

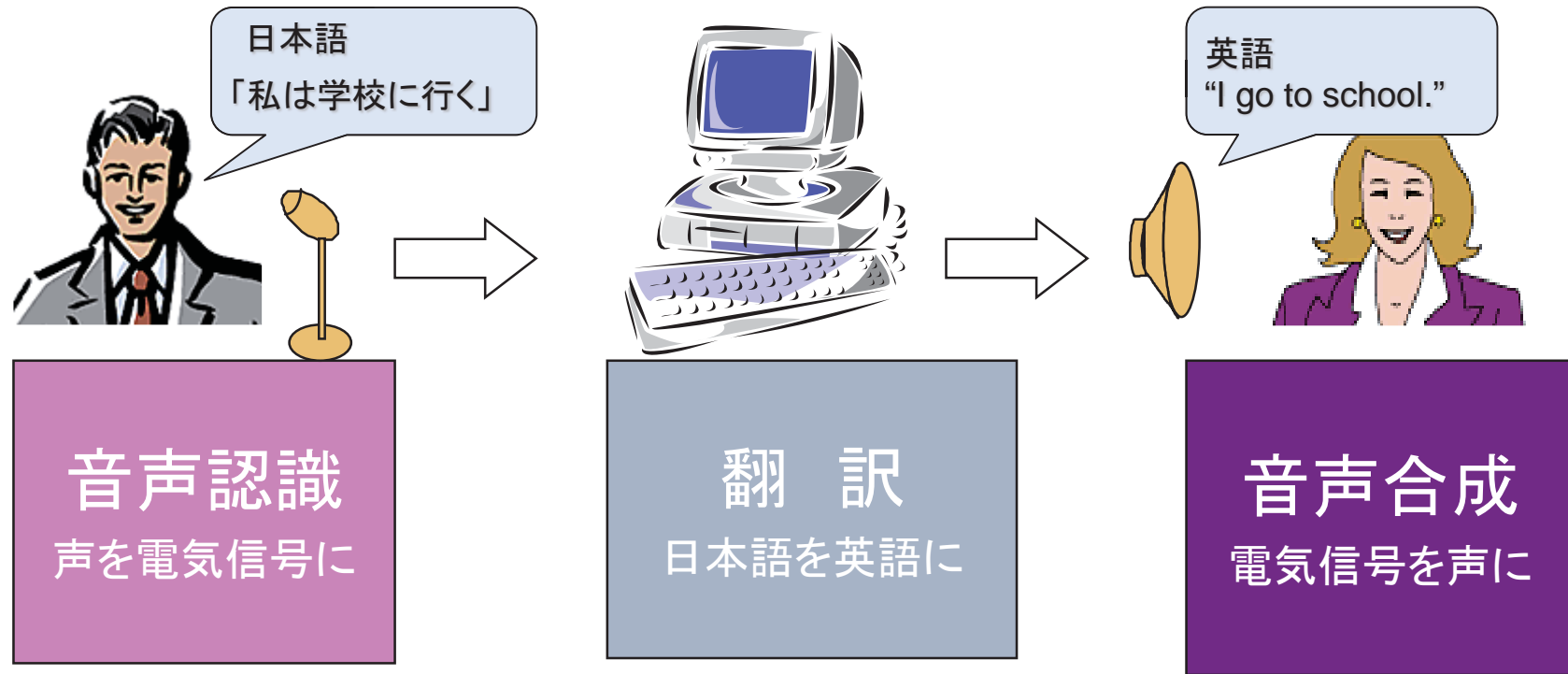
音声翻訳研究の変遷と課題

中村 哲

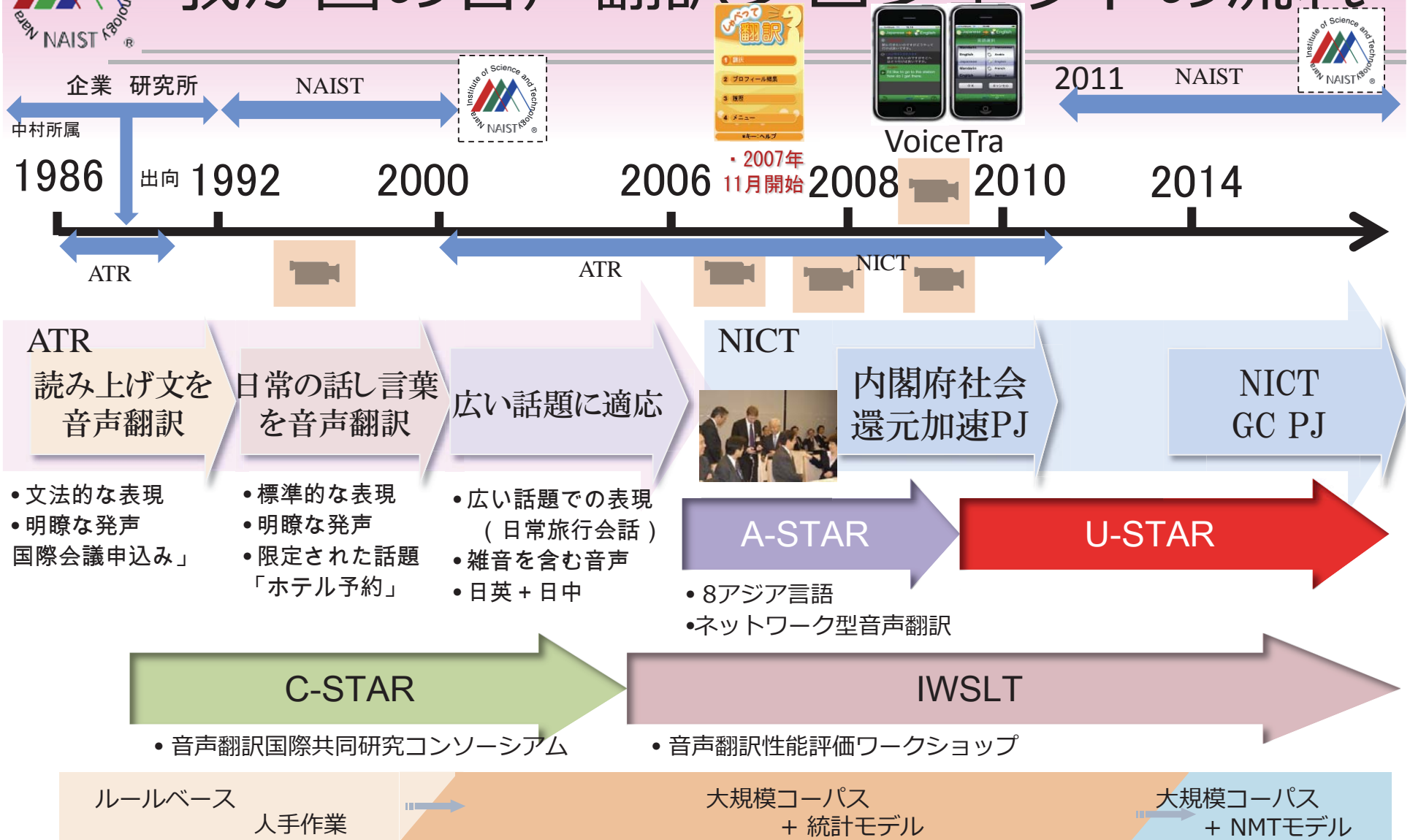
Satoshi Nakamura

奈良先端科学技術大学院大学

自動通訳：音声翻訳のメカニズム



我が国の音声翻訳プロジェクトの流れ



- ATR BTEC 旅行会話日英(100万文対)、日中(50万対訳)、ATRA-Set 音素バランス5240単語20名、ATR B-Set 音素バランス503文10名、ATR C-Set バランス216単語、503文、男女各約150名、ATR F-Set 音声合成用、ATR SDB 日常会話、音素バランス503文194話者、ATR SDB 日英通訳仲介会話、ATR 5240名多数話者データ、APP(対話)、APP-BLA(バランス文)、APP-DIC、ATRECCS 音声合成用データ、ATR中国語、英語音声データ(各500名)
