

ライフサイエンス分野の統合データベース整備事業  
『統合医科学データベース構築方式の開発』



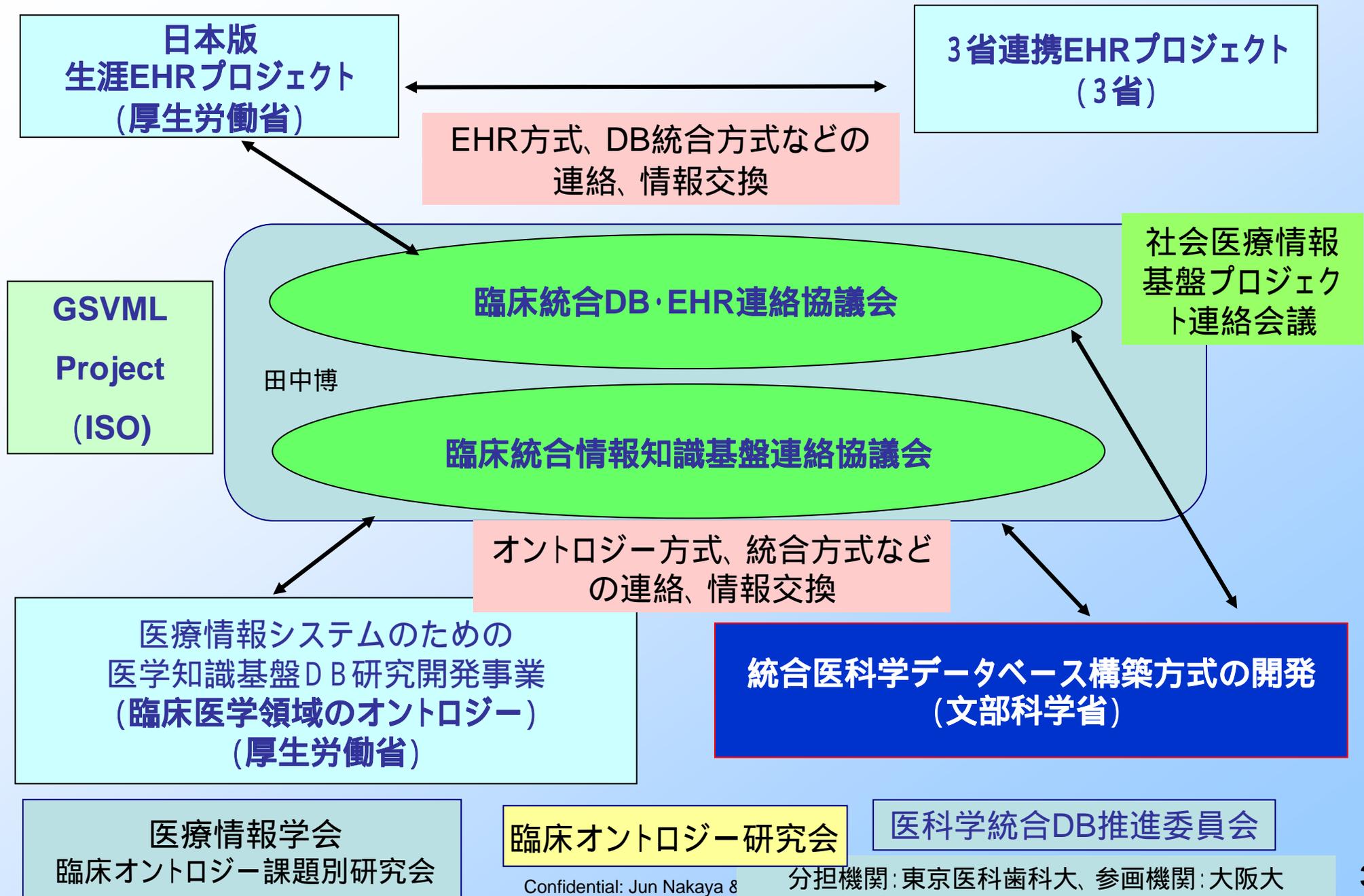
東京医科歯科大学 田中 博

平成20年4月3日

## 事業目標

- わが国の医科学分野のデータベースの統合化を行う
  - 2大疾患(がん・神経難病)から全疾患への対象拡大
    - 臨床DBの全国調査
- 疾患DB統合利用システムの構築
  - わが国に存在する臨床各分野の疾患データベースを統合的に利用可能とする情報環境を構築する。そのための方式を開発する
    - 他の関連プロジェクト(厚労省、総務省など)との連携
    - 医科学統合化推進委員会の設置(疾患DBの担当者の連絡会)
  - 疾患臨床オントロジーの開発連携
- 医科学データベースの高度化
  - オミックスと臨床・病理情報を統合した新しい疾患区分による統合
- ヒト症例データベース公開における社会倫理課題の検討

# 統合医科学データベースプロジェクト体制と関連主要プロジェクトとの連携



## 平成19年度 成果目標 と 実績成果

- 臨床疾患領域における、**統合DB利用システムの基本設計**を確立する。
- 日本において多施設からの**臨床情報の所在**を公開の社会的制約に関する**倫理的、法的問題**を調査・整理する。
- 疾患統合DBシステムの概念を提示した**事例DB**を公表する

具体的な目標

平成19年度の実績成果

1. **要件定義(1/2年)**
2. **平成19年度中に試験公開を行う**
  - **がん症例(100例程度)**
  - **パーキンソン病症例(100例程度)**

1. **要件定義書(1/2年)**
2. **平成19年度中に試験公開中**
  - **がん症例(100例程度)**
  - **パーキンソン病症例(100例程度)**
3. **国内臨床DBおよび倫理基準の調査**
4. **統合医科学検索プロトタイプの試作**

# 国内臨床疾患DBの調査方法

目的: 要素DBの把握、要素倫理基準の把握、国内技術傾向の把握と要素標準技術の選定

