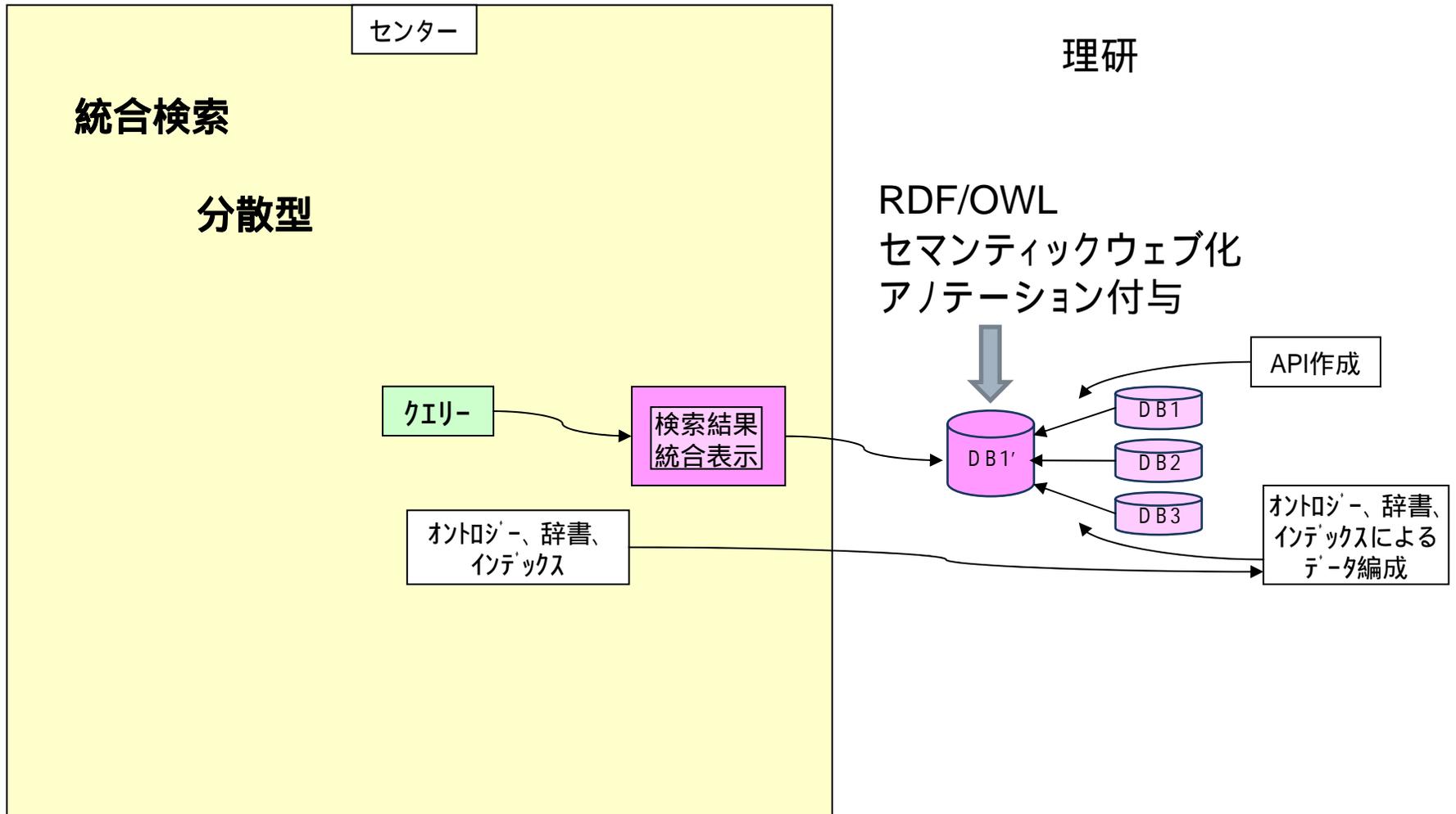


統合データベースプロジェクト作業部会(第3回)
植物オミックス情報および
蛋白質構造情報

- ◆独立行政法人理化学研究所

データベース検索のタイプ



シロイヌナズナトランスクリプトーム

理研DB

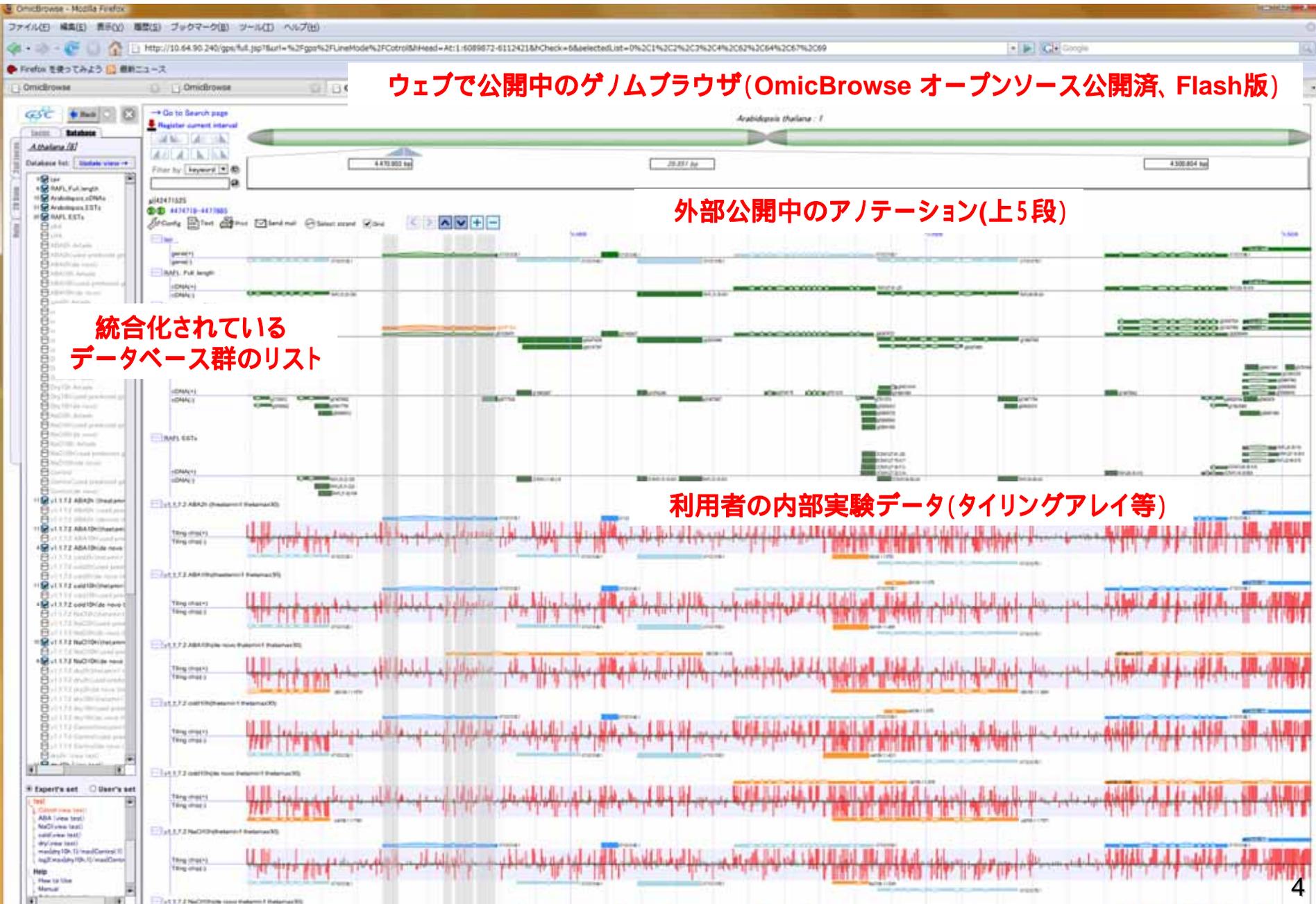
RARGE & RARTF



(1)生物種	シロイヌナズナ
(2)試料・ライブラリー等の種類、数	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19種類（各々FおよびRアレイを用いた計6回のハイブリ実験を行う） GEOデータベースに登録されている、シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データの内、3回以上繰り返し実験を行ったものは、わずか4種類のみである。
(3)測定方法	
(4)データの内容	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19種類（各々FおよびRアレイを用いた計6回のハイブリ実験を行う） 実験内容：播種後2週間目の植物体を用いた乾燥、低温、塩ストレス、ABA処理、再吸水処理による乾燥ストレスからの回復過程など

(1)生物種	シロイヌナズナ
