

11. データシステム部会報告

11-1. 背景

地球の自然系サブシステムと人間系サブシステムそれぞれの、および相互作用の複雑性やその変動の多様性を考慮すると、地球観測データは極めて多様である。また、衛星観測の活用機会が今後ますます増大し、数値モデルの高性能化に伴いその出力が急激に増加することを考慮すると、地球観測で取り扱うデータは必然的に大容量となると予想される。多様で多量のデータを統合化して、地球システムの理解を深め、変動の予測能力を高め、政策決定に資する品質管理された有用な情報へと変換し、その利用を効率的・効果的に促進することが急務である。

11-2. ニーズと現状

地上観測、高層観測、航空機・船舶による観測等の地上ベースの観測や衛星観測から得られるデータを利用して、地球システムの理解を深め、その変動に対する予測能力と対応能力を高めて社会経済的な利益につながる政策や意思決定を支援していくことが国際的に求められている。そのためには、膨大で多様な地球観測データを、収集・アーカイブし、データの品質を管理し、統合的に利用して、有用な科学的知見を抽出し、予測精度を高め、管理や政策決定に有用な情報へと翻訳し、その情報を国際的に共有できるシステムを構築しなければならない。しかし、現状では以下の問題点がある。

11-2-1. 観測データの公開性、流通性、統一的利用性が十分ではない

(1) 現業的な観測データについて、例えば気象や海洋等の分野では、観測の標準化、データ収集、品質管理、統一フォーマット化、アーカイブ、公開方法などが国際的枠組みの下に合意され、実行されているが、水資源に関する分野など特に各国間の利害に深く関連する分野ではデータの共有化（公開性・流通性）は遅れている。また、これまで国際的な枠組みの下で進められてきた高層観測等の直接観測が質・量ともに近年低下している国や地域もある。

(2) 研究的な観測データについては、先端的な観測機材の投入や一部地域、期間の強化観測によって多くの発見的成果を得ているものがあり、また生態調査や人間活動が環境に与える影響に関するデータ収集も進められているものの、計測方法やデータフォーマットが多様で、データの品質、加工度のレベル、精度にも格差があり、統合的利用のボトルネックになっている。また、有用なデータが得られても、その公開性、長期継続性については大きな問題がある。

(3) 現業的な観測データの場合はオンラインでのリアルタイム利用が可能な場合があるが、研究的な観

測データはオフラインの遅れモード利用がほとんどであり、予測や危機管理に適さない場合が多い。

11-2-2. 観測データの統合的利用が図られていない

(1) 地上観測は時間的連続性を有しているが点観測の場合が多く、また航空機、船舶は線的、衛星は面的な観測ではあるが、いずれもスナップショット的な観測の場合が多い。この時空間的特徴の違いが、観測データ相互の統合的利用を困難にしている。とりわけ衛星観測データの場合は、そのデータ量が膨大であり、時空間分解能や投影図法等が多様で、観測時の付随情報の活用が複雑であることが、他の観測データや数値モデル出力との統合的利用の阻害要因となっている。

(2) 一般に時間的、空間的に連続なデータを観測のみから得ることはできず、地球システムの変動性の理解や予測には不十分な場合があり、離散的データを統合するモデルが必要となる。

(3) 地球観測データ量は膨大であり、かつそれは急激に増加しており、2010年頃には数～数十ペタ（10の15乗）バイトに達するといわれている。コンピュータの計算機性能についても同程度の割合の向上が見込まれるものの、膨大なデータの取り扱いと解析のための人的資源の投入とソフトウェア開発の早急な対応が講じられないと、観測データの有効利用が図られないばかりか、データの収集、アーカイブすらできない可能性が懸念される。

11-2-3. 危機管理や資源管理等の政策や意思決定のための情報が共有されていない

(1) 危機管理や資源管理などの政策や意思決定において、地球観測データや数値モデル出力等の自然科学的な情報が活用されるニーズは非常に高いが、現在これら場面で有効に利用されていない。実際の危機管理や資源管理では、広域の現状の把握のみならず、地域的な特性に依存したローカルな将来動向に関する情報があわせて必要な場合が多く、実際の管理ニーズにあわせた有用な情報への翻訳が必要となる。現在この翻訳システムの開発が致命的に遅れている。

