

プロジェクト名:

分野横断型融合研究のための情報空間・情報基盤の構築

プロジェクトディレクター： 東倉 洋一

1．研究目標

本プロジェクトは、生命システム、地球システムを具体的な対象として、大規模で複雑な巨大システムの解明に向けて、「生命」、「地球」、「統計数理」、「情報」の4分野横断型の融合研究を加速的に推進するための情報空間・情報基盤を確立することを目標とする。

2．研究概要

従来、統計数理や情報と他研究分野との融合は、統計数理や情報におけるさまざまな理論、手法、技術などを、他研究分野に取り入れることによって、研究の推進や新しい展開を意図したものが一般的であった。しかし、統計数理や情報を研究の手段として利活用する融合研究では、必ずしも十分な分野融合が期待できず、その成果も限られた範囲に留まることが少なくない。本プロジェクトでは、異分野を十分に融合させ、融合研究の深化と効果的な推進を加速させることを重視した新しい形の融合研究を追究する。具体的には、融合研究によって、従来の手法では見ることができなかった生命や地球の新しい姿を明らかにするための情報空間・情報基盤の確立、および、統計数理や情報の理論、手法、技術における新しい展開の2つの視点の両立を図ることを重視する。

研究姿勢として、「現場を知り」「現場に学び」「現場を分かり」「現場を創る」ことをコンセプトとして研究を進める。「現場」とは、生命・地球システムに代表される巨大システムである。この巨大システムへの挑戦には、システムを構成する大規模で異分野・異種の情報が複雑に絡み合った情報空間の構造を解明し、ここに含まれる情報を有効かつ効率よく活用するための手法が必要となる。本プロジェクトでは、このような大規模・異種情報に対して、「連想」と「リンケージ」という切り口で取り組む。また、生命・地球システムの研究では、生命情報や地球情報に関する多様で大規模なデータが必要であり、量と質に関して十分に要求条件を満たす統合データベースや統合的情報基盤、および、これらの活用手法の構築を目指す。さらに、「生命」、「地球」、「統計数理」、「情報」の4分野にわたる横断型の共同研究や研究協同を進めるためのコラボレーションやコミュニティ形成を促進する情報共有基盤が不可欠であり、これをバーチャルラボとして構築する。

すなわち、生命や地球の研究に新たな発見をもたらすとともに、統計数理や情報の研究における理論、手法、技術の壁や限界を突破すること、さらには未踏研究領域の開拓、未来価値の創成が、本プロジェクトで推進する融合研究の究極的な狙いとなる。

上記のプロジェクトを推進するため、具体的には、サブテーマ1：大規模・異種情報の収集・解析・結合・分類の手法および知識基盤の構築、サブテーマ2：地球・生命などの巨大システム解明のための統合的情報基盤の形成と活用手法の確立、サブテーマ3：コラボレーションとコミュ

ニティ形成のための情報共有基盤とバーチャルラボの構築の3つのサブプロジェクトを実施する。

3. 年次計画

テーマ	16年度	17年度	18年度	19年度	20年度	21年度
	予備研究	プロジェクト初年度		中間評価		
サブテーマ1	←				→	→
サブテーマ2	←				→	→
サブテーマ3	←				→	→

上記には、3つのサブテーマである 大規模・異種情報の収集・解析・結合・分類の手法および知識基盤の構築、地球・生命などの巨大システム解明のための統合的情報基盤の形成と活用手法の確立、コラボレーションとコミュニティ形成のための情報共有基盤とバーチャルラボの構築に関する年次計画を示す。

平成16年度（予備研究）

サブテーマ1に関しては、異種情報の結合・分類手法の基礎となる連想検索の研究開発を進め、文書情報の類似性を高速に計算する汎用連想計算エンジンGETA (Generic Engine for Transposable Association)の開発と、そのオープンソース形態で配布を行ってきた。また、GETAを利用した図書情報検索サービスWebcat Plusによる、900万冊の本の連想検索の提供をはじめとした多様な情報サービスへの適用によって、要素技術としての充実を図ってきた。