(2)試料・ライブラリー等の種類、数	454シーケンサーを用いたsmall RNAの大量解析データ11種類 454シーケンサーを用いたシロイヌナズナのsmall RNAデータは、これまでに6種類報告されている。
(3)測定方法	
(4)データの内容	454シーケンサーを用いたsmall RNAの大量解析データ11種類 用いた植物材料：播種後2週間目の植物体を用いた乾燥、低温、塩ストレス、ABA処理、無処理の植物体など

シロイヌナズナトランスクリプトーム



シロイヌナズナメタボローム



(1)生物種	シロイヌナズナ
(2)試料・ライブラリー等の種類、数	<ul style="list-style-type: none"> ◇ 野生型および単一遺伝子欠損変異体およそ50サンプルの網羅的な代謝物質量プロファイル ◇ 物質の同定に用いる、標準物質のマススペクトルデータ10,000スペクトル(約1000物質)
(3)測定方法	ガスクロマトグラフィー質量分析計 液体クロマトグラフィー質量分析計 キャピラリー電気泳動質量分析計
(4)データの内容	質量分析計より出力されるマススペクトル
(5)その他、特記事項	計測方法が確立しているため、シロイヌナズナの完全長cDNAの過剰発現体または遺伝子欠損変異体約数百種類のデータを取得することも可能である。また各植物体について出来るだけ多く(数個体以上)のサンプルを計測することが望ましい。

シロイヌナズナフェノタイプ



(1)生物種	シロイヌナズナ (<i>Arabidopsis thaliana</i>)