(2) 気象や海洋情報等、国際的な情報共有が進められている場合もあるが、一般にはこれらの情報を国際的に共有するための情報ネットワークとこれを運営する枠組みが未整備である。

11-3. 今後10年で取り組む方向性と目標

非均質情報源からの超大容量で多様な地球観測データから、有用な科学的知見を抽出し、危機管理や資源管理等の政策や意思決定のための情報を得て、これを国際的に共有するために、次の考え方が必要である。

11-3-1. 地球観測データの公開性，流通性，統一の利用性の向上を目指す

(1) 統合的利用のインセンティブを促すサービス

国際的合意の下に，データの収集ネットワークの構築，品質管理，データフォーマットの統一化を進める．しかし，その調整にいたずらに時間と労力を割くのではなく，まずはメタデータの共通化を目指し，その上でデータ提供側のインセンティブを促すようなサービス(例えば，ライフサイクル管理された類似するデータの参照支援，品質管理支援やデータフォーマット変換支援を行うシステム等)を提供することによって，データ提供側とユーザ側がともにデータの統合的利用のメリットを共体験することが肝要である．その上で，さらに理想的なデータ収集・品質管理・アーカイブに関する国際協力体制の構築へと進むことが望ましい．

(2) 段階ごとのデータポリシーに関する合意

データポリシーに関しては理想的な像を描きつつも，それぞれの地球観測分野の状況に合わせて，実行可能で詳しく記述した方策を合意し，国際的にデータ交換・統合利用の実績を共有することを積み上げながら，より理想的な段階へステップアップする戦略を取るべきである．

(3) 国際協力の推進

地球観測データは，複数の国，地域が協力して，全地球的でより総合的なシステムを階層的に構築していくことが望ましい．地球観測におけるわが国の観測，研究，調整能力や国際的なリーダーシップ，地域特性における強みに鑑みて，独自性を発揮しつつも，地球観測データの公開，流通，統一の利用の促進の国際的責任を担える分野，階層において，データセンターとしての機能を有する組織を設立し，国際的貢献を果たすべきである．

11-3-2. 現象の包括的理解を促進するシステムを開発する

地球システムは様々な事象が相互に関連しており，その包括的理解には個々のサブシステムを理解するだけでなく，全体システムを理解し，それに基づいて予測の高度化を進めなければならない．

例えば熱帯域の森林火災は地域の乾燥化が引き金であるが，乾燥化をもたらしているのはエルニーニョなどの地球振動現象であり，また森林火災によって放出されるエアロゾルは日傘効果によって日射量を減少させて人々の健康や農作物生産に影響を与え，また間接的な効果により雲量を増大させるが降水量は減少させるといった特異な現象を引き起こす．

これらの包括的な理解には，大気，陸域，海洋の物理系，生物系，化学系のプロセスと人間圏の応答に関する観測データ，社会経済データを統融合し，データ同化やデータマイニングなどの新たな手法の導入によって複雑系を解析する能力を球けることが

不可欠であり，得られる包括的理解によってより精度の高い予測が可能となる．

(1) 地球観測データ統合化システムの開発

観測者(データ提供者)自身が容易に品質管理できて，統一フォーマットに自動変換できる支援システムを開発して，そのサービスを広く提供し，多種多様な地上ベースの観測データのオープンなデータベース化を推進する．これにより，データ提供者およびデータ利用者のデータ統合利用を支援する．

衛星データに関しては，地球観測衛星委員会(CEOS)との協力の下に，データカタログ機能の高度化，ネットワークによるデータ転送，フォーマット変換，幾何補正，多種データの重ね合わせ，複合アルゴリズムの適用，データベース化などの高度なデータ処理解析技術(ソフトウェア)を開発する試みがある．こうした研究開発的な作業をひろげ，数値モデルの出力を地上ベースの観測データや衛星観測データによって検証するシステムを開発することにより，研究機関によるモデル改良やモデル予測精度の向上を目指す．また，これにより，地上ベースの観測データの品質が確認される上，(品質管理された)数値モデルの出力自体の流通促進を図ることが可能となる．

