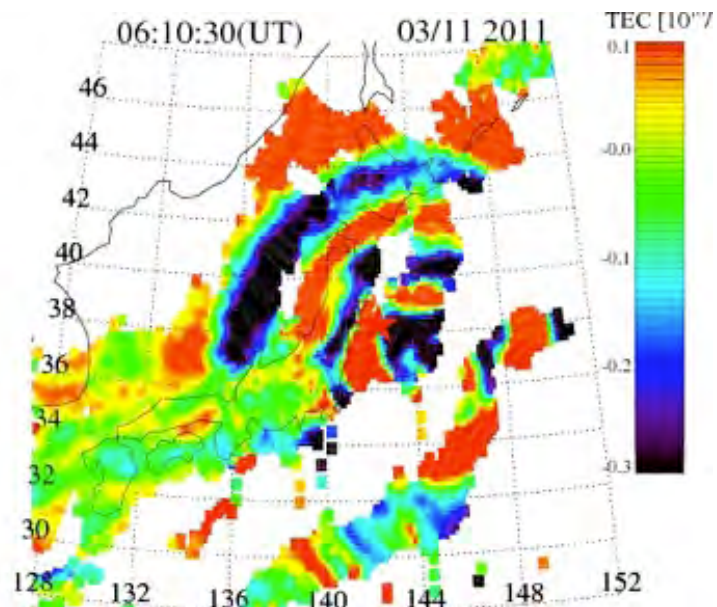
一般研究者の作成するデータ及び デジタル・リソースの恒久的オープン化

齊藤 昭則（京都大学理学研究科地球惑星科学専攻）

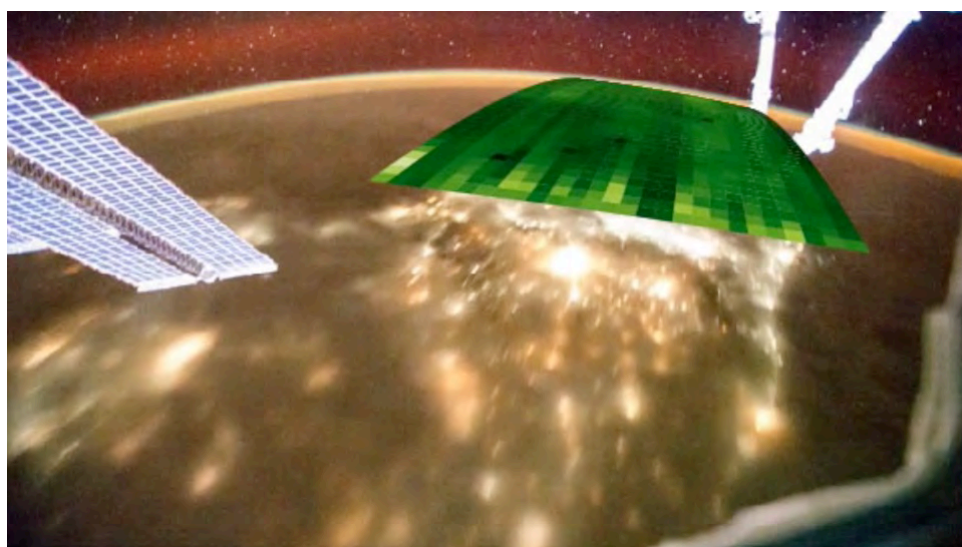
一般研究者（データベース・サービスを行う研究組織に属さない研究者）が作成するデータ（及びデジタル・リソース）を恒久的にオープン化出来る仕組みがあれば良い。

1. (私の)3つのデータベースとデータベースの4つのPhase
 - ・ GPS全電子数データベース
 - ・ ISS-IMAPデータベース
 - ・ ダジック・アース・データベース
2. データベースのPhase-up
3. 提案：共同利用機関によるデータベースPhase-up公募