## 外部公開情報の収集

- |               |  |
|---------------|--|
| <b>方法</b>     | <b>対象</b>  |
| ・インターネットによる検索 | ・各研究機関のHP及びDB                                    |
| ・関連機関からの資料入手  | ・各省庁の科学研究費成果報告一覧DB<br>例) 文部科学省科学研究費補助金採択DB公開状況一覧 |
|               | ・研究機関が発行する調査報告書<br>例) 学術情報DB実態調査報告書(国立情報学研究所)    |
|               | ・関連ポータルサイト                                       |
|               | ・その他、個別調査により判明したDB                               |

## 非公開情報の独自調査

- |              |           |
|--------------|-----------|
| <b>方法</b>    | <b>対象</b> |
| ・アンケート郵送調査   | ・医学系学会    |
| ・アンケートWEB調査  | ・大学付属病院   |
| ・アンケートFAX調査  | ・国立病院     |
| ・ヒアリング調査(一部) |           |

## 実態調査のまとめ

- 方法**
1. 調査項目に沿って、各データベースについて外部公開情報を収集する
  2. 非公開情報のアンケート回答を調査項目に沿って収集する
  3. 両者をまとめた上で総括し、実態と今後の活動課題に即した活動方針を策定する

## 調査項目

### 概要

- DB概要 (データ件数、内容、方式、データ形式、アクセス件数)
- ポリシー (公開基準、倫理規定、個人情報保護基準)
- 匿名化 (上記個人情報保護基準に準じた具体的方法)
- 拡張性 (多施設連携性、標準化基準(ISO,HL7等))
- オミックスデータ連携性(オミックスデータの有無)
- データの中身 (内容、構造、項目)