(2) データ同化システムの開発

地上ベースの観測データ，衛星観測データと，衛星観測原理である放射伝達モデル，地球システムを記述する数値モデルとを組み合わせることによって，観測データの空間的，時間的な代表性の限界を補うことのできるデータ同化手法を開発し，時空間的，物理的に整合性のある地球統合化データセットを作成するとともに，精度の良い予測や力学解析を行う上で重要な四次元再解析データセットを構築する．さらに生物・生態系プロセスや農業的な土地利用といった人間活動プロセスも取り込んだより統合的な同化手法へと展開を図る．

(3) データマイニング手法の開発

ペタ(10の15乗)バイト級におよぶ地球システム情報は従来の手法では解析しえない．そこでデータマイニングシステムを開発し，例えば研究者の様々な解析パターンをコンピュータで具現化する等，地球統合化データセットの高度な解析支援システムを開発する．

11-3-3. 社会経済利益に貢献できる情報を提供する

社会経済的な利益分野(災害，健康，エネルギー，気候，水資源，気象予測，農業，生態系，生物多様性など)の政策や意思決定において，地球観測データや数値モデル出力等の自然科学的な情報に対するニーズは非常に高いが，有効に利用されていないのが現状である．実際には，広域の現状の把握のみならず，地域的な特性に依存したローカルな将来動向

に関する情報との融合と、実際の管理ニーズにあわせた有用な情報への翻訳が必要となる。現在この翻訳システムの開発が致命的に遅れている。また得られる有用な情報を国際社会で広く共有することも必要となる。

(1) 自然科学的データと人文・社会的データの統合

地球環境問題に対して自然科学的視点と人文・社会科学的視点を統合してアプローチするためには、性質の異なるデータを重ね合わせることを可能とする基盤を構築する必要がある。そのためには、ベースとなる世界共通の基礎的地理情報を共有することがまず必要である。

我が国が中心となって国際協力の下に地球地図プロジェクトが推進されている。現在、18カ国の交通網、境界、水系、人口集中地区、標高、植生、土地利用、土地被覆の8項目のデジタル地理情報が公開されているが、この整備体制を強化し、データの信頼性や解像度の向上、収集項目などの一層の高度化を進める必要がある。

さらに、基礎的な地理情報を利用して多様な自然科学的データや人文・社会データを国際的に流通・共有化するためのデータ標準化やメカニズムの整備を進めるとともに、これらの活動を積極的に評価、支援する環境を整備する。

(2) 高度予測情報の提供

近年の数値計算技術の発達と計算機能力の増大により、詳細な過程を含む高分解能モデルや、変分法を用いた高度な同化手法などが開発されてきており、数値予報にも応用されつつある。また過去の地球システムの状態を再現しその変動トレンドをより正確に把握するための長期再解析も実施されるようになってきている。また、衛星観測手法、数値モデル、同化スキーム等を組み合わせ、地球規模から地域規模の変動を組みこんだローカルな現象の予測システムの開発も着手されている。

国内各機関との協力により、このような高度な数値予測モデルと地球観測データ統合化システムとを組み合わせ、高度な予測情報提供システムを開発し、政策決定機関による各地域における政策や意思決定を支援する。

(3) 情報の国際的な共有システムの構築

自然科学のみならず人文・社会データも含んだ情報の流通や多様な利用のためには、データの標準化などに加えて、各国、各地域の要求に応じた自然科学的、人文・社会科学的コンテンツやそれらのメタデータを、センサーネットなどのネットワーク技術によるローカルで、高密度、冗長・頑健なシステムによるデータ収集を実現し、ネットワークを通して情報の共有を可能とするデータセンターが必要である。また UNESCO や FAO など国際的な枠組みに提案し、その普及を図る。