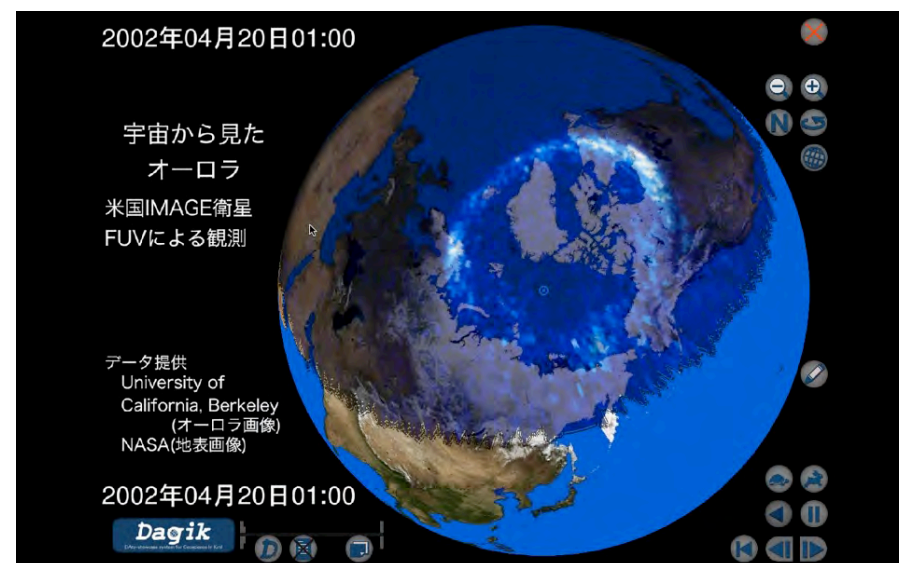
3つのデータベース



GPS全電子数 (2001-): Phase-D



ISS-IMAP (2012-): Phase-B



ダジック・アース (2009-): Phase-C

データベースの4つのPhase

一般研究者が作成するデータが、4つのデータベースのPhaseを経て発展していくことにより、有益なデータが保持されるようになるのではないか。

Pre Phase A : Source Data (検索性もない)

Phase A : Raw Database (検索性はある)

Phase B : Close Database (限定されたユーザーの利用)