(2) 試料・ライブラリー等の種類、数	シロイヌナズナのトランスポゾン・タグライン18,000系統と、全てのラインに関するトランスポゾン挿入位置情報。シロイヌナズナ26,000遺伝子のうち5,000以上の遺伝子に関する変異を含んでいると推測される。
(3)測定方法	トランスポゾン挿入部位近傍の塩基配列の決定
(4) データの内容	変異体番号、トランスポゾン挿入位置情報、近傍遺伝子情報

(1)生物種	シロイヌナズナ (<i>Arabidopsis thaliana</i>)
(2)試料・ライブラリー等の種類、数	シロイヌナズナの4000遺伝子の変異体に関する表現型情報。シロイヌナズナ26,000遺伝子のうち4,000遺伝子に関する変異体を調べている。
(3)測定方法	系統的な表現型解析
(4) データの内容	変異体番号、トランスポゾン挿入位置情報、挿入変異遺伝子情報、表現型の画像データ

(1)生物種	シロイヌナズナ
(2)試料・ライブラリー等の種類、数	シロイヌナズナ完全長cDNA遺伝子高発現型変異体は、理研オリジナルの変異体であり、約1万の遺伝子リソースを網羅する。これは、現在報告されている遺伝子の40%にあたる。シロイヌナズナActivation tagging変異体系統は、7万系統あり、シロイヌナズナのほぼすべての遺伝子の活性化をしている数と考えられる。
(3)測定方法	塩基配列決定による遺伝子情報 目視及び計測機器（光合成、色素吸収）による変異形質の情報
(4)データの内容	種子番号 遺伝子番号 遺伝子アノテーション情報 形質情報（光合成、色素、形態） 変異体画像情報