11-3-4. 革新的情報科学技術の導入、開発に積極的に取り組む

地球環境科学と情報科学技術との相互協力によって、地球観測データ、社会経済データ、数値予測モデルの出力を統融合し、新たな情報価値を創出して国際的に共有する。

(1) 観測システム、計算機能力とバランスしたデータシステムの開発

これまでの情報科学技術分野の一つの流れは計算機能力（プロセッサ能力）の改善であり、現在のテラ（10の12乗）フロップスから GEOSS 時代にはペタフロップスの時代へ入ると考えられている。地球環境科学分野でもこれまで計算機能力の改善による数値予測モデルの性能改善のみに目を向けられてきたが、GEOSS 時代には衛星観測データならびに数値予測モデルの出力の量が飛躍的に増大し、現在のテラ（10の12乗）バイトからペタバイトあるいはエクサ（10の18乗）バイト時代を迎えようとしている。

したがってデータシステムにおける情報科学技術分野の革新的な貢献がない限り、包括的な観測データやモデル出力を有効に社会経済的な利益分野（災害、健康、エネルギー、気候、水資源、気象予測、農業、生態系、生物多様性など）に役立てることは不可能である。それどころか、現在ですら観測データおよび数値予測モデルの出力がこれら社会経済的利益分野に十分に利用されていないという事実は、観測システムや計算能力と比較してデータシステムの開発、整備がいかにも遅れていたかを示しているといっても過言ではない。

(2) データの政策的ライフサイクル管理システムの開発

データマイニングや情報融合などによって多量で多様なデータから新たな情報価値を創出するためには、超大規模データベースシステムや情報ライフサイクル管理などの情報科学技術における研究開発が不可欠で、これらの技術によって衛星観測データや数値予測モデルの出力といった超大量のデータの収集、編纂、検索機能を飛躍的に向上させる必要がある。

同時に地上観測や社会経済的データ等の不均質な情報源からの多様なデータを収集し、品質を管理するための情報科学技術の開発が必要である。また、特に研究的な観測データや社会経済的データについては、情報提供を待っているだけでなく、広く世界をモニターし必要な情報を探索する能力の開発も重要な要素である。

(3) データおよび情報を 100 年管理する情報科学技術の開発

地球環境科学と情報科学技術の相互協力を考えるときに重要視しなければならないのは、地球環境科

学では特に情報の蓄積（積分）が重要でそのスケールが 100 年に及ぶのに対して、情報科学技術では開発の速度（微分）を重要視してきたということである。

ハワイで 1950 年代に始まった二酸化炭素観測が今日の地球温暖化問題を認識させ、人類へ警鐘を鳴らし、国際的取り組みへの引き金となったことを考えれば、継続した観測とそのデータの蓄積が必要であることは明らかであろう。現在、気象の中期予測において世界でもっとも高精度を誇るのは欧州中期予報センターと評されているが、その秘訣はセンター創設以来過去 20 数年にわたって予測に用いたデータと予測結果をすべて収録、編纂して、適切に管理することにより、それらを随時最新の予測モデルに反映しているからといわれている。わが国で現在進めている気象の長期再解析も、この欧州中期予報センターの資産に負うところが極めて大きい。

地球環境科学におけるデータおよび情報の蓄積の重要性に比較して、これまで情報科学技術分野は 2000 年問題に代表されるように、長期的視野というよりは開発の速度（微分）を重要視してきた。地球観測のデータシステムの開発には、100 年のスケールを見据えた情報科学技術分野の革新的展開が望まれる。

11-4. 今後 10 年間程度の取り組みの重点事項

11-4-1. 非均質情報源からの超大容量で多様な地球観測データの統合的管理

- (1) 先進的ストレージ技術による超大規模データの政策的ライフサイクル管理
- (2) メタデータの標準化と利用環境の開発
- (3) データ提供側、特に人文社会科学分野のデータ公開、共有化を促すようなサービスの提供やサポート体制の整備
- (4) 実行可能で詳しく記述されたデータポリシーの策定

11-4-2. 超大容量で多様なデータ、モデルの出力の情報融合による価値創生

- (1) データ同化システムの開発
- (2) データマイニング手法の開発
- (3) 多様なデータの重層的可視化を可能とする先進的データベース技術の利用
- (4) データ辞書の整備・交換（オントロジーの整備・公開）

