

個別タスクの実施計画及び成果イメージ(案)

1. データベース統合戦略立案および評価

情報・システム研究機構 事務局

1. データベース統合戦略立案および評価

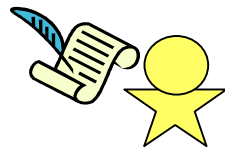
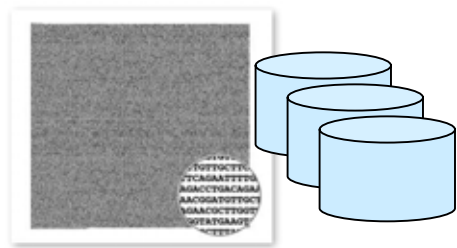
業務計画書項目名	個別課題名	成果の概要
ゲノム注釈とデータベース間の連携における課題	統合データベース間の連携と課題の整理	ゲノムアノテーション間の詳細な分析比較による質情報およびアノテーションの統合利用の方法に関する報告
国内外DBの俯瞰と質的量的比較	国内外DBの俯瞰調査	主に分子に関する国内外の主要なデータベースのリストおよびその維持管理状況やデータサイズなどに関する報告書
ライフサイエンス分野の研究の俯瞰調査	ライフサイエンス分野の俯瞰のためのタイトル処理システム	癌学会・生化学会・分子生物学会を含む国内主要学会の過去10年間の要旨に基づく分野研究の俯瞰マップと閲覧システム
	ライフサイエンス分野の俯瞰のための連携システム開発	ライフサイエンスの知識を俯瞰するためのデータや知識の整理法を開発する目的で、動物の脳に関する機能的、形態的、分子的、進化的なデータ・知識を集め、教科書的な知識と最新の知見をおりませで伝えるシステムの構築
検索アルゴリズムを含めた知識情報技術の動向調査	統合処理技術の動向調査	調査報告書
臨床情報や医療統計の現状調査	医療情報の現状調査	疾患に関わる各種情報の存在調査と共同利用への課題調査報告書
	医療統計の現状調査	臨床情報や医療統計に関するアルゴリズムの調査

「統合データベース間の連携と課題の整理」実施項目(計画)

1. イネならびにシロイヌナズナを対象として、代表的モデル研究植物である両者におけるゲノムアノテーション型のデータベース統合状況に関する詳細な調査を実施する
 - 各公開データベースの品質とアクセシビリティを検討し生物学と情報学の見地から、インターネット上の利用可能なデータベースに対し網羅的な評価を行う
 - 個々のデータベースの開発者ならびに想定されるユーザに聞き取り調査を行い、開発側とユーザー側との双方からの課題を整理する
 - 上記の調査及び分析結果により、代表的なモデル生物ゲノム解析情報を基盤とした統合型データベース群を対象としたデータ統合と提供における種々の課題を明らかにし、今後のデータベース統合の方策についての調査報告書を作成する
2. 調査結果を踏まえ、データベースの更なる統合を目標とした優れたインターフェイス等のプロトタイピングを行い、必要な機能を実証的に検討する
 - 活用可能な公開プログラムについても網羅的な調査を行い、最大限に活用する方策を検討する
3. 引き続き、より高度なDB統合のための課題調査を行い、統合データベースの枠組みとして既存のプログラムでは不足する領域の開発を進める(今後の課題)

「統合データベース間の連携と課題の整理」の成果イメージ

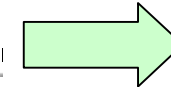
Oriza sativa (イネ) 統合型データベース



実施者



統合DB間の連携状況
に関する調査報告



より進んだDB統合
のための課題調査
と開発
(今後の課題)



より高度なDB統合のため
のプロトタイピング

Arabidopsis thaliana (シロイヌナズナ)

統合型データベース

「国内外DBの俯瞰調査」の実施項目(計画)

1. 内閣府の調査とJSTのDB情報を融合するための調査

[融合対象]

- 1) 内閣府の調査: ライフサイエンスDBの網羅的収集と多次元的な分類
- 2) JST: 専門家によるライフサイエンスデータベースの解説や分類活動

2. 大型のライフサイエンスデータベースの詳細調査

[調査内容]