シロイヌナズナのリソースデータ

RIKEN BIORESOURCE CENTER
EXPERIMENTAL PLANT DIVISION 実験植物開発室

News: 植物培養細胞の形質転換と保存に関する技術研修のお知らせ
理研GSCから委託されたリソースのMTA更新について
cDNAクローンのBlast検索が可能になりました

Populus nigra (樹木のモデル)
Arabidopsis thaliana
Lotus japonicus (マメ科のモデル)
Physcomitrella patens (コケのモデル)
Nicotiana tabacum By-2 cell (培養細胞標準株)

あなたは 081557 人目の訪問者です。
(2002.03.01から)

連絡先: 理化学研究所 バイオリソースセンター
リソース開発部 実験植物開発室
tel 029-526-9007 fax 029-526-9053
plant@bric.riken.jp

(1)生物種	シロイヌナズナ
(2)試料・ライブラリー等の種類、数	申請者の保有データが、特定の分野・生物種においてどの程度カバーしているかの自己評価も記述 完全長cDNAクローン(RAFL clone) シロイヌナズナ(エコタイプ: Columbia)のほぼ全ての転写領域をカバー
(3)測定方法	cDNA塩基配列の全長もしくは両端を決定
(4)データの内容	記録しているデータ項目(例えば、試料番号、遺伝子名、発現データ(画像)等) リソース番号、クローン番号、塩基配列のアクセッション番号、遺伝子コード領域のAGI番号、塩基配列

リソース種	収集保存数 (累計)
植物 シロイヌナズナ種子, cDNA, 培養細胞株	251,553株

ストラクチュローム (微生物由来)

(1)生物種	<p>(試料調整・結晶化・回折実験データベース) 9 種類 (<i>Thermus thermophilus</i> HB8, <i>Pyrococcus horikoshii</i> OT3, <i>Escherichia coli</i> K-12, <i>Aeropyrum pernix</i> K1, <i>Sulfolobus tokodaii</i> strain7, <i>Aquifex aeolicus</i> VF5, <i>Geobacillus kaustophilus</i> HTA426, <i>Thermotoga maritima</i> MSB8, <i>Methanocaldococcus jannaschii</i> DSM 2661)</p> <p>(変異体構造解析データベース) 2 種類 (<i>Thermus thermophilus</i> HB8, <i>Pyrococcus horikoshii</i> OT3)</p> <p>(重原子データベース) 不明だが多数</p>
(2)試料・ライブラリー等の種類、数	<p>(試料調整・結晶化・回折実験データベース) 総レコード数11190のうち発現プラスミドがあるもの11021 (9 8 %)。回折画像数300 (変異体構造解析データベース) 変異体総数241種類のうち、99種類を構造決定済み (4 1 %)。 <i>Pyrococcus horikoshii</i> OT3由来PH0725蛋白質 (265残基) については、変異体179種類をプラスミド構築し79種類の結晶構造を決定済み。 <i>Thermus thermophilus</i> HB8由来TTHB049蛋白質 (177残基) については、変異体62種類をプラスミド構築し20種類を解析済み。両蛋白質を通じ、セレノメチオニン化のためのLeu-Met変異はほぼ網羅している。</p> <p>(重原子データベース) 重原子を結合した蛋白質の情報784件を収録する。特に水銀に関しては351件と豊富なデータを有する。現在のところ22種類の重原子をカバーしている。</p>
(3)測定方法	<p>(試料調整・結晶化・回折実験データベース) 実験データベースであるため多種多様である。特に発現精製は多様な機器の出力データ等を含む。結晶化は主に結晶化ロボット T E R A の出力データである。</p> <p>(変異体構造解析データベース) X線結晶構造解析により変異体蛋白質の結晶構造を決定。</p> <p>(重原子データベース) 重原子を結合した蛋白質の結晶構造をX線結晶構造解析により決定。</p>
(4)データの内容	<p>(試料調整・結晶化・回折実験データベース)</p> <p>(基礎情報)</p> <ol style="list-style-type: none"> 試料蛋白質の由来生物種名、遺伝子名、吸光係数、分子量、等電点など、いずれも計算結果等の二次データ。 構築プラスミドデータ (自前のデータ)。ホスト、ベクター、予備発現結果 (5 段階の発現ランク)。 <p>(発現精製)</p> <ol style="list-style-type: none"> 培養情報 (自前のデータ)。発現データ (画像と 5 段階の発現ランク)、発現の諸条件 (誘導状態、培養時間、培養温度、培養量溶液量等)。 精製情報 (自前のデータ)。各タンパク質について精製の諸条件 (懸濁方法、超音波破碎方法、遠心時間、各タンパク質に最適なカラム情報) 加えて精製タンパク質の吸収スペクトルの画像、精製蛋白質のSDSおよびNativeページ画像、精製蛋白質の収量等。精製タンパク質のDLS測定結果 (画像および数値)。選択カラム情報の詳細について：カラム名、緩衝液名、フラクションサイズ、フロー速度、グラジエント方法、溶出濃度、カラムチャート (A280、イオン強度等) の画像、各フラクションのSDSページ画像。 <p>(結晶化)</p> <p>結晶化ロボット T E R A の出力 (自前のデータ)。精製タンパク質の結晶化スクリーニング情報 (結晶化条件、観察画像、 1 0 段階のスコア)。</p> <p>(回折実験)</p> <p>回折実験データ (自前のデータ)。回折画像データとその計算処理のための各種パラメータ。</p> <p>(変異体構造解析データベース) 二種のタンパク質について網羅的な変異体構造解析データベース。結晶化データ、回折実験データ (画像データ、分解能、測定条件等)、回折画像処理データ、精密化データ、構造座標データ (PDB)。</p> <p>(重原子データベース) 自前の構造解析から得たデータ。さらに、文献データおよび登録されているPDBからの計算等による二次データ。内容は重原子実験データで、タンパク質名、重原子名、重原子試薬名、実験方法、沈澱剤名、緩衝液名、pH、文献名、重原子結合サイトの二次構造等。インターフェースとして重原子選択予測機能等。</p>