11-4-3. 社会的利益に資する情報の提供と国際的な共有

- (1) 地理情報、データ標準化による自然科学的データと人文社会的データの統合
- (2) 高度数値予測モデルと地球観測データ統合化システムとの組み合わせによる高度予測情報の提供
- (3) Web サービスによる情報の国際的な共有システム

ム構築

- (4) センサーネットなどによるローカルで、高密度、冗長・頑健なシステムによる地上データ収集と有用な情報への変換と情報の共有

11-4-4. 地球観測のデータセンター機能の提供

地球観測におけるわが国の観測、研究、調整能力や国際的なリーダーシップ、地域特性における強みに鑑みて、独自性を発揮しつつも、地球観測データの公開、流通、統一の利用の促進の国際的責任を担える分野、階層において、以下の二つのデータセンター機能を確立し、国際的な連携及び府省連携のもと、効果的、効率的に研究、開発、サービス提供を行う。

(1) 集中型データ統合センター

データ提供機関の独自性を確保しつつ、非均質情報源からの超大容量で多様な地球観測データの統合を効果的に行い、新たな情報価値の創生とその国際的共有を目的として、4-1~4-3 の成果を兼ね備えた集中的なデータ統合センターを設立し、データ GRID、WEB サービス等を介した国際的な機関協力を目指す。

(2) 分散型データ統合センター

非均質情報源からの多様なデータを、ユーザがネットワークを用いて手軽に、しかも統合的に利用できる分散型データ統合センターとしての機能を開発し、サービスを提供する。

11-5. 留意事項

11-5-1. 分野横断的協力が必要

多量で多様なデータを多機関から収集し、アーカイブするには、最新の情報科学技術分野におけるデータベースおよびネットワーク技術を導入し続けることが不可欠で、人文・社会系も含めた地球観測の各分野と IT 分野の有機的で建設的な研究・技術開発・定常運用体制を確立しなければならない。

11-5-2. 人的資源の重点的な投入が必要

地球観測分野における情報科学技術基盤の整備には、計算やアーカイブ資源の整備が不可欠ではあるが、データ収集と品質管理、データ統合化、データから地域支援情報への翻訳のための研究・技術開発・定常運用のために相応の人的資源が必要となることから、人的資源の重点的な投入が最も重要であることに留意しなければならない。

11-5-3. 現業機関と研究グループの連携が必要

データが現業機関と研究観測から提供され、またデータシステムが IT 分野の先端的な研究開発事項であるとともに現業的な長期運用が必要なことを考慮して、研究グループ、技術開発グループ、現業機関による連携体制を築くことが、持続的発展が可能な地球観測システムの基本的戦略である。

11-6. 課題分析表

(「重点化の必要性」の 印は重要度の高いもの)