サブテーマ2に関しては、ライフサイエンスメタデータベースシステム構築の研究を推進するための枠組みとなるバイオポータルの研究開発を国立情報学研究所と国立遺伝学研究所が共同で推進し、これを公開した。最新の研究成果を「科学者がわかりやすい日本語で語る」ことを基本として、専門用語の変換翻訳、用語辞書、解説などの整備を行い、研究成果の背景にある生命の原理に触れるための道具として提供した。

サブテーマ3に関しては、情報共有基盤の基礎となるグループウェア「NetCommons」を構築し、全国約70団体で運用実験を行ってきた。また、大学院教育を主とする高等教育向けe-LearningシステムWebELS (Web-Based Learning System)の出発点となるWebLSの研究開発を行い、試用を行ってきた。

平成17年度（プロジェクト開始）

プロジェクト開始にあたって、プロジェクトの目標と方向性の明確化、サブテーマにおける研究項目と研究計画の具体化を行い、3つのサブテーマ毎の要素技術に関する研究開発を、サブテーマ間の連携と生命・地球システム解明を目指した情報基盤構築という目標を共有して推進する。プロジェクト運営において、特に、機動性と柔軟性を重視した予算配分・経費投資および人材雇

用・補強を重視する。

サブテーマ1では、分野を横断して存在する異種データの情報内容の類似性に基づく連想的な結合・統合と自動分類などの手法の検討とシステム実装、および、実証実験に用いる各種情報を収集・整備する。これらの情報を使って、異種情報源の関連づけとその提示が可能な利用環境のプロトタイプの開発を行う。また、リンケージ情報を機械的かつ大規模に収集するための機械学習・データマイニング等の要素技術の研究開発を行う。

サブテーマ2においては、ライフ系統合データベースの核となるメタデータベースシステムを構築のための基本設計、および、必要なハードウェアやソフトウェアの整備を行い、国内データベース充実に力点を置いたシステム構築を開始する。また、国立極地研究所が収集してきた極地生物資料のうち、植物関係の資料を対象とし、従来から付記されている情報に新しい情報を加えた統合データベース構築のための実験、データベースの基本設計、および、必要な設備等を整えらるとともに、貴重資料に対して適切な手法の開発を開始する。さらに、多種多様な地球環境データの統合について検討し、統合に関する方向性と基本的な考え方を明確にするとともに、いくつかの統合形態について、Web上のデータを使った実験的な検討を行う。

サブテーマ3においては、バーチャルラボ構築支援システム NetCommons を開発・公開するとともに、広報のためのユーザカンファレンスを開催し、システムの多言語化への取り組みとオンラインマニュアル整備を行う。また、国立遺伝学研究所の生命情報・DDBJ 研究センターをはじめとする複数機関の情報共有基盤として導入し、情報共有の促進と試用を図る。さらに、WebELS に関して、コンテンツアップロード用インターフェースの開発、コンテンツへの音声・ビデオ付加機能や多地点オンライン・ミーティング機能の付加など、システムの機能充実に図る。

平成18年度

サブテーマ1においては、連想的に結合された異種データの有用性を示すため、環境問題を例題に取り、多様な情報をテーマ別に自動分類する方式について研究する。また、研究者間のコミュニケーションがますます困難になりつつある生命科学分野で、困難さの原因が固有名称の多用や遺伝子の機能構造に関する自然言語表現にあることに注目して、それらを自動的にアイコン化して分野間のギャップを埋めるジーンアイコン(遺伝子象形文字)プロジェクトを推進する。さらに、科学研究費補助金データベースとCJP および Thomson Scientific 社の SCI (国際的な論文誌) の2つの引用索引データベースとの間のリンケージ法の検討に重点を置いて研究を進める。数十人(～数百人)程度の規模の研究者をサンプルとして研究者や所属機関履歴に識別子を付与したオーソリティデータを作成し、リンケージ情報抽出技術とリンケージ情報分析技術の融合を試みる。

サブテーマ2では、利用者主導型メタデータ DB システムの構築と運用に向けて、データ収集を精力的に進めるとともに、レコードの整理を進め、データベースの所在情報に加えて、機能情報とその評価とを連結させる。このため、さまざまな分野の研究者で構成する共同研究者グループを中核とした評価体制を確立する。また、極限環境生物統合データベース構築のために必要な