- NICT 音声、言語データ(ALAGIN) 日本語高齢者音声100名、小学生音声100名、中国語100名、非母語英語音声40名、京都観光対話100名、日英中バイリンガル独和約50名

技術的なブレークスルー

▶ コーパスベース

- 大規模なコーパスと統計的手法により学習とデコードが同一の枠組みで

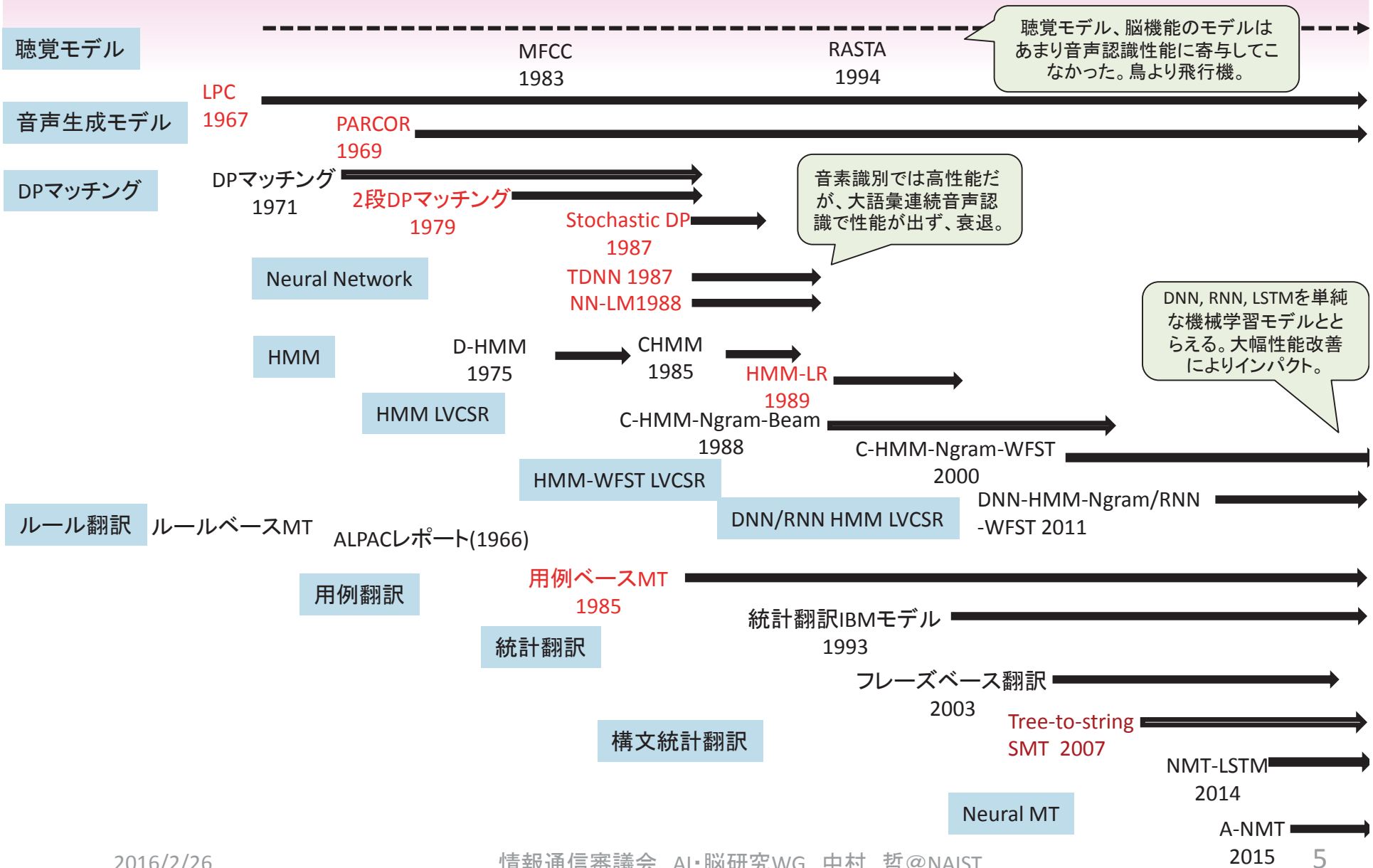
▶ 音声認識

- LPC分析 [Itakuraら 1967]、PARCOR分析 [Itakuraら 1969]
- DPマッチング [Sakoeら 1971]
- 確率モデル
 - Stochastic DP [Nakagawaら 1988]
 - 隠れマルコフモデル、N-Gramモデル [Mercer, Jelinek, Bahl, 1983]
 - 大量のデータの収集、試行サービスによるデータ収集