ニーズ (重要度)	現状	ギャップ(問題・課題)	具体的な取り組み	重点化 の必要 性	重点化の視点(留意事項)
<p>非均質情報源からの超大容量で多様な地球観測データの統合的管理</p>	<p>気象、海洋の現業観測データはそれぞれ国際的枠組みの下でデータの共有化が進展 先端的な観測機材の投入や一部地域、期間の強化観測によって多くの発見的成果 基本的には各機関で独自のデータ管理が行われ、管理体系が不明瞭な場合が多い 生態調査や人間活動が環境に与える影響に関するデータ収集が進展 現業的な観測データの場合はオフラインでのリアルタイム利用が可能な場合がある</p>	<p>現業データの共有化において分野による格差が大 国や地域により現業観測データの観測量、質が低下 基本的には各機関で独自のデータ管理が行われているが、統合管理の体系は不在な状況であるので、統合的利用が進まない 研究観測データは計測方法・フォーマットの多様性、品質格差、長期継続性が、統合的利用のボトルネック 研究観測データの散逸の危険が高い 研究観測データは遅れモード/オフラインデータのため緊急時の利用が不十分 生態学、地理学、人文社会科学分野にシフトするにつれデータ流通が滞り、分野横断的・体系的取り組みの障壁となっている</p>	<p>個々の機関における独自の専門的サービスの継続と、超大容量で多様な地球観測データの統合アーカイブによる新たな価値創生を実現するセンター機能の確立 先進的ストレージ技術によるボリブームの超大規模データのライフサイクル管理 データの標準化と利用環境の開発 データ提供側、特に人文社会科学分野のデータ公開、共有化を促すようなサービスの提供やプラットフォーム体制の整備 実行可能で詳しく記述されたデータポリシーの策定</p>		<p>国際的合意の下に、データの収集ネットワークの構築、品質管理、データフォーマットの統一化 全地球的でより総合的なシステムを国際協力で段階的に構築 データ提供側とユーザー側がともにデータ統合化のメリットを共体験 国際的な実績を共有することにより、より理想的な段階へステップアップ わが国の国際的リーダーシップ、地域特性における強み、独自性を重視 分野による取得データに関する権利意識や価値判断など「文化」の差を考慮した柔軟な取り組みを進める</p>
<p>超大容量で多様な地球観測データの情報融合による価値創生</p>	<p>地上観測からは、時間的連続性はあるが点観測のデータが、一方、航空機・船舶・衛星からは、線的・面的ではあるがスナップショット的な観測データが取得される 地球観測データ量は膨大で、急激に増加(2010年頃には数~数十ペタ(10の15乗)バイト)</p>	<p>機動的な情報融合に耐えるシステムの不在 時間的、空間的特徴の違いにより、観測データの統合的利用が困難 観測のみから時間的、空間的に連続な情報を得ることは不可 衛星観測データは、データ量膨大、時空間分解能・投影図法等が多様、観測時の付随情報が複雑という特徴により統合的利用が困難 生態学、地理学、人文社会科学分野にシフトするにつれ、不定形・記述的なデータが増加</p>	<p>超大容量で多様な地球観測データの情報融合による価値創生を実現するセンター機能の確立 データ同化システムの開発 データマインガ手法の開発 多様なデータの重層的可視化を可能とする先進的データアクセス技術の利用 データ辞書の整備・交換(マトリックスの整備・公開)</p>		<p>データ提供側とユーザー側がともにデータ統合化のメリットを共体験 国際協力による相互利用性(interoperability)の向上 観測、データシステム、モデル、社会的利益分野ユーザーとの相互協力体制の確立 わが国の国際的リーダーシップ、地域特性における強み、独自性を重視</p>

<p>社会的利益に資する情報の提供と国際的な共有</p>	<p>危機管理や資源管理などの政策や意思決定の場での自然科学的な情報のニーズは高い</p> <p>危機管理や資源管理では、地域的な特性に依存したローカルな将来動向に関する情報があわせて必要</p> <p>地域における適応戦略や対策を企画・評価するにあたり必要な人文社会的なデータもきわめて不足</p> <p>数値計算技術の発達と計算機能力の増大により、高分解能力モデルや高度な同化手法などの開発、長期再解析に着手</p>	<p>実際の管理ニーズにあわせた有用な情報への翻訳システムが未整備</p> <p>国際的な情報共有のためのネットワークとその運営体制が未整備</p> <p>情報を国際的に流通・共有化するためのデータ標準化やメタシステムが未整備</p> <p>自然科学的、人文社会科学的コネクツや、それらのメタデータの収集・構築が未整備</p> <p>人文社会学的なデータは国などによる統計データを除き、これまでほとんど流通していない</p>	<p>地理情報、データ標準化による自然科学的データと人文社会的データの統合</p> <p>高度数値予測モデルと地球観測データ統合システムとの組み合わせによる高度予測情報の提供</p> <p>Webサービスによる情報の国際的な共有システムの構築</p> <p>センサーネットなどによるローカルで、高密度、冗長・頑健なシステムによる地上データ収集と有用な情報への変換と情報の共有</p>	<p>地理情報、データ標準化による自然科学的データと人文社会的データの統合</p> <p>高度数値予測モデルと地球観測データ統合システムとの組み合わせによる高度予測情報の提供</p> <p>Webサービスによる情報の国際的な共有システムの構築</p>
------------------------------	--	--	---	---