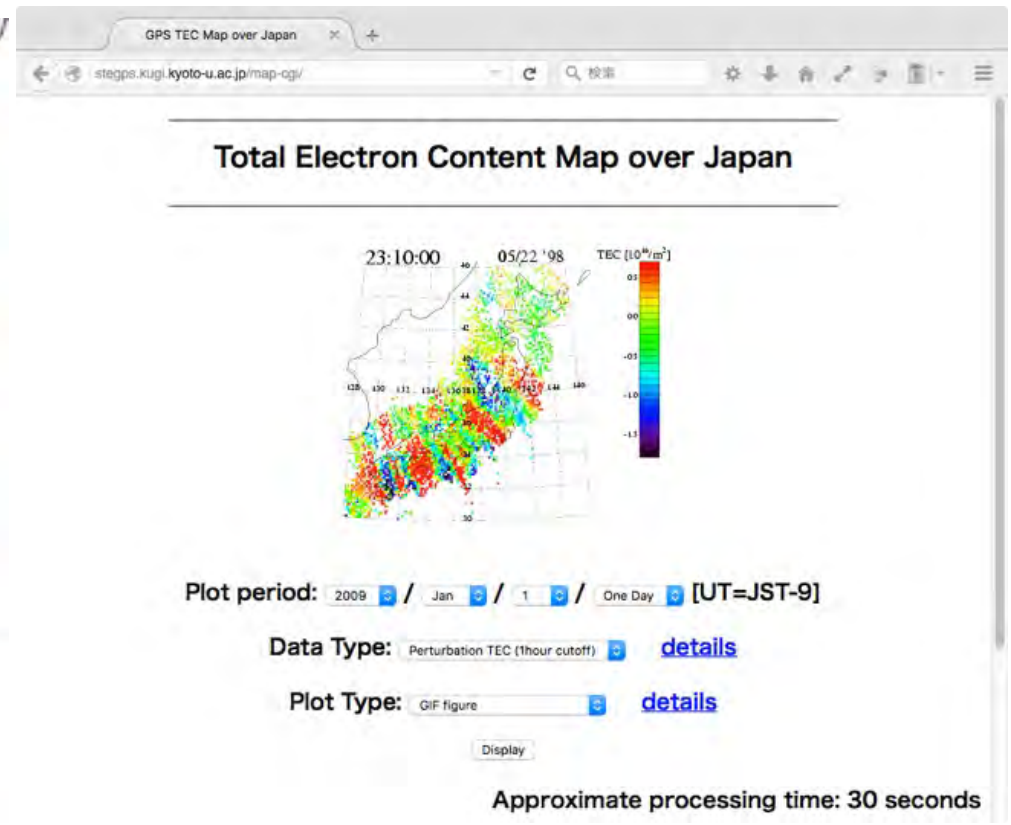
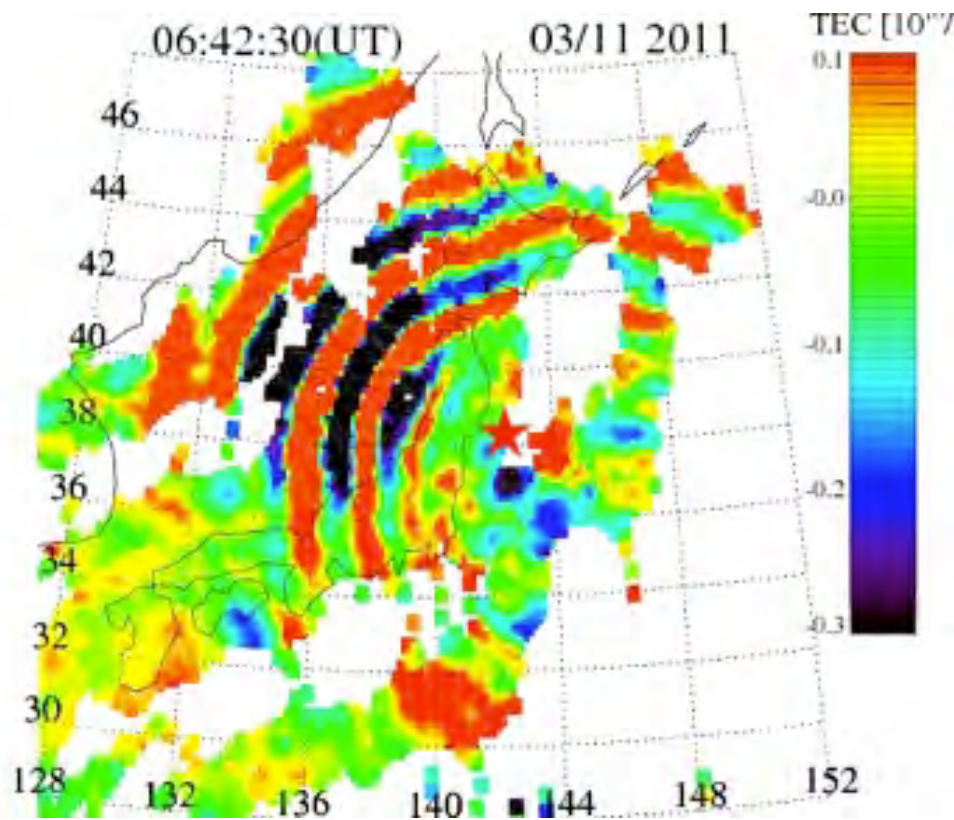
Phase C : Open Database (誰でも利用可能)

Phase D : Permanent Open Database (機関による恒常的サービス)

(Phase E : Open Open Database (誰でもデータ・サービスが可能))

GPS全電子数データベース: Phase-D

- ・ 国土地理院が運用するGPS受信機網GEONETのデータを用いた電離圏全電子数データ
- ・ 国土地理院は測位情報のみで電離圏情報の提供は行っていない
- ・ データは準リアルタイムに処理
- ・ 1997年から毎日4 G byte(非圧縮)のデータ量
- ・ 現在は7 T byteのサイズ。