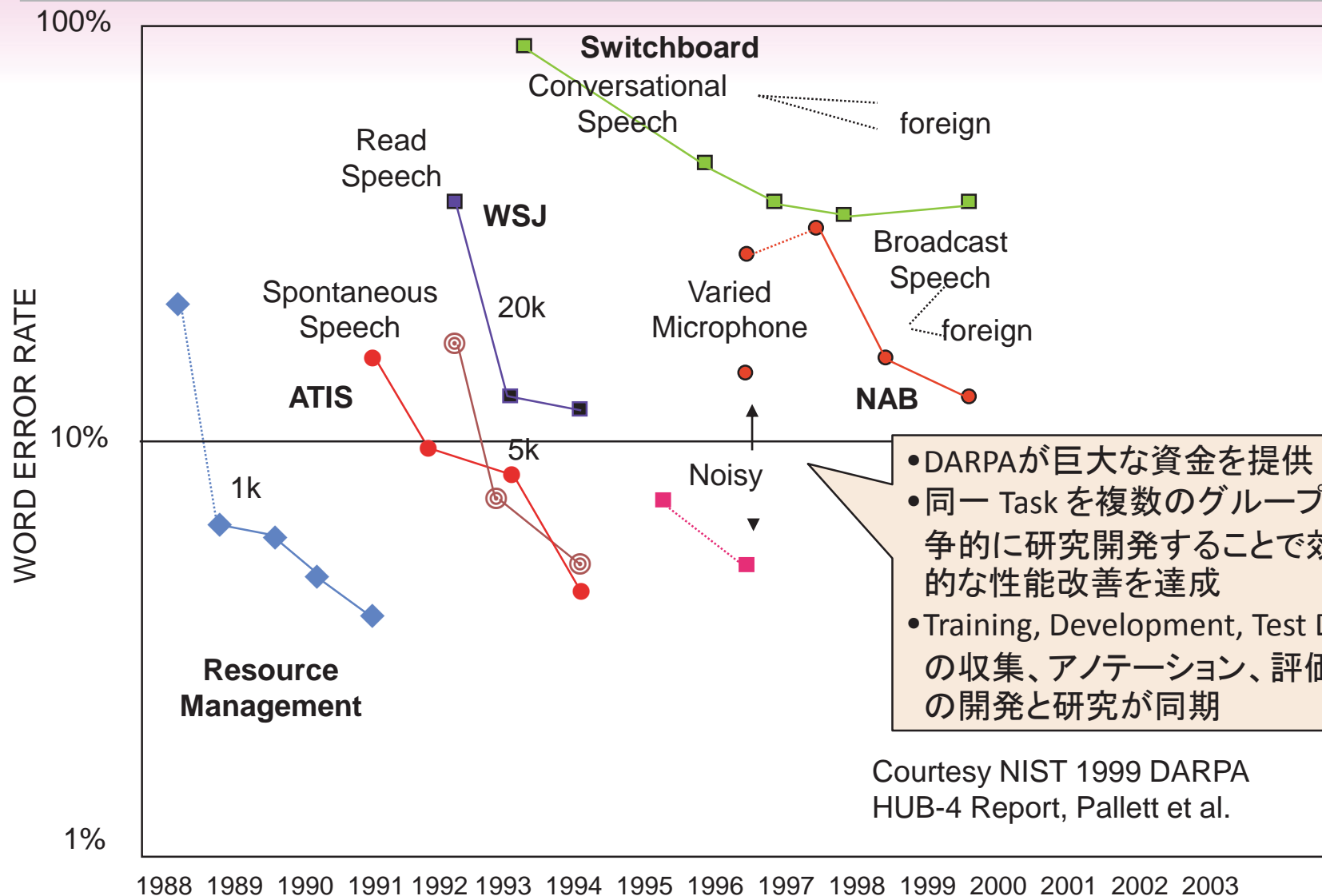
▶ 機械翻訳

- ルールベース：
言語学者、言語学の知識のある作業者がルールを注意して作成
- コーパスベース：
 - 用例ベース (Example-Based)
ルールを自動的にコーパスから抽出 [M.Nagao, J.Tsujii, 1985]
 - 統計ベース翻訳 (Statistical Machine Translation)
ルールが頻出するかの確率を学習。Noisy Channel Model [P.F.Brown, et.al., 1993]
(Jelinek「ルールライターが一人やめるたびに性能が上がる」)

音声翻訳を構成する研究に関する歴史



評価型プロジェクトとコーパス



分析

▶ 30年間にわたり研究を継続、音声翻訳サービスまで達成

– 目標

- 旅行会話音声翻訳という、明確なニーズがあり、音声認識、機械翻訳、データ収集が容易な目標を設定。ブレずに継続。

– 研究、予算

- 総務省からの研究資金が継続、当初は民間出資も。予算の使用が柔軟。
- 国際的マネージメント（東倉先生ベル研、鹿野先生CMU）。
- 1986から音声研究者と言語研究者で共同PJ。
- ATR視聴覚機構研（人間情報研）と協力。
- 膨大な資金を投入して音声、言語、対訳コーパスを構築（ATR+NICT）
- 第一線の研究所として認知された
 - HMM, NN, Expert System, Unification Grammar, Example-based MT, Phrase-based MT
- ATR→NICT内閣府社会還元PJ, グローバルコミュニケーションPJ, ITU-T標準化
- 2007年に最初の携帯電話用のネットワーク型の音声翻訳サービス、2010年にVoiceTra.

– 人材、体制

- NTTを始め多くの出資企業からの出向者やポスドク。キャリアアップ。
- どの企業よりも高い給料設定。海外からの研究者にも十分な年俸。
- 海外から多くの若手研究者を招聘、必要に応じて著名人を招聘(3割) 大量の学生インターン。
- 家具付き宿舎があり、海外から来ても即研究ができる。所内に外国人生活サービス。
- A.Waibel(CMU), A.Black(CMU), S.King(Edinburgh), M.Schuster(Google), Li Deng, K. Yao(Microsoft)
- 2007年 シャベって翻訳 (NTT ドコモ用) は、ATRの人脈