基礎データの一層の充実を図る。

さらに、地球情報に関しては、ウェブサービスを用いて国立情報学研究所と国立極地研究所のシステムを接続するとともに、「鉛直地球ポータル」および「南極 GIS ポータルサイト」を一般向けに利用可能な形で公開する。また、気象シミュレータの稼動による気象シミュレーションを行い、「鉛直地球ポータル」における蓄積・処理データの定常的な生成を行う。

サブテーマ 3 では、NetCommons の国立極地研究所への一斉導入を行い、運用上の問題点・改善点を探り、特に、各研究室の情報公開用 Web 基盤として、また、研究室内の情報共有基盤としての活用を図る。教育分野への展開においては、特に、セキュリティ面・運用面について検証を行うとともに、利用者アンケートを実施する。また、アジアの各種言語化への取り組みの一環として、タイ語化を行う。

WebELS については、総合研究大学院大学の「テイラーメイド e-Learning(略称)」に協力し、生命科学研究科、複合科学研究科、高エネルギー加速器科学研究科等での実利用を開始し、評価・改良を行う。主として、システムの信頼性向上、セキュリティ対策等とともに、グローバル展開のための多言語対応インタフェースを開発する。

平成 19 年度 (中間評価)

サブテーマ 1 では、環境関連情報と生命科学分野の研究情報を例にとり、異種情報源から情報内容の類似性に基づいて関連情報を収集し、それらを概観しやすい形で提示する情報システム、および、ジーンアイコンを活用して、文献の深い理解に必要となる遺伝子等の情報について、最新の関連データの内容を略図表示するシステムを試作する。また、研究者のオーソリティデータの規模や分野を拡大し、分野やセクター間での連携・融合の分析を進めるとともに、対象を他の情報源に拡大し、リンケージの新たな可能性の検討を行う。

サブテーマ 2 は、生命と地球に関する統合的データベース構築を目指しているため、生命分野や地球分野の進展を注視し、その最先端の状況を反映しながら研究開発の方向性を決めていくことが重要となる。さらに、両分野における国策的な統合データセンター設立の動きにも関連する可能性があるため、中間時点までの進捗状況と成果および周囲状況を検討し、必要な組織の再編と計画の調整を行い、これらの検討結果に基づいて、最終目的実現のための研究を推進する。

サブテーマ 3 では、複数の NetCommons 間で認証を行い、コンテンツの移転・コピー・エイリアスの作成などを契約に基づいて安全かつ簡便に行うモデルを構築・実装するとともに、NetCommons を個人のバーチャルデスクトップ・ファイルサーバとして利用するための研究開発を行い、これを NetCommons ver2.0beta として提供する。

また、総合研究大学院大学の「テイラーメイド e-Learning」に協力し、実利用の促進、および分散統合システム化と高信頼システム化を研究開発するとともに、グローバルスタンダード化を目指すために、国内・国際展開を図り、共同研究開発体制の構築を開始する。

平成20年度

中間評価に基づいて再調整、再組織した研究計画と研究体制によって、後期プロジェクトを開始する。特に、実現性と有用性および学術性などの観点によって、プロジェクト終了時の具体的達成目標を明確にして研究を推進する。各サブテーマにおいて確立、構築した要素技術やプロトタイプシステムに関して、実用環境を前提とした実証実験と評価を行い、最終年度に向けた問題点の把握とその解決策を検討する。

大規模で異種の情報空間の結合、分類、分析研究に関しては、要素技術や新手法で構成するプロトタイプシステム分野融合研究への具体的な適用を図る。生命と地球のデータベースに関しては、ライフサイエンスメタデータベースシステムおよび「鉛直地球ポータル」を中心とした地球情報統合データベースシステムとして、そのプロトタイプを公開し、試用に供する。また、コラボレーションやコミュニティ形成のための情報共有基盤に関しては、情報共有環境の多言語化やユビキタス化とともに、分野融合研究をはじめとする情報・システム機構関連のすべてに対して提供できる体制を検討する。