- (1) ライフサイエンス分野の統合データベース整備事業(以下本事業)に対する認知状況
- (2) 本事業を知った情報経路
- (3) 回答者の所属する組織の臨床・疾患データベース(以下DB)の保有の有無とその個数
- (4) 保有する臨床・疾患DBの種別  
(対象疾患・オミックスデータ保有の有無)
- (5) 保有する臨床・疾患DBの概要
- (6) 保有する臨床・疾患DBのデータボリュームと関連する事項
  - (6-1) 登録症例数
  - (6-2) 症例の項目数
  - (6-3) 症例追跡期間
  - (6-4) 登録遺伝子数
  - (6-5) 遺伝子の項目数
  - (6-6) データ容量
  - (6-7) アクセス数
  - (6-8~10) その他任意記載項目
- (7) 保有する臨床・疾患DBのデータ保存フォーマット
- (8) 保有する臨床・疾患DB利用時のアクセス手段
- (9) 保有する臨床・疾患DBの公開有無とそのURL
- (10) 保有する臨床・疾患DBデータの更新頻度
- (11) 保有する臨床・疾患DBの該当する法令又は運用指針の有無および生命倫理面・安全面又は個人情報の保護のために講じている措置
- (12) 本事業によるDB参照の許諾の可否
- (13) 上記(12)で参照不可となった場合の理由
- (14) 他の組織とのDB連携の有無とその方法
- (15) 保有する臨床・疾患DBの医科学オントロジーへの準拠の有無とその内容
- (16) 回答者の所属する組織以外の臨床・疾患DBの紹介依頼項目
- (17) 本事業への期待やご意見、ご質問等の自由記述欄

# 実際に用いられた Web調査票(一部)

臨床・疾患データベースの統合化に関するアンケート調査票

東京医科歯科大学 統合データベース整備事業「疾患DB実情調査」  
文部省統合データベースプロジェクト  
ライフサイエンス分野の統合データベース整備事業

平成19年10月吉日

平成19年度文部科学省「ライフサイエンス分野の統合データベース整備事業」  
アンケートへの回答のお願い

文部科学省「ライフサイエンス分野の統合データベース整備事業」分担機関  
東京医科歯科大学 情報医科学センター長 田中 博

拝啓 時下ますますご清栄のこととお慶び申し上げます。

東京医科歯科大学では我が国のライフサイエンス関係のデータベースの統合的な利用環境の構築を図るため文部科学省の委託を受け、平成19年度より「ライフサイエンス分野の統合データベース整備事業」を実施しています。(http://lifesciencedb.mext.go.jp/)

東京医科歯科大学は本事業の「臨床疾患統合データベース開発」を担う分担機関として、散在する臨床・疾患データベース(DB)を統合的に利用できる環境を開発して、分子から個体にいたる多階層的で疾患横断的な臨床医学データの有効利用を広く可能にする総合的な利用システムの開発を進めています。

本アンケートは、我が国の臨床・疾患データベースの現状把握と統合化のための意見収集を目的として実施しております。

11. 貴組織が実運用中または計画中のデータベースについて、該当する法令又は指針がある場合には、該当するものすべてにチェックをつけてください。法令、指針等に定められる手続以外にも貴組織が必要と認めて講じる生命倫理面・安全面または個人情報等の保護についての具体的な措置がございましたら、「18. その他」の欄にご記入ください。

- 1. 医療法
- 2. 医師法
- 3. ヒトゲノム・遺伝子解析研究に関する倫理指針
- 4. ヒトに関するクローン技術等の規制に関する法律
- 5. ヒトに関するクローン技術等の規制に関する法律施行規則
- 6. 特定胚の取扱いに関する指針
- 7. ヒトES細胞の樹立及び使用に関する指針
- 8. 遺伝子組換え生物等の使用等の規制による生物の多様性の確保に関する法律
- 9. 遺伝子治療臨床研究に関する倫理指針
- 10. 疫学研究に関する倫理指針
- 11. 臨床研究に関する倫理指針
- 12. ヒト幹細胞を用いる臨床研究に関する指針
- 13. 動物の愛護及び管理に関する法律
- 14. 実験動物の飼養及び保管並びに苦痛の軽減に関する基準
- 15. 研究機関等における動物実験等の実施に関する基本指針
- 16. 医薬品の臨床試験の実施の基準に関する省令
- 17. 医薬品の臨床試験の実施の基準に関する省令の一部を改正する省令

# 非公開情報の調査対象(詳細)

外部公開情報の収集

方法  
・インタ  
検索  
・関連機  
入手

## 医学系学会

・調査対象...日本医学会(1件)及びその分科会として加盟している**医学系学会102件**

・回収数: 現在のところ、**50件(49%)**

・調査期間...郵送によるアンケート期間  
平成19年10月31日～11月12日

WWW上でのアンケート期間  
平成19年11月13日～11月21日

## データのまとめ

方法

1. 調査項目に沿って、各データベースについて(二次データ)を収集しまとめる

非公開情報の独自調査

方法

- ・アンケート郵送調査
- ・アンケートWEB調査
- ・アンケートFAX調査
- ・ヒアリング調査(一部)