うまく行かなかったこと

▶ 同じ研究を継続することによるデメリット

－ 目標

- ATR時代は7年毎のプロジェクト再申請（会社の廃止、新設）が必要だった
→プロジェクトの目標を再設定、研究者もリセット。継承が困難に。

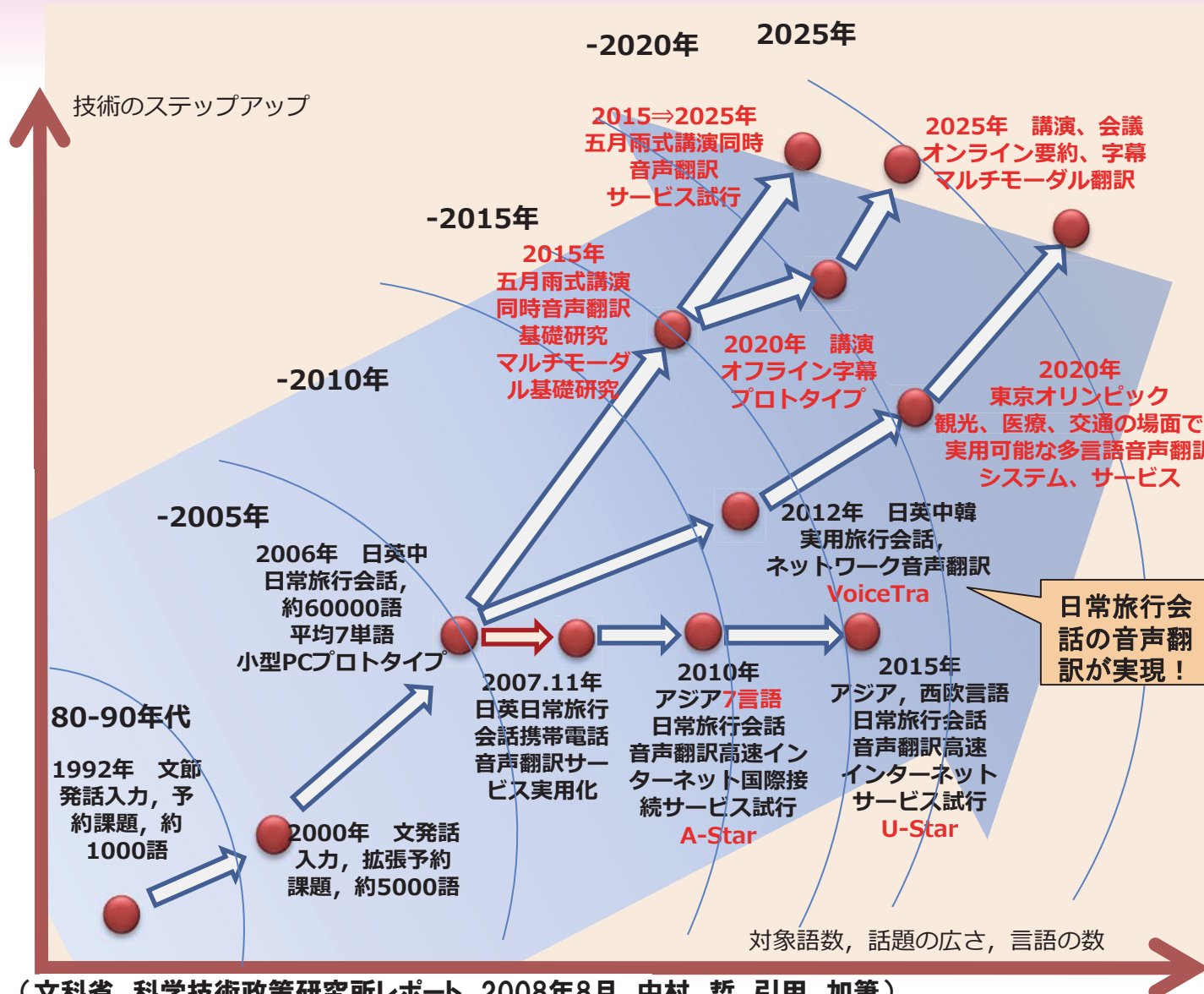
－ 研究、予算

- 2000年頃戻った際には予算は半減。基盤センター（投資会計）の行き詰まり。
- 良くも悪くも不連続。人も変わり継承できない。
→（2001年にプラットフォーム部隊を発足。研究者からソースをもらい、コード整備、テスト、ドキュメント作成を行う。→システム構築、技術移転対応部隊として機能）
- 旅行会話音声翻訳研究がガラパゴス化傾向。
- 個別研究が多く目標感が希薄に。研究は個別新規研究の追求。新規性か、性能か。
- 2007年以降、実証実験、システム構築、評価、デモが急増し、基礎研究時間が縮小。
- HMM, SMTのような理論体系的な強い研究がない。一般化され含されてしまう。
- 世界的な認知、社会還元に至らない。vs. Google, Microsoft

－ 人材、体制

- ATR → NICT移行。位置づけ、体制、運用、人事の大きな変化。安定化。
 - － NICTにより内閣府社会還元PJ, グローバルコミュニケーションPJ, 国際標準化を達成。
 - － NICT ASTRECでは出向者を受け入れることで、再度、復活を目指す。
- 雇用の自由度が低下し、人材の流動化が弱まる。

音声翻訳研究の展望 (2008⇒2015)



翻訳対象の種類

- ① コミュニケーション用翻訳
 - 音声翻訳、チャット翻訳、会議翻訳
- ② コンテンツの翻訳
 - ニュース、講演、コンテンツの翻訳
- ③ メディア間翻訳
 - テキスト、音声、画像、動画の翻訳

コンテンツ翻訳にも今後注力すべき

音声翻訳研究のポイント

- ① 短い発話⇒講演⇒会議
- ② 発話終了後翻訳⇒同時通訳
- ③ 音声翻訳 + 翻訳字幕

音声翻訳研究のポイント

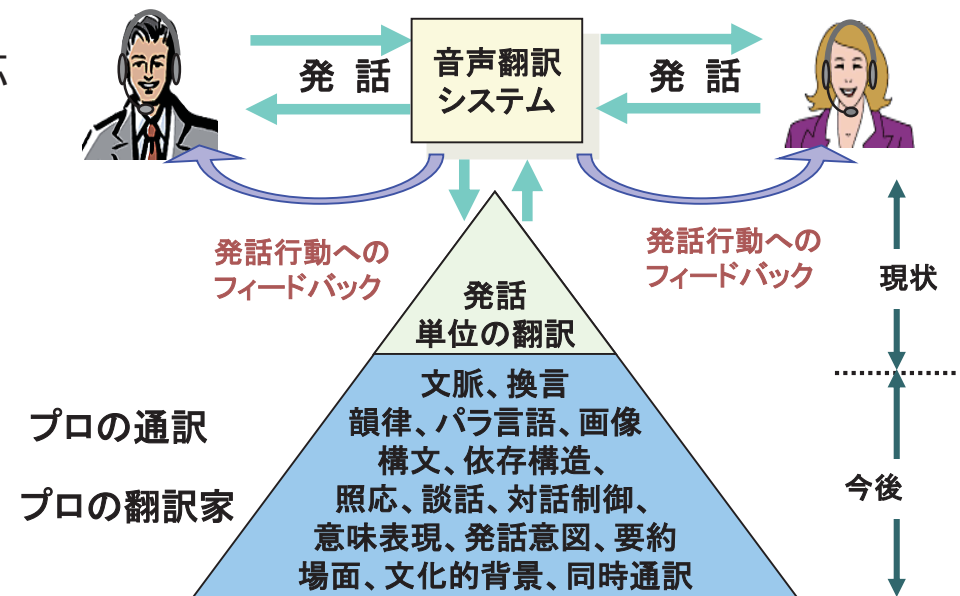
- ① 音声、テキスト、パラレルテキストコーパスの収集
- ② 実データの収集
- ③ 固有名詞、アクセント
- ④ 他のモダリティ、情報との統合