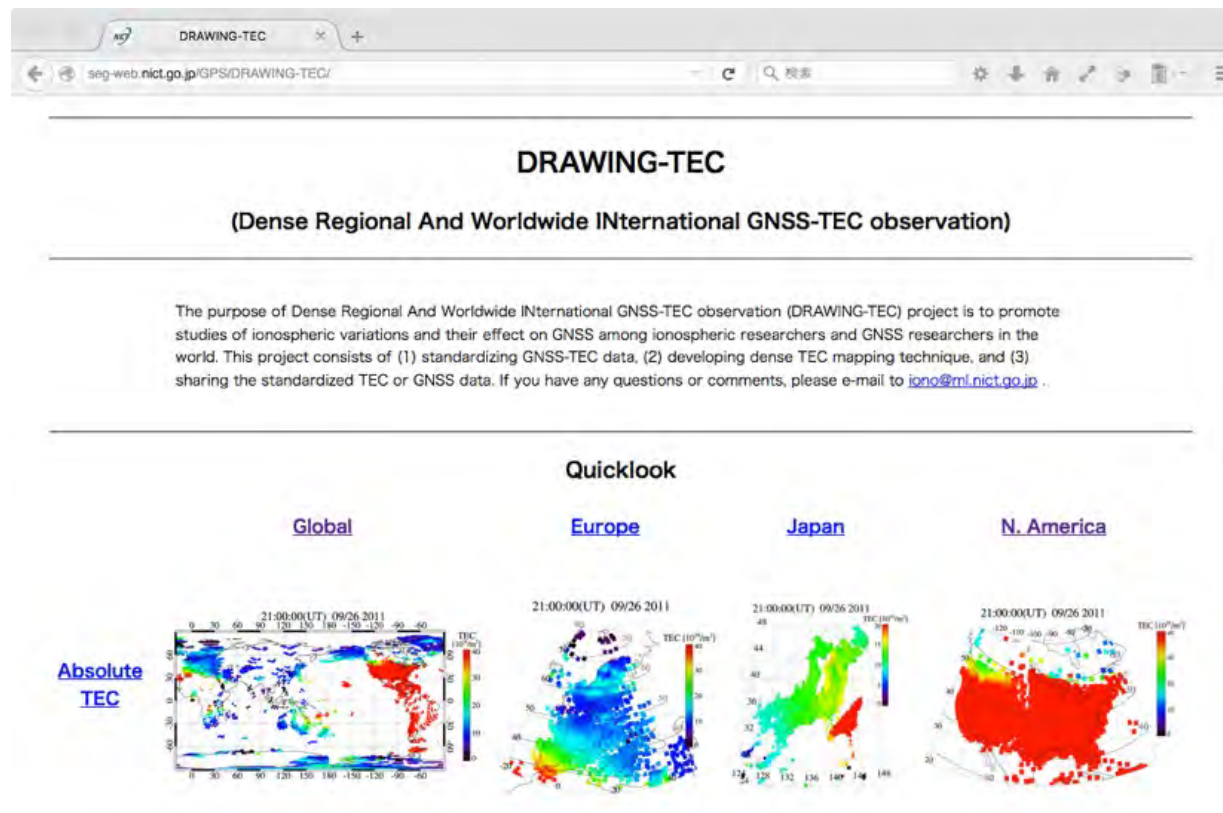


Phase-A : 1998-1999 : 全電子数アルゴリズムの開発

Phase-B : 2000-2001 : 1997年からのデータをデータベース化

Phase-C : 2001- : Webインターフェースをつけてデータ提供開始。科研費 研究成果公開促進費「データベース」(2002年から2009年まで)で維持。

Phase-D : 2010? : NICTが同じアルゴリズムとより広いデータを用いたデータベース・サービスを開始



<http://seg-web.nict.go.jp/GPS/DRAWING-TEC/>

GPS全電子数データベースから学んだこと (1/3)

なぜ始めたのか？

当時の計算機環境ではデータサイズが大きく、共同研究者が利用出来るデータベースを作り、共同で使うのが効率的だった。当時、国土地理院もNICTもGPS全電子数データベースを作ろうとはしていなかった。Phase-B の(close database) とPhase-C (open database) の間の技術的な手間は小さく思えたので、公開した。