対象

- ・医学系学会
- ・大学附属病院
- ・国立病院

## 国公立病院/国公立大学附属病院

・調査対象...WWW上にホームページを公開している国公立病院/国公立大学附属病院**300件中**、メールでコンタクトできた**139件**

・回収数...現在のところ、**10件(7.2%)**

・調査期間...WWW上アンケート期間  
平成19年11月22日～12月4日

## 調査結果 (一部抜粋)

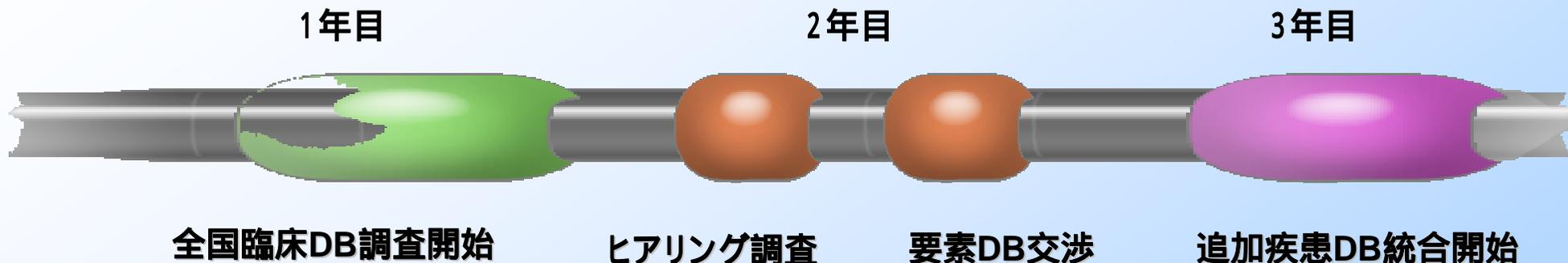
国内の臨床・疾患関連データベースの実態調査															東京医科歯科大学							
															最終更新日: 2023年12月13日							
No.	研究機関名称	部門/学部等	DB名称(和名)	DB名称(英名)	URL	提供機関	所属	調査方法	目的	内容	方式	データ形式	システム	アクセス	公開範囲	更新頻度	連携性	匿名化	セキュリティ	データの連携	データの共有	備考
N030	国立感染症研究所	エイズ研究センター	HIV 感染症臨床データベース	DBV-DB	https://dbiv.ri.go.jp/index.html	国立感染症研究所	国立	Web調査	研究	HIV感染症の臨床経過に関するデータベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース
N031	東京医科歯科大学	無し	無し	CGED (Cancer Gene Expression Database)	http://dbicrion.sc.gcf.ac.jp/	国立	Web調査	研究	がん遺伝子発現データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N015	北海道大学	無し	無し	無し	無し	国立	Web調査	研究	がん遺伝子発現データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N017	大阪府立成人病センター研究所	感染症部門	無し	無し	http://www.frc.or.jp/	国立	Web調査	研究	感染症臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N027	東京都立総合医療センター	がん研究センター	無し	無し	http://www.tokai.ac.jp/	国立	Web調査	研究	がん臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N037	名古屋市立大学	がん研究センター	無し	無し	http://www.nyu.ac.jp/	国立	Web調査	研究	がん臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N034	九州大学	がん研究センター	無し	無し	http://www.kyushu-u.ac.jp/	国立	Web調査	研究	がん臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N036	名古屋大学	がん研究センター	無し	無し	http://www.nagoya-u.ac.jp/	国立	Web調査	研究	がん臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N038	国立がんセンター研究所	がんゲノムデータベース	Genome Medicine Database of Japan (GMDJ)	https://www.genome.jp/gmdj/	国立がんセンター研究所	国立	Web調査	研究	がんゲノムデータベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N032	早稲田大学	医学部	ヒト胎児発育データベース	Human Fetal Development Database	公開範囲: 一部公開	早稲田大学	国立	Web調査	研究	ヒト胎児発育データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース
N050	大阪大学	医学部	Japanese National Cancer Database	http://www.jncdb.org/	大阪大学	国立	Web調査	研究	がん臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N034	日本神経学会	神経学	神経学臨床データベース	CADIS (Cerebral Arteriovenous Disease Information System)	https://www.cadis.or.jp/	日本神経学会	協会	Web調査	研究	神経学臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース
N053	早稲田大学	医学部	Genetic Disease Information Center Database	http://www.kuhp.kyushu-u.ac.jp/genetic/index.html	早稲田大学	国立	Web調査	研究	遺伝性疾患データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N061	日本循環器学会	循環器科	日本成人冠動脈造影データベース	Japan Adult Coronary Artery Database	http://www.jacardb.com/	日本循環器学会	協会	Web調査	研究	冠動脈造影データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース
N047	日本産科婦人科学会	産科	産科臨床データベース	http://www.jstcc.or.jp/	日本産科婦人科学会	協会	Web調査	研究	産科臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N052	国立がんセンター研究所	がんゲノムデータベース	GMDJ (Genome Medicine Database of Japan)	http://www.genome.jp/gmdj/	国立がんセンター研究所	国立	Web調査	研究	がんゲノムデータベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N043	早稲田大学	医学部	臨床情報データベース	http://www.kuhp.kyushu-u.ac.jp/clinical/index.html	早稲田大学	国立	Web調査	研究	臨床情報データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N044	日本産科婦人科学会	産科	産科臨床データベース	http://www.jstcc.or.jp/	日本産科婦人科学会	協会	Web調査	研究	産科臨床データベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	
N053	国立がんセンター研究所	がんゲノムデータベース	Genome Medicine Database of Japan (GMDJ)	https://www.genome.jp/gmdj/	国立がんセンター研究所	国立	Web調査	研究	がんゲノムデータベース	Web	CSV	Web	公開範囲: 一部公開	更新頻度: 不定期	連携性: なし	匿名化: なし	セキュリティ: なし	データの連携: なし	データの共有: なし	特設ページあり	研究内容に関するデータベース	