平成21年度

プロトタイプシステムおよびデータベースを、その実用性、有効性、適用範囲を明確にして、分野融合研究およびその他の研究・教育の情報基盤として提供する。

大規模で異種の情報空間の結合、分類、分析に関する要素技術や新手法で構成するプロトタイプシステムに関しては、情報空間から知識を汲み出す知識基盤として確立する。また、ライフサイエンスメタデータベースシステムおよび地球情報統合データベースシステムを公開し、融合研究分野をはじめとする学術コミュニティや一般に提供する。さらに、コラボレーションやコミュニティ形成のための情報共有基盤に関しても、これを公開するとともに、研究・教育以外の広範な分野への応用展開を図る。

知識基盤、情報共有基盤およびデータベースに関しては、その継続的な維持、改善を安定的に確保するための運用体制の確立が不可欠であり、このような体制の検討と実施に向けた準備を行うとともに、可能なものについては事業展開を図る。また、本プロジェクトの推進過程において生まれた新しい知見や課題を出発点とした新規プロジェクトを検討し、これらの実施に向けた企画・立案を行う。

4 . 平成17年度研究実施体制

本プロジェクトの目標である分野横断型融合研究のための情報空間・情報基盤の構築を効果的に推進するため、以下の3つのサブテーマを設定して研究開発を行う。

大規模・異種情報の収集・解析・結合・分類の手法および知識基盤の構築

地球・生命などの巨大システム解明のための統合的情報基盤の形成と活用手法の確立

コラボレーションとコミュニティ形成のための情報共有基盤とバーチャルラボの構築

また、これらのサブテーマを推進するための研究実施体制は、平成17年度末の時点で以下のと

おりである。

【プロジェクト総括】

プロジェクト・ディレクター：東倉洋一

【サブテーマ】

研究代表者：[国立情報学研究所] 高野 明彦

共同研究者：[国立情報学研究所] 西岡 真吾、佐藤 真一、相澤 彰子、根岸 正光、
安達 淳、西澤 正己、高須 淳宏

[国立遺伝学研究所] 大久保 公策

[統計数理研究所] 馬場 康維、石黒 真木夫、土屋 隆裕

[新領域融合研究センター] 高久 雅生（国立情報学研究所在勤）

【サブテーマ】

研究代表者：[国立情報学研究所] 藤山 秋佐夫

共同研究者：[国立情報学研究所] 武田 英明、市瀬 龍太郎、北本 朝展、佐藤 真一、孟 洋

[国立遺伝学研究所] 菅原 秀明、城石 俊彦

[国立極地研究所] 神田 啓史、野木 義史、土井 浩一郎、宮岡 宏、
本山 秀明

[統計数理研究所] 長谷川 政美、樋口 知之、上野 玄太、尾形 良彦、
村田 泰章、種村 正美、石黒 真木夫、島谷 健一郎

[東京理科大学] 宮崎 智

[Johns Hopkins 大学] 大谷 晋一

【サブテーマ】

研究代表者：[国立情報学研究所] 新井 紀子

共同研究者：[国立情報学研究所] 上野 晴樹、ウッチチャイ・アンポンナラバス、張 超、
アナ・ハディアナ

[国立遺伝学研究所] 嶋本 伸雄、宮川 宗博、佐々木 裕之、桂 勲

[新領域融合研究センター] 舩川 竜治（国立情報学研究所在勤）

[国立情報学研究所/チュラロンコン大学] タワ・クアンパチュア

[総合研究大学院大学] 高畑 尚之

[(株)オーム社] 森 正樹

5 . 平成 1 7 年度の研究進捗

プロジェクト開始にあたって、プロジェクトの目標と方向性の明確化、サブテーマにおける研究項目と研究計画の具体化を行い、3つのサブテーマ毎の要素技術に関する研究開発を、サブテーマ間の連携と生命・地球システム解明を目指した情報基盤構築という目標を共有して推進した。