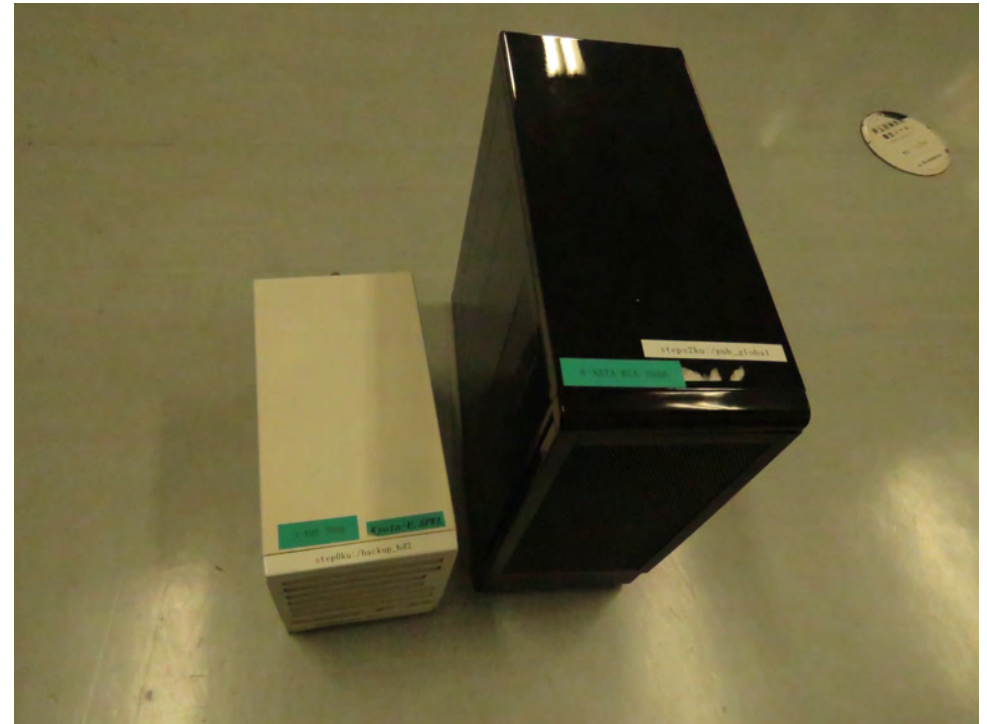
良かった点は？

- ・ 自分たちにとってもデータ利用は便利になった。
- ・ 幾つかの共同研究、幾つかの共著論文、GPS全電子数を使う研究者グループの構築。しかし、これらの幾つかはPhase-Bのデータベースでも得られたかもしれない。
- ・ いわゆる“Long-tail”なマイナーな使われ方がPhase-Cデータベースのメリットだが、概ねそういう使われ方は目につかない。
- ・ ハードディスクの容量の増加はデータの増加よりも速かった。

GPS全電子数データベースから学んだこと (2/3)

難しかった点は？

- ・ データのバックアップは必須。ハードウェアの故障は避けられない。ディスクだけでなくサーバー自身もバックアップを用意するべき。しかし、個人規模ではなかなかやりくりが難しい。
- ・ 予算は限られている。科研費「データベース」はハードウェアには使えない。そして、毎年申請書を出して10年間予算を得るのは簡単ではない。
- ・ ソフトの改良・更新、データエラーへの対応などデータベース維持のために日々の管理は必要。
- ・ 10年間で本人の科学的興味は変わっていく。他のプロジェクトとの関係でデータの優先度は変わっていく。



GPS全電子数データベースから学んだこと (3/3)

どうやったらデータベース維持の負担を減らせるか？

- ・ 「クラウド」システムはハードウェアの維持の労力を減らせる。
- ・ しかし、継続的な予算を取ることは一時的予算を得るよりも難しい。

結論

- ・ 有用なデータでも、一般研究者が10年以上、Phase-C データベース(Open database)を維持することは難しい。
- ・ NICTによるPhase-Dデータベース (Permanent open database)の開始に助けられた。
- ・ 有益なデータは大手機関が継続的にデータ提供をしてくれるとありがたい。