(文科省 科学技術政策研究所レポート 2008年8月 中村 哲 引用 加筆)

研究開発課題

- 高速化、高精度化、大語彙化、多言語化、未知言語対応
- 持続的なデータ自動収集と教師無し学習
- 全体最適化
- 方言・アクセント対応
- 雑音・残響対応、遠隔認識
- 韻律利用、パラ言語・非言語制御
- 構文・依存構造の利用
- 同時音声翻訳
- 主語省略への対応、照応への対応
- 対話構造・談話構造の考慮
- 意味表現とその利用
- エンティティリンク
- クロスカルチャー対応
- 対話システム

現在、利用可能な大規模データが存在しない。

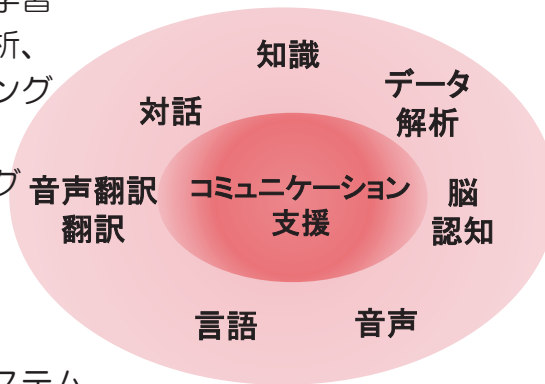


▶ 音声翻訳、機械翻訳、自然言語処理

- 同時音声翻訳、同時通訳の評価尺度
- 強調、個人性等の非言語情報を保持した音声翻訳
- 音声翻訳性能による音声認識最適化
- 構文構造を考慮した機械翻訳
- 中間言語を用いた多言語翻訳
- 構文解析器の自動学習
- 機械翻訳の誤り分析、
- アクティブラーニング
- 意味理解
- 自動プログラミング

▶ 音声対話

- 情報案内対話
- 説得、協調対話システム
- チャット対話システム
- マルチモーダル対話システム
- 対話を介した概念獲得、画像-言語の関係学習
- 対話における個人性モデリング
- 対話中の感情トリガーの抽出、応答生成
- 対話におけるエンタテインメント
- 言語横断QA, Non-factoid QA



▶ 認知

- 脳波信号分離、違和感リアルタイム検出
- 文の誤りの脳波による検出
- 眼球運動からの未知語推定
- 自閉症児の対話システムによるトレーニング
- 軽度認知症、統合失調症の自動検出

▶ 音声認識

- 超低資源言語用教師なしモデル学習
- 雑音、残響下音声認識
- 環境音、音声同時認識
- 感情音声の認識
- 非母語音声認識用発音辞書、音響モデル学習

▶ 音声合成

- テキスト音声合成の高品質化、声質変換
- 調音パラメータからの音声合成

▶ ビッグデータ解析(学内横断PJ)

- 構造理解に基づく大規模文献情報からの知識発見
- バイオ・マテリアルインフォマティクス
- 匿名化、ソーシャルメディアからの評判分析
- けいはんな学研都市スマートシティ

提言 1 スマートシティの例

(H27 戦略WG第3回NAIST中村資料等参照)

▶ 現状の課題

- グローバルビジネスで劣勢
- 政府委託研究予算も減少
- 民間での新規研究開発が減少
- 優秀なグローバル人材不足、人材の流動性不足

▶ 研究開発実施上の課題

- 研究人員 > 研究開発人員も課題
- 持続的イノベーション > 破壊的なイノベーション
- 起業も十分でない
- 全体の中で、どこが投資で、どこで収益を得るかの戦略 (Google, TEMASEC)

▶ 今後の社会的課題と発展のチャンス

- 社会課題：超少子化、超高齢化、社会インフラ老朽化、気候変動、都市集中、エネルギー源枯渇、世界人口増大
- 機会：
 - オリンピック、地方創成 (インバウンド観光)

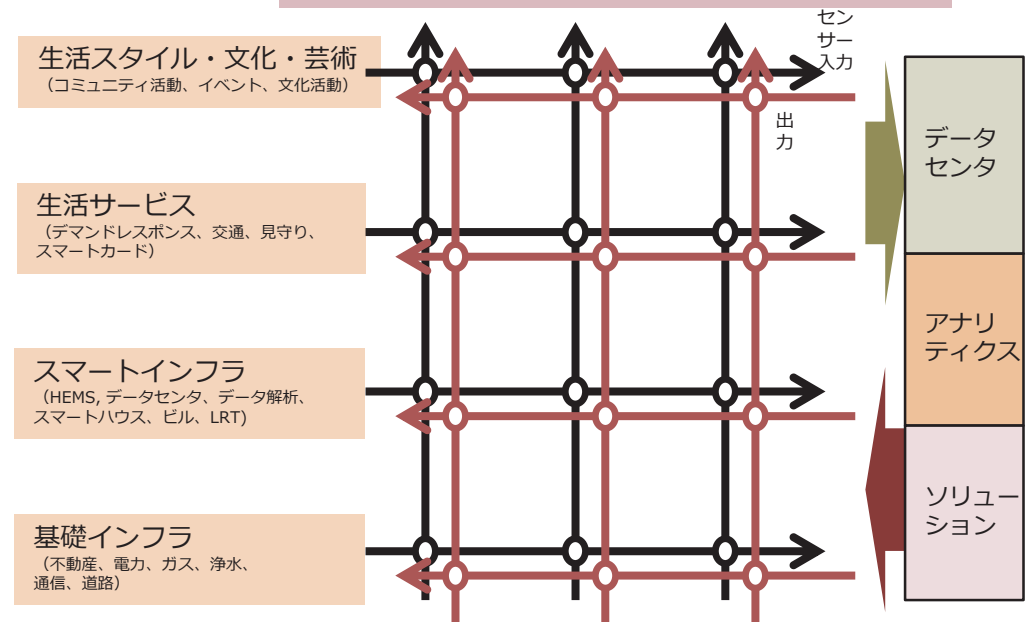
▶ Smart X

- 生産性向上 ⇒ グローバル社各企業が考える
- 生活向上 ⇒ 国民、納税者還元：政府で担当
- 新商品開発 ⇒ 両方を実現する可能性有り

▶ インテリジェントスマートシティ 知のネットワーキング

- **納税者が幸せになる、在日企業が幸せになる**
- これからのグローバルビジネスを創出
- 持続的、成長的に、人材と技術を生み出す
- 感動、知識、能力、安全、安心、コミュニティなどのサービスを要求に応じて、リアルタイムにシティ、国レベルで最適に提供する仕組み