- 内容まで良く見て臨床DBかどうか判断した
- 各DBの中の内容を可能な限りよく見て吟味した
- アカウントの取れるものは、実際に使用してみた
- 相手に十分説明し相手との意識合わせをしたうえで、質問に回答していただいた

間違いなくこれから発展してゆく分野  
なので期待している (数件)

# 調査まとめとロードマップ

## 疾患DB調査ロードマップ



統合対象のスクリーニング(1年目)  
572件

	DB_A	DB_B	DB_C	.....					
連携統合に向けた詳細調査 (2年目以降)	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Red	Red	Red	Red	.....	White	White	White	White
	Purple	Purple	Purple	.....	White	White	White	White	
	Purple	Purple	Purple	.....	White	White	White	White	

45件

国立病院・国立大学病院(300件)では有効回答(10件)中7割、医学会総会(102件)の調査では有効回答中(50件)の2割が臨床疾患DBを保有していた。また、情報を外部に公開している医学系DB(170件)のうち1割弱は臨床疾患DBの要素を持っていた。

総数572件を対象とした1次スクリーニングを終えたが、より具体的な統合の可能性の検討と統合の推進、詳細の検討のために、より深い調査ヒアリングが今後45件のDBについて必要である。



## 統合医科学データベースWEBサイトのご紹介(デモ)

### 統合医科学データベース(IBM D)

<http://ibmd.tmd.ac.jp>

- 統合医科学データベースの説明
  - 概要、ミッションステートメント、倫理規定、組織体制、お問い合わせ、関連リンク(文科省、中核機関)
- 要素データベース検索
  - 試験公開
  - 各要素データベース(がん、パーキンソン病)への**自動ログイン**
  - 各データベース独自の機能の利用(検索等)

# 試作プロトタイプによる統合検索デモ

(注: 今年の成果目標ではなく、2年後の成果目標です)