シロイヌナズナオミックス注釈

- ◆ 理研は、シロイヌナズナを題材に、トランスクリプトーム、メタボローム、フェノーム、リソースデータのアノテーションを、新規の実験データを含めて実施し、データベース化する。その際、ブラウザでの可視化と共に、ダウンロードできる形でDB化する。
- ◆ かずさ、理研、中核との間で、遺伝子名称辞書、オントロジー、ID、データバージョンやアノテーションフロー等の共通化を実施する。
- ◆ かずさと理研の課題は互いに相補的になっているので、かずさのゲノムと理研のトランスクリプトームの間を統合化することによって、ゲノムから、トランスクリプトーム、メタボローム、フェノーム、リソースデータに至る統合化を図る。
- ◆ 統合化方式の枠組みが、本プロジェクトにおいて確立されれば、これをその後、別の植物や植物以外にも適用可能になることを意識して開発を行う。

タンパク質構造注釈

- ◆ ターゲットタンパクPJ (PDBjも)と理研補完課題間の連携を、オントロジーや辞書、API等の共通仕様の観点で行う。
- ◆ 両PJ間の形式上の切り分けが必要。例えば、ターゲットタンパクPJでは、実験情報のオントロジーと辞書構築といった共通仕様に関することを実施し、理研では、具体的なタンパク3000データの実験情報のDB構築を行う。今後検討する。
- ◆ 中核と各機関との連携
オントロジーや辞書、API等の共通仕様の観点で連携する。
テキストマイニングツールの構築と利用に関して連携する。
上記、オントロジーや辞書、API等の共通仕様の観点での連携では、逐次、情報や仕様の交換を実施する。
- ◆ 今後の連携の具体的方法
具体的な連携を行うための各機関担当者によるチームを編成する。
理研(豊田、横山、国島)、九工大、阪大、中核機関それぞれ1名