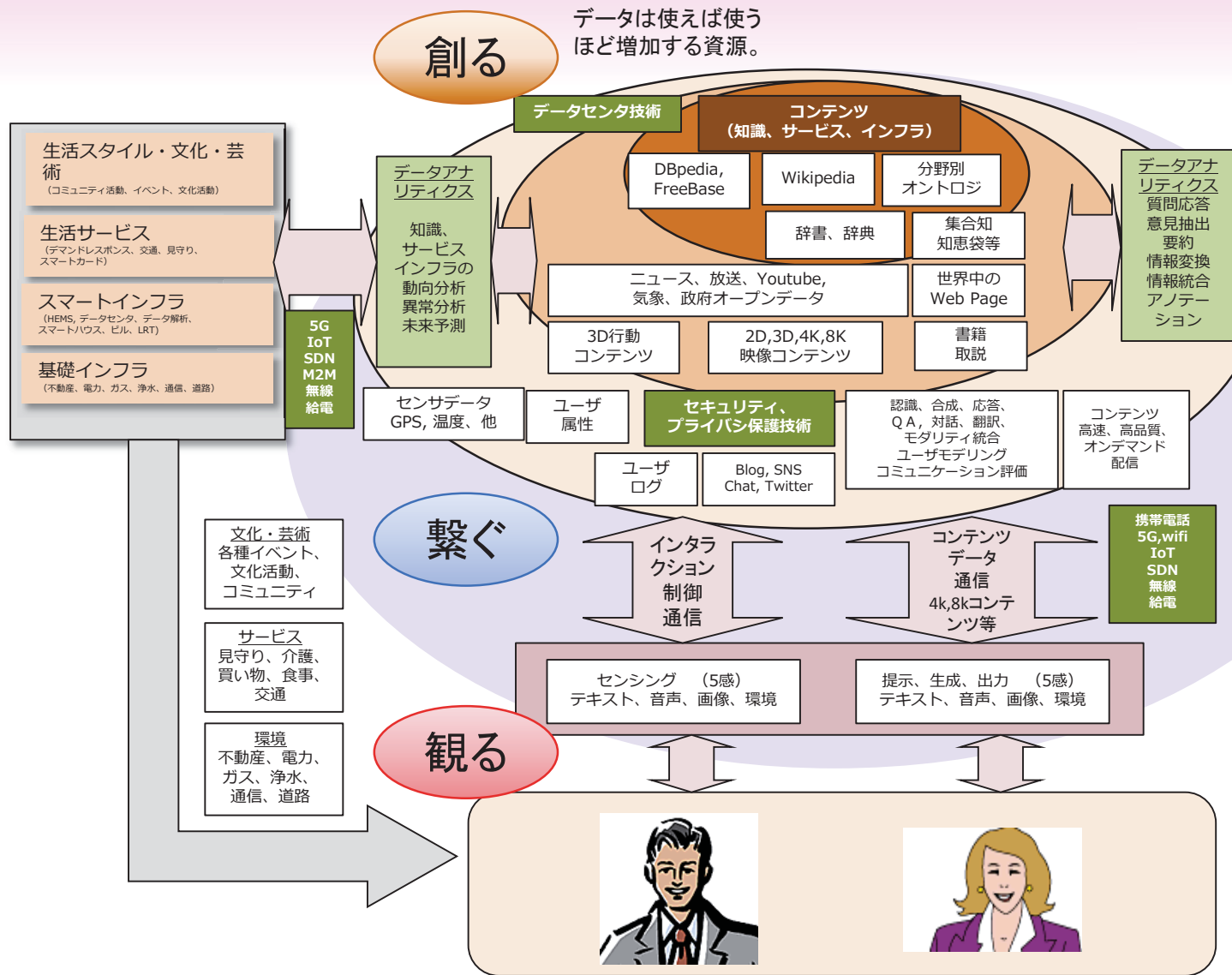
住民、スマートシティに関するグローバルマーケット



(H27 戦略WG第1回MRI中村様資料等参照)

コンテンツ、サービスを中心にした社会

(H27 戦略WG第3回NAIST中村資料等参照)



- 知のネットワークング技術とは**
- ▶ データアナリティクス
 - 多次元、多種類、多言語、大規模、超成長
 - 可視化、統合、関連づけ
 - フィルタ、同定、予測、変換、理解、生成
 - 推論、データマイニング、Non-factoid QA
 - ▶ データセンタ
 - 超並列分散設計
 - 超分散アルゴリズム
 - クローリング、フィルタリング
 - ▶ 言語解析
 - 言語解析、意味解析、メディア間の意味づけ、関連づけ
 - 日本のための日本語処理
 - グローバル化のための多言語処理と翻訳
 - ▶ インタラクション
 - テキスト、音声、画像、環境の認識、合成
 - 意図、意味を伝える、受け取る、理解する
 - QA、対話
 - マルチモーダル、脳
 - ▶ 感動、付加価値通信
 - 臨場感、立体、感動
 - ハイブリッドキャストによる付加価値通信
 - インタラクティブ通信
 - ▶ 動作仕様
 - リアルタイム、利用者に適応、
 - プライバシー保護
 - 低価格