## • シナリオ1

### – 検索条件「症状:便秘」

#### • がんDB

- <D> 症状に直接「便秘」と記載のある 3症例がヒット
- <S> 0症例

#### • パーキンソン病DB

- <D> 症状に直接「便秘」と記載のある 35症例がヒット
- <S> 0症例

<D> Direct

<S> Semantic

# 試作プロトタイプによる統合検索デモ

(注: 今年の成果目標ではなく、2年後の成果目標です)

## • シナリオ2

### – 検索条件「症状: 消化器症状」

#### • がんDB

- < D > 0 症例
- < S > シソーラスオントロジーにより、「消化器症状 便秘」と意味変換した 3 症例がヒット
- < S > 同じく、「消化器症状 食欲不振」 2 症例がヒット

#### • パーキンソン病DB

- < D > 0 症例
- < S > 同じく、「消化器症状 便秘」 35 症例がヒット
- < S > 「消化器症状 食欲不振」 0 症例

# 試作プロトタイプによる統合検索デモ

(注: 今年の成果目標ではなく、2年後の成果目標です)

## • シナリオ3

### – 検索条件「症状:うつ状態」

#### • がんDB

- <D> 0症例
- <S> うつ状態のオントロジーによって、「うつ状態 食欲不振」という意味変換によって症状に食欲不振を持つ2症例がヒット。 2症例
- <S> 同様に意味変換し、「うつ状態 からだのたるさ」 4症例(うち1症例は食欲不振と重複しているので、全5症例の表示となる)

#### • パーキンソン病DB

- <D> 0症例
- <S> うつ状態のオントロジーによって、「うつ状態 動作緩慢」という意味変換を実行し、初発症状に動作緩慢持つ結果がヒット。 6症例

# 試作プロトタイプによる統合検索デモ

(注: 今年の成果目標ではなく、2年後の成果目標です)

## • シナリオ 4

### – 検索条件「症状:動作緩慢」

#### • がんDB

- <D> 0 症例
- <SS> 「動作緩慢 うつ状態 食欲不振」という二段の意味変換によって症状に食欲不振を持つ2症例がヒット。 2症例
- <SS> 同様に二段変換、「動作緩慢 うつ状態 からだのたるさ」4症例(うち1症例は食欲不振と重複しているので、全5症例の表示となる)

#### • パーキンソン病DB

- <D> 症状に動作緩慢をそのまま持つ症例がヒット。 6症例
- <S> 0 症例

<SS> 二段Semantic

# 試作プロトタイプによる統合検索デモ

(注: 今年の成果目標ではなく、2年後の成果目標です)

## • シナリオ5

### – 検索条件「すべて:うつ状態」

検索条件のカテゴリ指定を行わず、全検索を行います

#### • がんDB

– <D> 0症例

– <S> 一段の意味変換 **症状**「うつ 食欲不振」 2症例

– <S> 同様に、**症状**「うつ からだのだるさ」 4症例

(うち1症例は食欲不振と重複しているため、全5症例の表示となる)