- 1) データの種類 (内容)
- 2) データ利用のための機能
- 3) 維持管理状況やデータサイズ

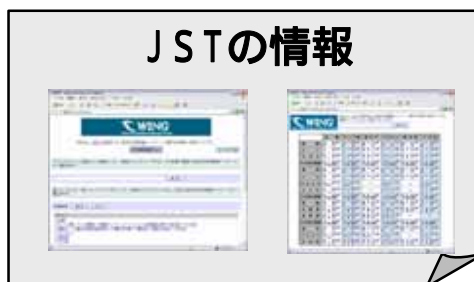
3. データベースカタログの作成

内閣府の調査とJSTのDB情報を融合したデータベースカタログの作成
主に分子に関するデータベースカタログ

「国内外DBの俯瞰調査」の成果イメージ



国内4省庁の250DB
世界の主要な83DB



国内の230DB
世界の主要な140DB

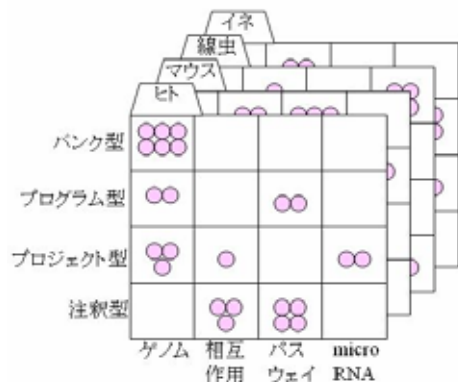


データベース調査情報の融合

DB基本情報

名称	名称	提供機関	URL
HUGE	Human Unidentified Gene-Encoded Large Proteins Analyzed by Kazusa cDNA Project	かずさDNA研究所	http://www.kazusa.or.jp/huge/
HPRD	Human Protein Reference Database	Pandey Lab at Johns Hopkins University	http://www.hprd.org/
mtSNP	human mitochondrial genome single nucleotide polymorphism database	JST/GiB	http://www.gib.or.jp/mtsnp/index_e.html

索引情報



日本語解説

DB	特徴
HUGE	ヒト長鎖 cDNA (KIAA cDNA) 解析情報データベース
HPRD	ヒトのタンパク質について、そのドメイン構造、翻訳後修飾、相互作用ネットワーク、病気との関連性などの情報を集約したデータベース。
mtSNP	ヒトミトコンドリアゲノムの全塩基配列の解析に基づいて、ミトコンドリアゲノムに分布する多型をデータベース化し、多型の相互連鎖、疾患との関連、多型の機能的相違に関する総合的情報を提供

DB特性値

名称	サイズ	アクセス数	更新
HUGE			9/15
HPRD			2/16
mtSNP			9/12

調査報告書

データベースカタログ

データベースカタログの更新

更新(今後の課題)