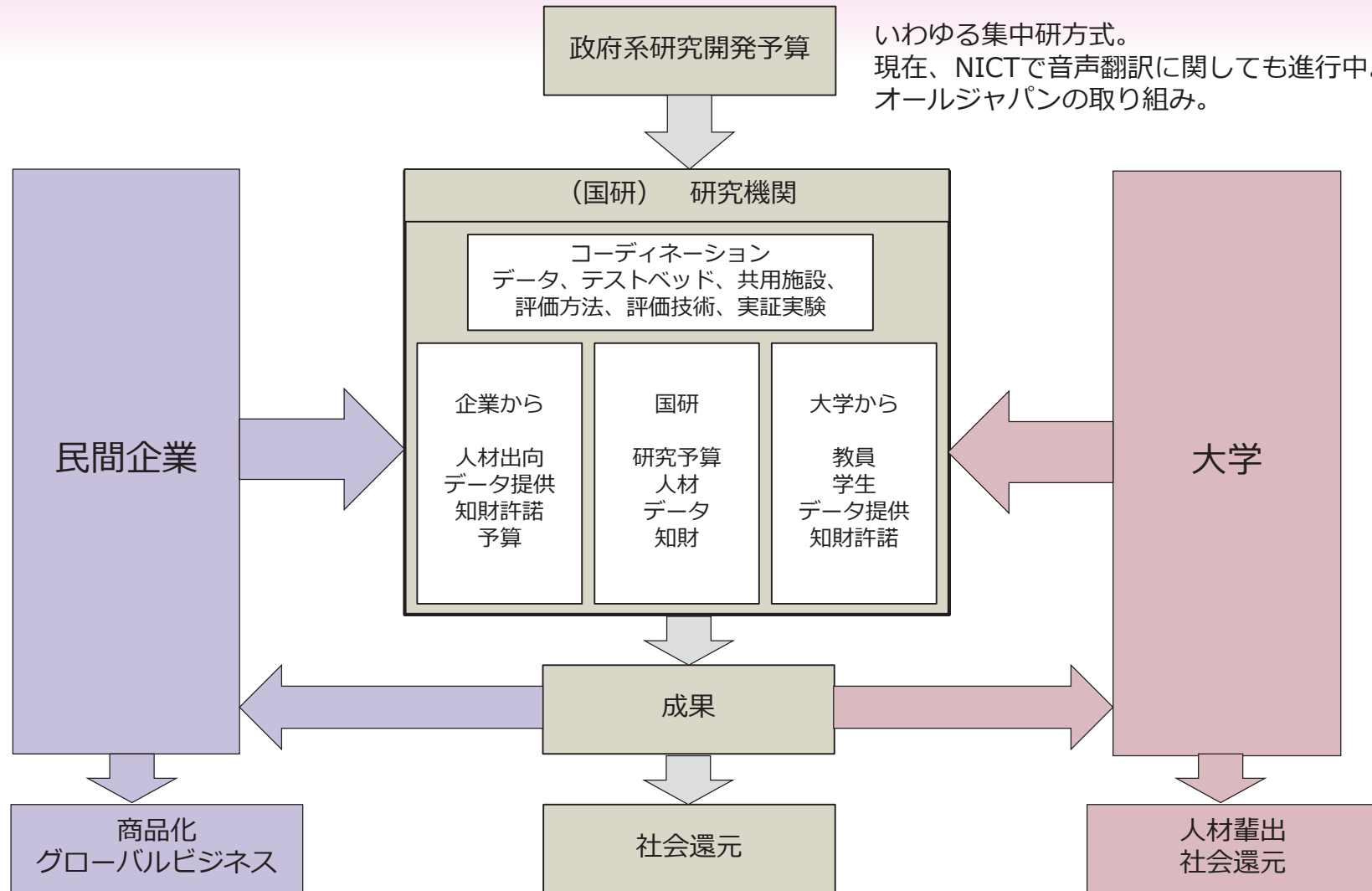
提言 2

▶ 研究のタイプ、予算、体制

- 基礎研究：いつアイデアが出るか不明。 新規性で評価。
 - 担当：大学
 - 予算：安定予算であるべき。基礎研究には5年でも短い。競争的資金も採択、不採択でなく比率認定して受給を増やす。
→現在、校費を減らして競争的資金への流れがあるのは問題。
 - 対象：主として研究者の person 費。試験的な設備やコーパス費。
- 基盤研究： 目的を設定し計画研究。競争的に性能で評価。
 - 担当：国立研究所、企業、大学がグループを作って行う。
 - 予算：省庁直轄委託、NICT, JST, NEDO および科研など。
 - 対象：プロジェクトにつける。特に、大規模データ構築費用、オープンプラットフォーム、社会実証実験費。
- 当該分野の研究者、技術者の質、数を、確保、増加させる必要があるため、博士学生を育成、最大限に活用。さらに、人材の流動性を確保。