サブテーマ1の異種情報の結合・分類手法の研究においては、手法の有効性を検証する準備として、商用の環境情報ポータルで提供されている各種最新情報、環境問題資料集成として出版されている条約、法令、政策文書、議会議事録等の基本資料について、研究目的利用の許諾を得て、実証実験に着手した。これらの異なる情報源をユーザの興味により関連づけて提示する利用環境のプロトタイプを開発した。特に、特許情報と論文情報について、本利用環境の有効性を検討した。また、研究者間のコミュニケーションがますます困難になりつつある生命科学分野で、困難さの原因が固有名称の多用や遺伝子の機能構造に関する自然言語表現にあることに注目して、それらを自動的にアイコン化して分野間のギャップを埋めるジーンアイコン（遺伝子象形文字）プロジェクトをスタートした。文書にデータを盛り込むこのアプローチは、科学的知識の共有に適した新しい文書表現の可能性を示唆している。

さらに、大規模リンケージ情報の収集・分析手法の研究では、リンケージ情報抽出のための基盤技術、リンケージ情報の分析を重点に研究を進めた。リンケージ情報抽出のための基盤技術に関しては、多様な記述形式の混在環境における高精度のリンケージの実現に必要な文字列情報の構造化のための学習手法の研究を進め、文字列データの解析と同一ラベルの要素同士の対応比較処理のための近似的な構文解析法の具体的な適用例として、引用文献文字列を分割し、書誌項目に該当する部分文字列をラベリングするための方法を考案した。また、大規模リンケージの取り扱いにおいて必要となる大量の文字情報からリンケージ候補文字列を高速に検出する処理として、複数の情報源による大規模データベース同士での重複エントリの検出に焦点をあて、サフィックスアレイと呼ばれるデータ構造を利用した高速リンケージエンジンの実現方法を検討した。さらに、リンケージ対象として特に重要な「人物名」を対象として、名前表記が等しい複数の人物名を、同一人物ごとにまとめるためのクラスタリング手法を開発した。

抽出したリンケージ情報の分析に関しては、日本の学会誌論文を対象とした「引用文献索引データベース」(CJP)に基づいて、名寄せ作業・所属機関の同定およびセクター分類方法の検討等を行った。具体的には、分析のための基礎データ作成と併行して、このデータを用いた論文の共著関係の定量的解析を行うことによって、研究ネットワークの実態分析に着手した。また、科研費採択課題における総合領域、複合新領域の分野細目間の関連を分析するための研究者履歴調査（500名弱分）を行い、研究成果データベースからの著者情報の同名他者の識別研究のための基礎データ作成、および、研究者 - 研究分野の解析に必要な日本統計学会を含む6関連学会の名簿情報のデータベース化に着手した。

サブテーマ2においては、主に地球・生命システム解明に向けた情報・統計統合処理と意味抽出の情報基盤形成とポータルサイトの構築を目指して、以下の研究項目を推進した。

ライフサイエンスメタデータベースシステム構築研究では、ライフサイエンス研究全体の効率化を目指した利用者主導型メタデータDBシステム構築を、「新世代バイオポータル」プロジェクトの成果を発展させる形で推進している。平成17年度は、データ量（約29,000件のレコードを収集済）の増大、多様化に継続的に対処可能なシステムの基本設計、および、これらに必要なハードウェアの整備を実施した。また、収集済レコードの整理を進め、データベースの所在情報

と機能情報を連結させる目的で約 40 件について精査し、新たに PubMed とのリンクを設定して日本語バイオポータル(<http://www.bioportal.jp>)から公開した。

極限環境生物統合データベースの構築研究では、国立極地研究所が南極大陸で採集した生物標本を対象に、高精度 3 次元画像、ゲノム情報、分子進化情報等を基礎情報に加えて統合データベース化し、極地を中心とした生物多様性についての新しい研究基盤の構築を進めている。平成 17 年度は画像処理技術の開発を進め、被写体を中心とする半球を画像上で 3 次元表示させるためのソフトウェア開発を行うとともに、撮影用機材の整備と開発を実施した。また、ゲノムと分子進化解析に関しては、極域限定種と、極地および世界各地で採集した汎地球種について葉緑体のゲノム解析を行うこととし、約 6 0 0 0 件のコケ葉緑体 DNA 配列情報の整理を行った。