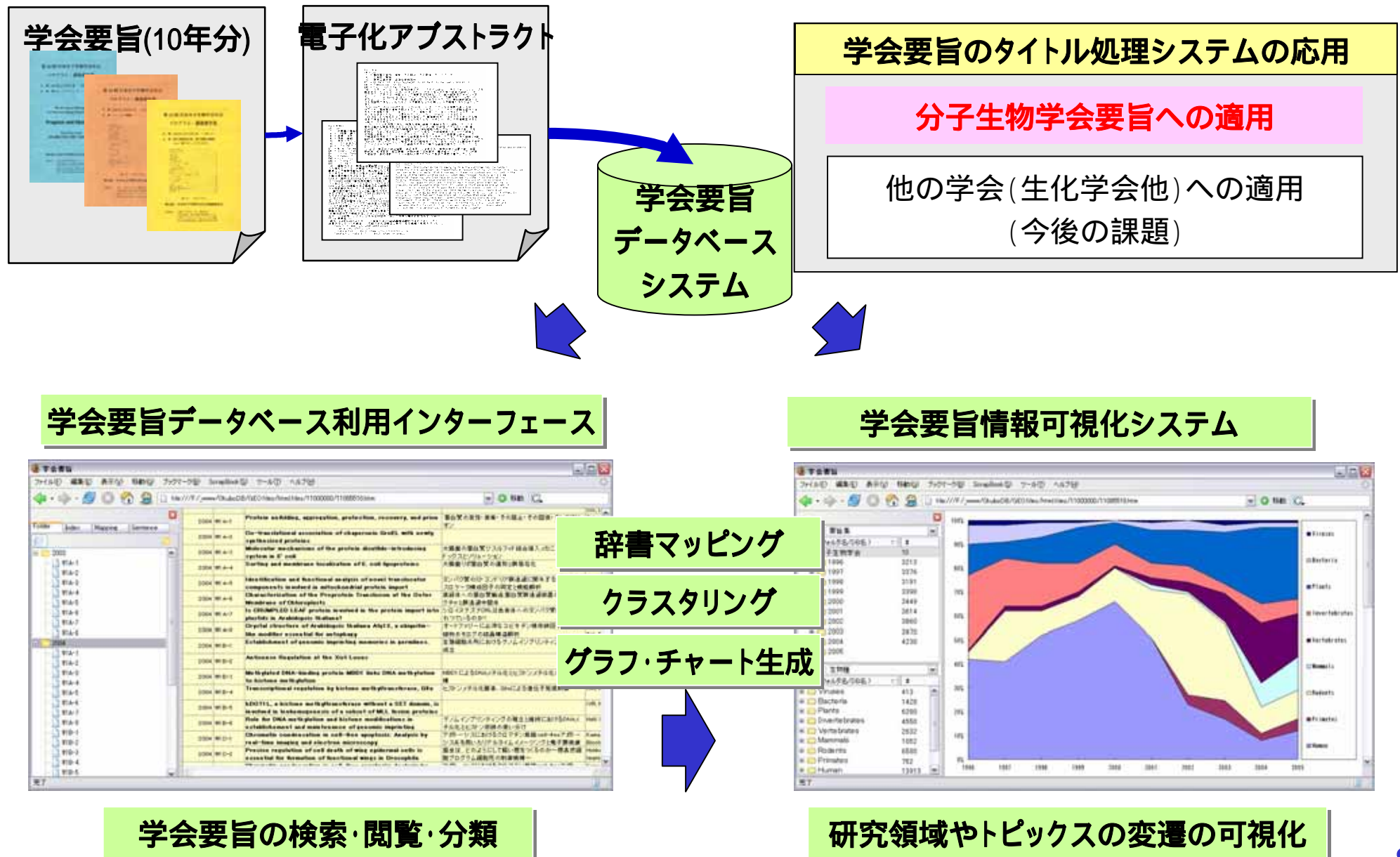
「ライフサイエンス分野の俯瞰のためのタイトル処理システム」の実施項目（計画）

1. **学会要旨データベースシステム**
 - 1) 学会要旨・索引データ登録・検索・閲覧機能
 - 2) 学会要旨クラスタリング機能

2. **学会要旨情報可視化システム**
 - 1) 学会要旨エントリーの情報をリスト表示
 タイトル・サマリー・エントリー表示
 リスト、エントリーイメージ、学会イメージ表示
 - 2) グラフ表示
 - 3) チャート表示

3. **学会要旨情報可視化システムのエンハンス**
 - 1) 解析・レポート機能の追加（今後の課題）

「ライフサイエンス分野の俯瞰のためのタイトル処理システム」の成果イメージ



「ライフサイエンス分野の俯瞰のための連携システム開発」の実施項目(計画)

1. 1,000個の細胞と約40個の遺伝子情報のマトリクスのデータベースを構築
 - 1) 1000個の細胞と40個の遺伝子情報のデータの集積
 - 2) 該当する細胞単位へのデータの反映及び登録
 - 3) 発現する細胞への色付け作業を実施

2. 細胞単位で発現を視覚的に把握することができる(俯瞰できる)ビューワー機能を構築
 - 1) XMLを使用し、3次元に近い、ビューワー機能を構築
 - 2) ユーザー側で、ビューの切り替えとレイヤー別に透明度を変更できる機能を構築
 - 3) ユーザー側で、細胞単位で発現する遺伝子を色づけで把握できる機能を構築

3. 既存サイトと今回拡張する機能も含め、サイトマップを作成する。

「ライフサイエンス分野の俯瞰のための連携システム開発」の成果イメージ

プラナリアの細胞単位で
発現を把握



1000個の細胞と
40個の遺伝子
のマトリクス
データベース

1,000個の細胞と約40の遺伝子情報の数値データ



各細胞の位置を3次元座標で定義

細胞単位の発現を見る

正面 | 背

発現データの3次元モデルへのマッピング



俯瞰できるビュー
機能

約40の遺伝子情報から発現
する細胞を検索

細胞検索 (遺伝子別に反)

細胞番号検索 : 数値入力 【検

遺伝子発現検索 : プルダウン 【約40から選択】 赤 【見る】
プルダウン 【約40から選択】 紫 【見る】
プルダウン 【約40から選択】 緑 【見る】

細胞を階層別に見る

見る階層の透明度の数値ををを入れてください。

1階層 透過度20% 2階層 透過度20% 3階層 透過度20%
4階層 透過度20% 5階層 透過度20% 【見る】

「統合処理技術の動向調査」の実施項目(計画)