#### • パーキンソン病DB

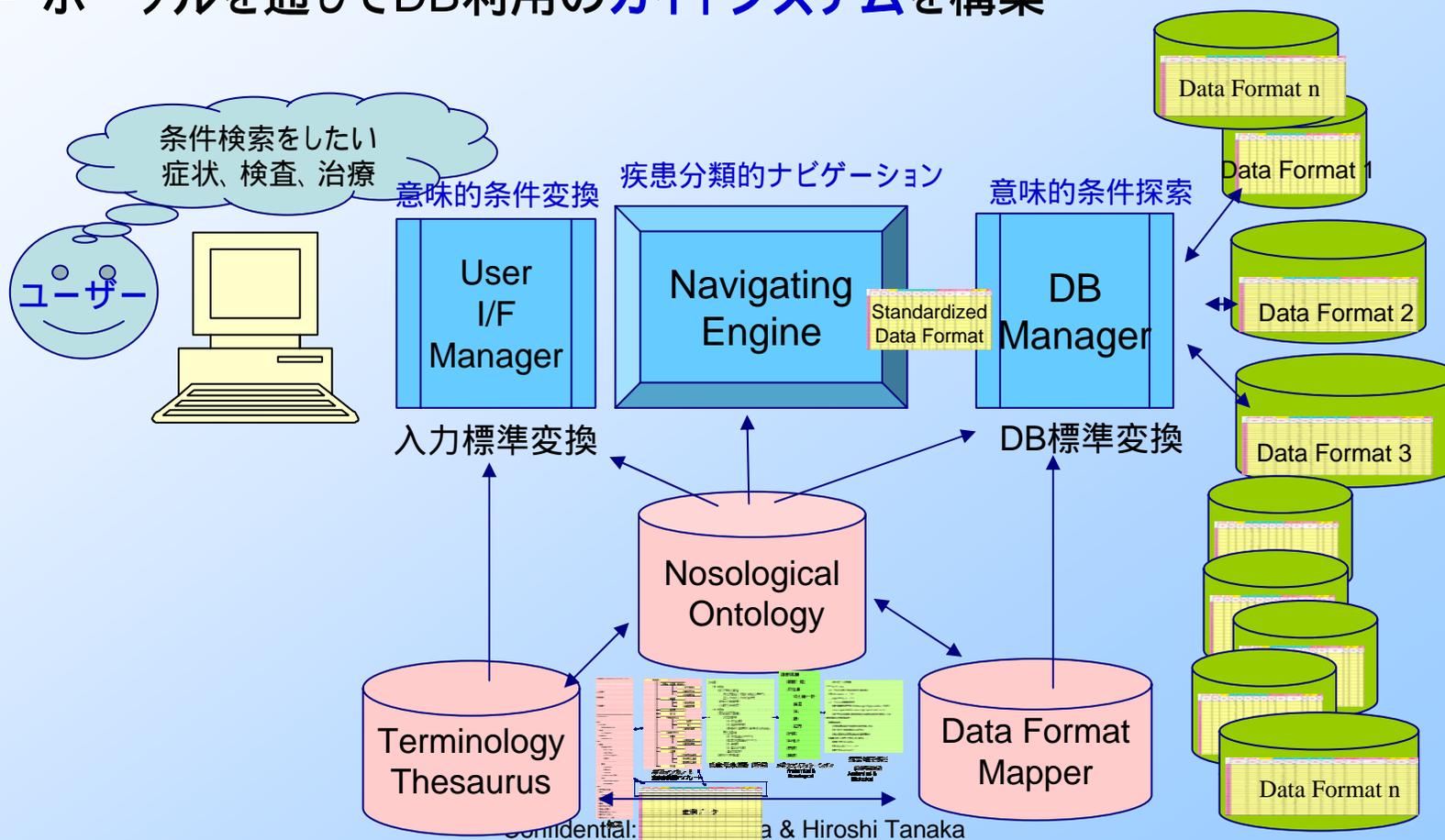
– <D> 0症例

– <S> 一段の意味変換 **症状**「うつ 動作緩慢」 6症例

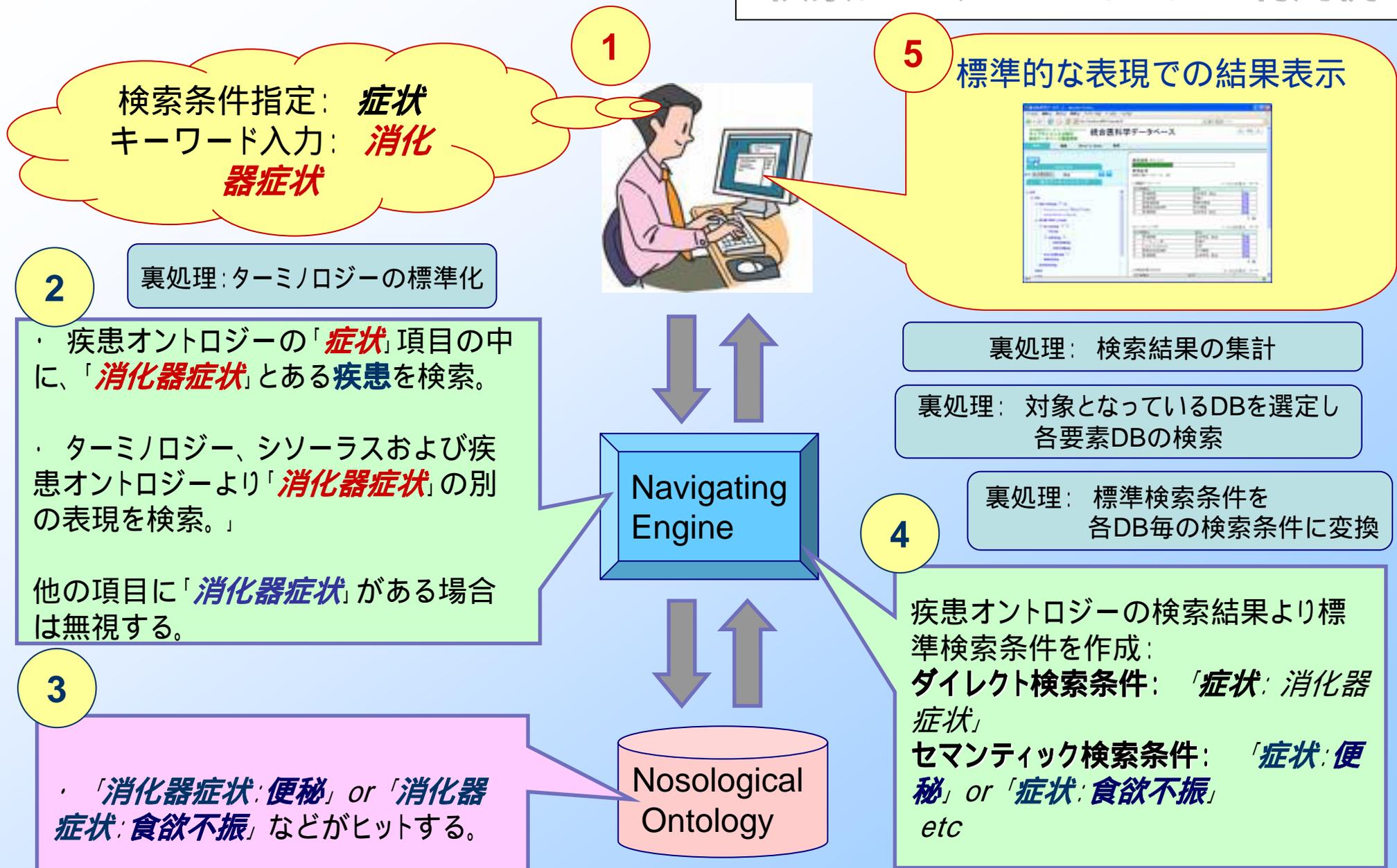
– <S> 一段の意味変換 **治療**「うつ 坑うつ薬(三環形)」 15症例

# 統合医科学DB利用ガイドシステムの基本設計

- コア疾患DB (2個 + (10個程度)を想定 : 全国調査により選定)
  - 統一的な検索を可能とする**統合利用システム**を構築
- 周辺疾患DB (数10個を想定)
  - ポータルを通じてDB利用の**ガイドシステム**を構築



# 検索におけるオントロジー利用例



### 今年度(平成20年度)の目標

- **要件定義書(2/2年)の完成**
  - 統合医科学データベース倫理案を創案する
  - 統合検索に向けた技術・社会倫理要件を、網羅整理する
  - 臨床データベース倫理規定の国際調査の継続と深化
    - HIPPA18項目、匿名化基準(SSA)
- **症例公開**
  - がん(肝臓がん、大腸がん、口腔がん)
    - 100症例      **200症例**
  - 神経疾患(パーキンソン病)
    - パーキンソン病 100症例      **神経疾患 200症例**
- **プロトタイプの進化と試験運用**
  - セマンティックリンクの拡充(意味変換語彙の整理拡充)
  - 臨床オントロジーの拡大(2疾患 全疾患)
- **要素データベースの拡充**
  - レセプトデータベース(レセプトデータ37万件)の連結を公衆衛生科学院と検討中
    - 背景: レセプトデータの電子申請義務化、特定検診の電子申請義務化

ありがとうございました