地球情報に関しては、まず、地球観測データを鉛直的に統合することが、地球システムを総合的に理解するために必要であるという視点に立ち、地球環境ポータルを「鉛直地球ポータル(Vertical Earth)」として具体化した。従来より地球システムの層構造(圏など)毎に蓄積されてきた地球観測データについて、複数の層をまたいだ重ね合わせや統合の実現を目指す。また、地理的に分散した機関をまたいだデータ統合を実現するためにウェブサービスを用いる。具体的には、地理情報システム(GIS)において広く使われつつある Web Map Service (WMS)等のウェブサービスを利用する。この基礎調査、および、利用のための試験的なアプリケーションの開発を行った。

また、地球環境データのの一つとして気象シミュレーションデータのための基盤システムを整備した。鉛直データ統合の実験とデモンストレーションに適したデータを生成する気象シミュレーションのための Linux クラスタ購入、および、気象シミュレータに関する基礎調査を行った。その結果、気象シミュレータとして著名な米国の MM5、その後継プロジェクトである WRF、あるいは気象庁が開発する非静力学モデル NHM などのソフトウェアをインストールして利用することに決定した。

さらに、地理情報システム(GIS)の導入を図り、そのフロントエンド(クライアント)とバックエンド(データベース)を整備することにより、南極に関する地理情報を整備・閲覧するための環境を構築した。構築した地理情報のウェブブラウザ閲覧を可能にするため、ポータルサイト(専用ウェブサイト)を立ち上げた。平成 17 年 8 月 3 0 日に国立極地研究所において、南極 GIS ポータルサイト利用講習会を開催し、同日から 9 月 2 2 日にかけてテスト公開を実施した。さらに、この期間に実際にポータルサイトを利用した所員からアンケートも回収し、今後のシステム構築に役立つための意見を収集した。地図データの整備に関しても、ADD(Antarctic Digital Database)やラスタ化された地形図を取り込んだほか、紙の地形図の一部もデジタル化した。その結果、地形図と地質図や航空写真、衛星写真などを重ね合わせて表示することが可能となった。

サブテーマ 3 のバーチャルラボに関しては、当初計画のどおり、最先端研究者が分散した環境において、距離の隔たりを感じることなく、協調して研究を行うことを可能にする「バーチャルラボ」構築ソフトウェア NetCommons1.0.0 を開発し、GPL オープンソースとして平成 17 年 8

月9日より一般公開した。これまでに本ソフトウェアは4000件以上ダウンロードされ、コアユーザは400人を超えている。

NetCommonsの用途は多岐にわたり、(1)共同研究のバーチャルラボ、(2)学協会活動のポータルサイト、(3)大学等高等教育機関の授業支援、(4)NPOのバーチャルオフィス、(5)学校ポータルサイト、(6)IT企業のイントラネット上のバーチャルラボやグループウェア等に広がった。特に、共同研究基盤として活用している機関としては、東京大学社会科学研究所、東京工業大学COEプロジェクト、国立遺伝学研究所生命情報・DDBJ研究センター、富山県立大学流体力学講座産学連携研究プロジェクトなどがある。特に、平成17年度には、NetCommonsの英語化を終え、国際基督教大学やタイのチュラロンコン大学が試用を開始した。また、国内グループウェアでは先駆けて携帯電話からのアクセス対応を行った。

高等教育向けe-Learningシステムでは、WebELSの充実を計画どおりに実施した。具体的には、パワーポイントコンテンツを容易にWebELSサーバにアップロード出来るインタフェースを開発し、コンテンツに音声やビデオを付加する機能を追加した。この音声・ビデオ付加機能はスライド(コンテンツ)を使用しながらの音声説明のマイク入力やカーソル使用におけるカーソルの動きの付加を可能とする。また、パワーポイントコンテンツを共有した多地点オンライン・ミーティング・システム(WebELS Meeting)を開発・公開した。このシステムは、コンパクト、無料、e-Learningシステムとシームレス、マルチプラットフォーム等の特徴とする。さらに、多地点参加型Whiteboardシステムを開発し、これをオンライン・ミーティング・システムと組み合わせることにより、議論をしながらコンテンツを作成するという作業をインターネット経由で行うことを可能にした。