1. 統合処理技術に関する以下の情報技術の調査

1 - 1. 検索システムの調査

- ・検索システム研究者にインタビューし、検索システムの動向をまとめる

1 - 2. データマイニングの調査

- ・データマイニング研究者にインタビューし、データマイニングの動向をまとめる

1 - 3. Web2.0の調査

- ・WWW, 書籍を利用して、Web2.0の技術動向をまとめる

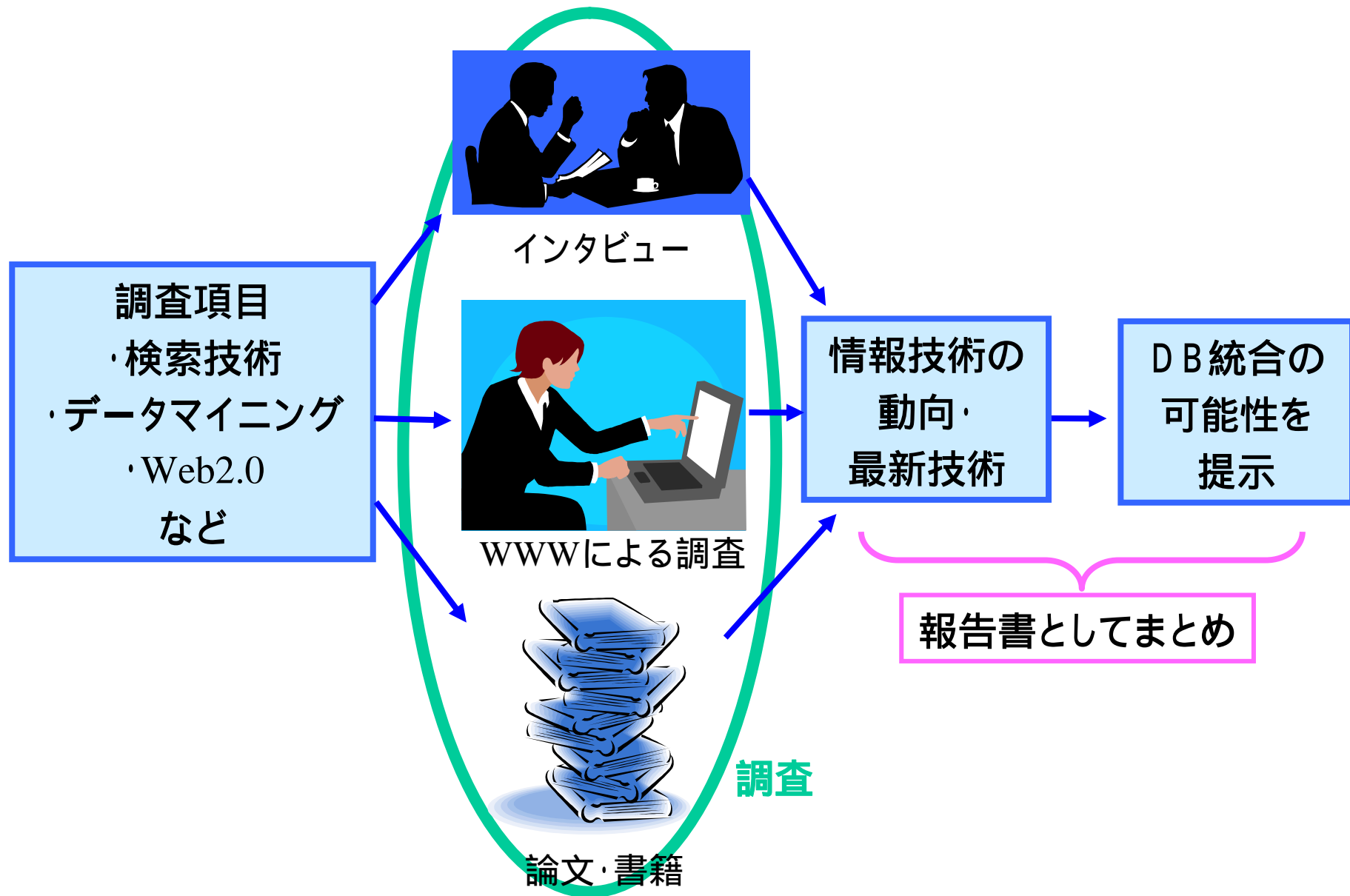
1 - 4. その他の情報技術調査

- ・上記三項目に限定せず、統合データベース構築に関わる可能性のある情報技術について調査する

2. 調査に基づくライフサイエンスDB統合に関する提案

- ・上記四項目の成果から、ライフサイエンスDB統合について、さまざまな角度から提案を行う

「統合処理技術の動向調査」の成果イメージ



「医療情報と医療統計の現状調査」の実施項目(計画)

1. 医療情報の現状調査

- 1) 目的 臨床データの活用を視野に、本プロジェクトにおける医療情報への取組みに関する方向性を探る。
- 2) 内容 電子カルテ、臨床データの規格化、医学統制用語等の調査。
医学文献など臨床研究で必要となる情報源の調査。
臨床データ活用の現状と活用上の問題点の洗出し。
- 3) 方法 専門家インタビューならびに文献、インターネット調査を行う。

2. 医療統計の現状調査

- 1) 目的 医療統計に関わるデータ産生の現状と解析手法の調査を行い、本プロジェクトにおける取組み指針を得る。
- 2) 内容 遺伝子多型解析に関わる研究手法の分類と解析アルゴリズムの調査。
遺伝子発現解析に関わる解析手法及び解析アルゴリズムの調査。
現行アルゴリズムの問題点と将来のニーズ、方向性の把握。
- 3) 方法 関連研究者による会議、専門家インタビュー、ならびに文献、インターネット調査を行う。