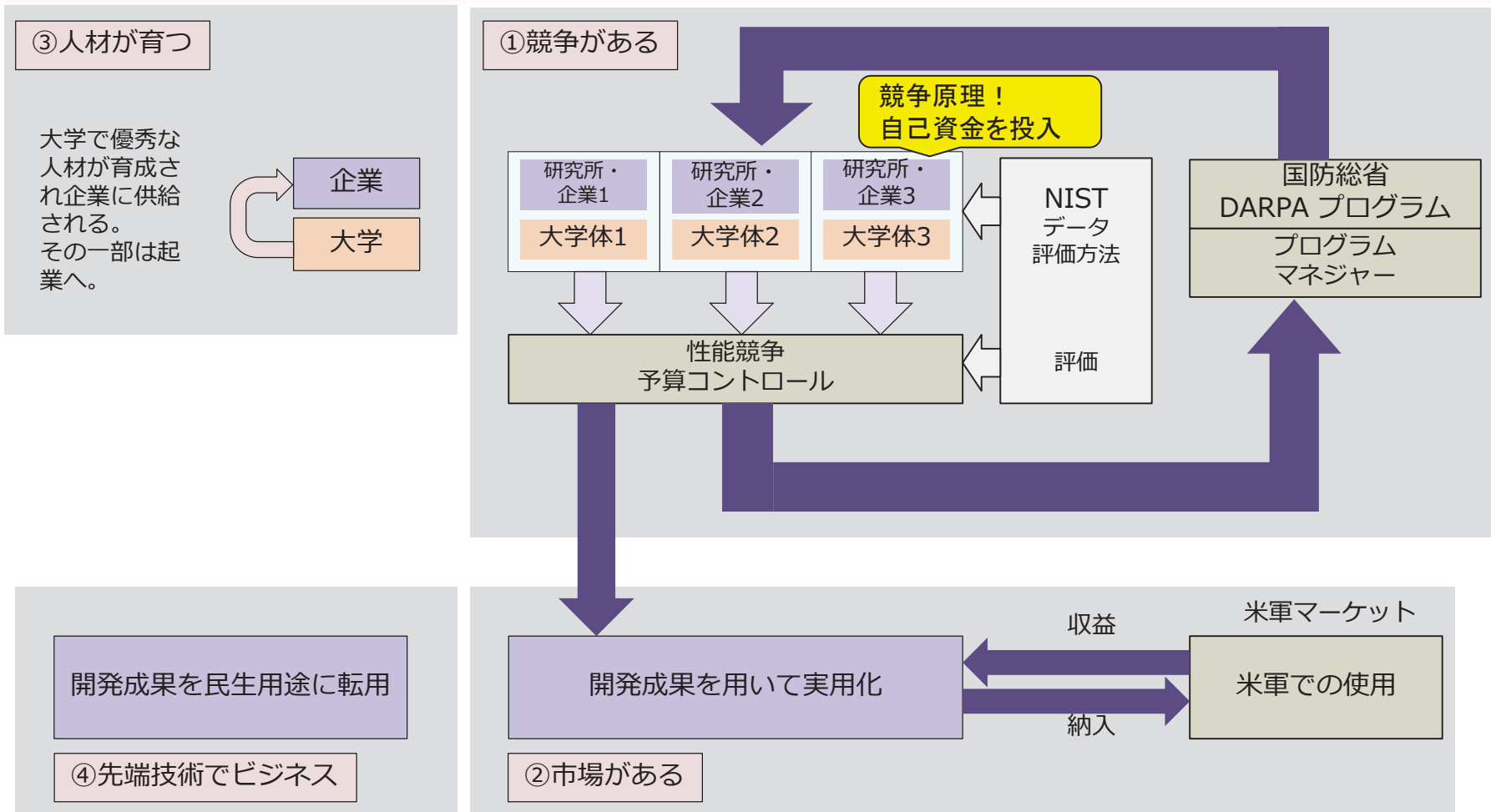
集中型モデル

(H27 戦略WG第3回NAIST中村資料等参照)



米国における技術開発、人材育成のモデル*

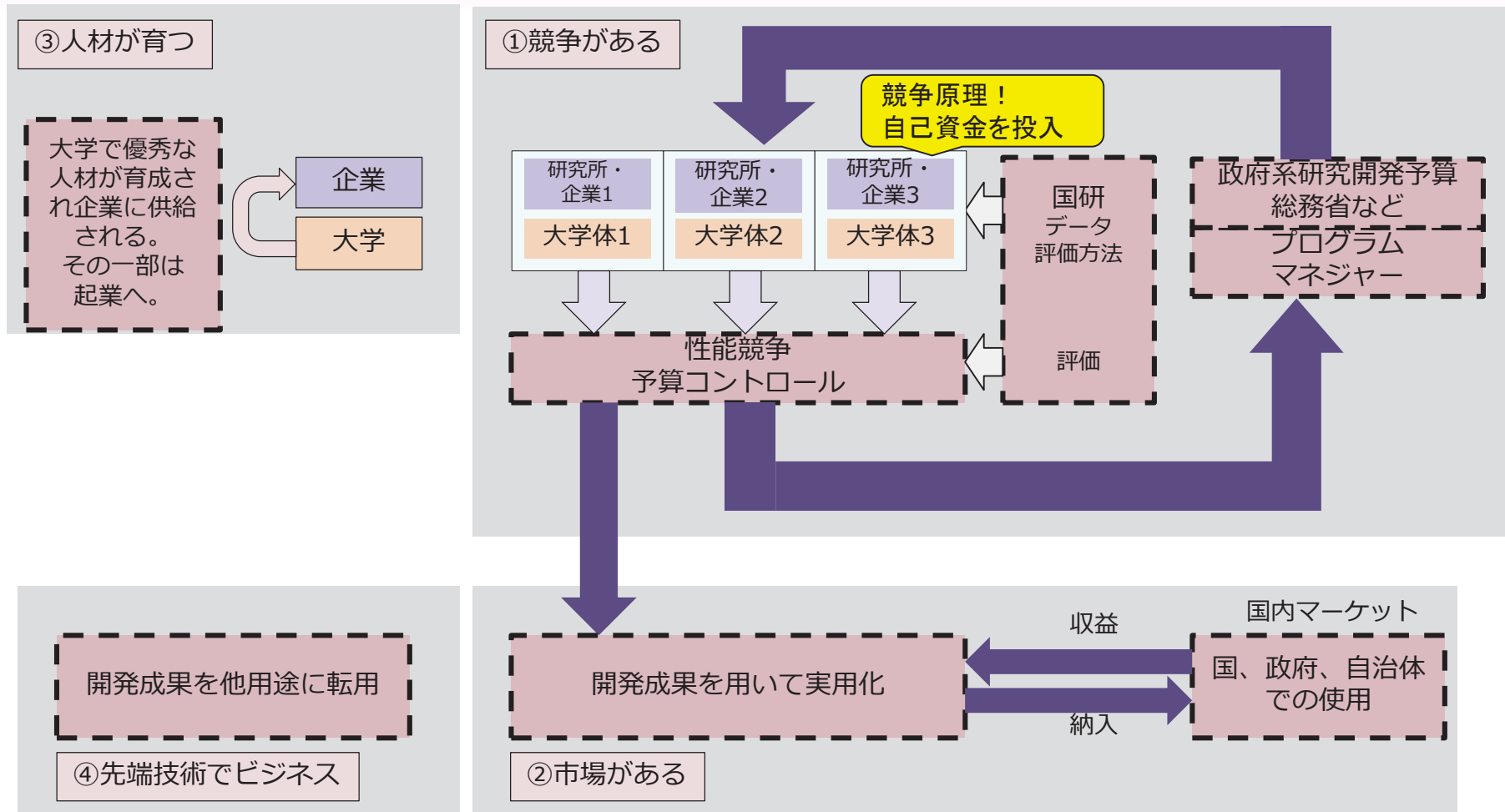
(H27 戦略WG第3回NAIST中村資料等参照)



* 著者の理解の範囲

日本でも可能なモデル

(H27 戦略WG第3回NAIST中村資料等参照)



提言 3

▶ ボード、アドバイザリ体制を作る

- 分野ごとに産官学の多様なメンバーからなるボードを長期的に設置し、中長期戦略を継続的に議論する（5～20年）。
- ボードは分野に責任を持ち評価を受ける。また、情報発信の責任も負う。ボードメンバーは国際的評価を考慮。
- 国際アドバイザリ委員会を設置し継続的に助言をもらう。
- 産官学コンソーシアムをつくる。定常的な議論、情報交換の場を創る。

▶ 予算を作る

- 分野ごとの目標、研究の質に応じてボードが提案。

▶ データやサービスを共有するための制度設計、知財管理

- プライバシー関連法律の設計
- データフォーマットの標準化
- 特許プールなど知財の合理的な活用

▶ 大規模データを利用可能にする

- 政府データのオープン化
- 全ての省庁、JST, JSPS, NICT, NEDO委託研究のデータのオープン化