6. 平成17年度の研究成果

(1) 知見・成果物・知的財産権等

- 「想・IMAGINE」のプロトタイプ

異なる情報源をユーザの興味により自在に関連づけられる新技術「想・IMAGINE」のプロトタイプを作成して、連想計算によるデータベースの動的連携の有効性を確認できた。

- ジーンアイコン(遺伝子象形文字)提案

ジーンアイコンの効果を検証するため、文書中に自動的にジーンアイコンを挿入・表示する簡易サーバを作成した。タンパク名などの専門用語が正しく認識された例については有効であることが確認できた。しかし、実際の論文では同一概念について多用な表現が使用されているため、現在使用中の専門用語辞書のみでは精度が低い。この実用上の限界の解決を目指し、精度をある程度保証できる遺伝子名称のアクロニムに限定して辞書を整備して、実用水準まで高めることを目指す。

- 日本語バイオポータル バイオデータベースショウケース

<http://www.ps.noda.tus.ac.jp/biometadb/db-j/index>

(上記アドレス、名称については変更の可能性有り)

- 地球観測データの提供
 - 台風中心位置を中心に切り取った台風画像（過去 11 年分、約 64000 件）
 - 台風経路データ（ベストトラックデータ（過去 55 年分、約 1470 件）
 - アメダス気象観測データ（過去 30 年分、約 3 億 2000 万件）
 - 台風関連の気象災害データ（過去 34 年分、約 3600 件）
 - 台風関連のニュース記事（過去 3 年分、約 7000 件）
 - ウェブログからのトラックバック（過去 2 年分、約 400 件）
- 過去に作成された地図をデジタル化と南極 GIS への格納
 - 1/25000 の南極 DEM（デジタル標高モデル）5m を約 200 枚デジタル化し、地理情報システム上で表示・検索可能。
- NetCommons ver1.0.0 から ver1.0.6
- WebELS ファミリーとその総合ポータルの公開
 - <http://weblsx.ex.nii.ac.jp/>
 - WebELS：全ての機能を備えたフルバージョンシステム
 - WebELS-NIG：総研大生命科学研究科の要望に特化したサブシステム
 - WebELSx：NIG バージョンを汎用化したもので普及を目指すシステム
 - WebELS Meeting：ユーザ登録不要の Internet 会議システム
- WebELS（WebLS を名称変更）の登録商標を申請。

(2) 成果発表及び著書執筆等

- Hai Yen Siew and Yasumasa Baba: “ Linear regression for a mixture of continuous and binary data, ” The 5th IASC Asian Conference on Statistical Computing, 2005.12.17, 香港大, 香港, 中国.
- Hai-Yen Siew and Yasumasa Baba: “ Regression with some observations below the threshold, ” Second German Japanese Symposium on Classification 2006, 2006.3.7, 日独センター, ベルリン, ドイツ.
- Atsuhiro Takasu, Kenro Aihara: “ Bibliographic Component Extraction from References Based on a Text Recognition Error Model, ” Systems and Computers in Japan, Vol.36, No. 7, pp.1-12, 2005.
- Atsuhiro Takasu: “ A Sequential Labeling Method Using Syntactical and Textual Patterns for Record Linkage, ” Lecture Notes in Computer Science 3686 (Proc. 3rd ICAPR), pp. 199-208, 2005.
- 藤山秋佐夫:「生命の秘密に挑戦するゲノムインフォマティクス～人間とチンパンジーの違いを読み解く」、情報通信ジャーナル, 24, 30-31 (2006)
- 藤山秋佐夫、菅原秀明ら:「ごぞんじですか? 日本語バイオポータルサイト」, 図書館情報, 印刷中, (2006)
- 藤山秋佐夫:「科学コミュニケーションと日本語バイオポータル」 国立遺伝学研究所共同

- 研究会「生物情報資源の相互運用性」, 2006.3.6
- 北本 朝展: "デジタル台風: 異種データストリームを融合した準リアルタイム台風データベース", 画像の認識・理解シンポジウム (MIRU 2005), pp. DE-14, 2005年07月
 - 北本 朝展: "デジタル台風 - 網羅的な台風データベースはどう「使える」のか? -", 東京工芸大学「21世紀 COE プログラム」オープンセミナー, No.37, 2005年11月 (招待講演)
 - 北本 朝展: "Vertical Earth: 地球の鉛直構造をまたぐデータベース/ポータルサイトへの構想", 「宇宙地球系情報科学研究会」「巨大データベース構築に関する研究会」合同研究会, 2005年12月 (招待講演)
 - 北本 朝展: "地理情報資源のデータベースと相互運用の試み", 国立遺伝学研究所研究会「生物情報資源の相互運用性」, 2006年03月
 - 北本 朝展: "デジタル台風: 地上と宇宙の多様なデータを網羅的に統合した台風データベース", 第15回生研フォーラム「宇宙からの地球環境モニタリング」, pp. 49-50, 2006年03月
 - Asanobu KITAMOTO: "Digital Typhoon: Toward the Hub of Typhoon Emergency Information Aggregated from Various Sources", NII International Workshop on Cyber Science Infrastructure, 2006年03月 (in English)
 - NetCommons ver1.0.0 から ver1.0.6
 - WebELS ファミリーとその総合ポータルへの公開 <http://weblsx.ex.nii.ac.jp/>
 WebELS: 全ての機能を備えたフルバージョンシステム
 WebELS-NIG: 総研大生命科学研究科の要望に特化したサブシステム
 WebELSx: NIGバージョンを汎用化したもので普及を目指すシステム
 WebELS Meeting: ユーザ登録不要の Internet 会議システム
 - WebELS (WebLS を名称変更) の登録商標を申請。

7. その他

(1) 実用技術重視の研究推進

特に、研究項目 では、論文、専門辞典、教科書、新聞記事データ、特許情報、条約・法令など多様な情報源の新しい利用環境を追求している。研究成果の実用性を検証しながら研究を進めるため、実際に専門家が業務で参照している高信頼な情報源を導入している。同種の研究においては、論文執筆を優先して、規模や内容が現実の情報源とは大きく異なるものを利用して実験する場合も多いが、本項目は提案技術の実用性を重視して、導入済みの現実の情報源を利用して研究を推進している。

(2) 研究会「大規模データ・リンケージ、データマイニングと統計手法」の開催

研究項目 に関して、2006年2月20日(月)、21日(火)統計数理研究所において標記研

研究会を以下のように開催した。

- 竹田隆治、高須淳宏：「時系列テキストストリームからの単語共起を使った新情報の検出方法」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.45-50，2006.
- 孫媛，西澤正己，根岸正光：「日本の引用文献索引データベースを用いた産学連携の現状分析」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.51-55，2006.
- 西澤正己、孫媛：「科研費採択課題における総合領域、複合新領域の関連分析」，第一回大規模データ・リンケージ、データマイニングと統計手法，2006（口頭発表）.
- 根岸正光：「発表論文からみた研究分野構成の類似度と引用度指標による大学間競争関係析出の試み」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.57-64，2006.
- 相澤彰子：「大規模異種データベース間でのレコード同定手法とその適用例」，大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.85-90，2006.
- 馬場康維：「研究者マップについて」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.91-92，2006.
- 石黒真木夫：「情報抽出」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.107-108，2006.
- 鈴木康平、正田備也、高須淳宏、安達淳：「書誌情報の共著関係を用いた著者同定に関する研究」，第一回大規模データ・リンケージ、データマイニングと統計手法予稿集，pp.109-117，2006.

（３）その他の研究会

- 藤山秋佐夫：「科学コミュニケーションと日本語バイオポータル」 国立遺伝学研究所共同研究会「生物情報資源の相互運用性」，2006.3.6（招待講演）

（４）その他の情報発信

- 「鉛直地球ポータル(Vertical Earth)」プロジェクト<http://earth.nii.ac.jp/>
- 記者発表
平成 17 年 8 月に NetCommons ver1.0.0 リリースに際して、記者発表を実施。
平成 18 年 3 月に、NetCommons を用いた遠隔教育の試みが朝日新聞科学面に掲載。

（５）国際交流

- WebELS に関して、チュラロンコン大（タイ）、ダッカ大（バングラデシュ）等との間に、実利用と共同開発のための研究交流協定を締結。