


学生研究奨励賞応募研究
「テキストマイニングを用いた
行政機関内の活動の可視化」

2016/10/28

筑波大学大学院 ビジネス科学研究科
狩野英司

目次

1. はじめに p.3
2. 先行研究および使用ツール p.10
3. 活動の特徴を示すキーワードの抽出 p.12
4. 組織内活動の関係性の可視化 p.20
5. まとめ p.29



1. はじめに

1-1. 背景と目的

- 背景：
 - 近年、政府や自治体では「オープンデータ」の取組みが活発に進められているが、データを真に役立たせるためには、各機関が「データマネジメント」（以下、DM）に取り組むことが必要
 - そのためには、DMに関する活動事例が発掘・共有されていく必要があるが、DMそのものを対象とした行政機関向けのリサーチは存在せず、手掛かりを掴むことも難しい
- 目的：
 - 公開情報にテキストマイニングを用いることで、組織内の活動の関係性の特徴、具体的には行政機関内でDMの活動が行われているかどうかを可視化する
 - これにより、目的とする組織内活動の特徴を効率的に探索し、組織の活動状況を定量的に評価することを可能とする

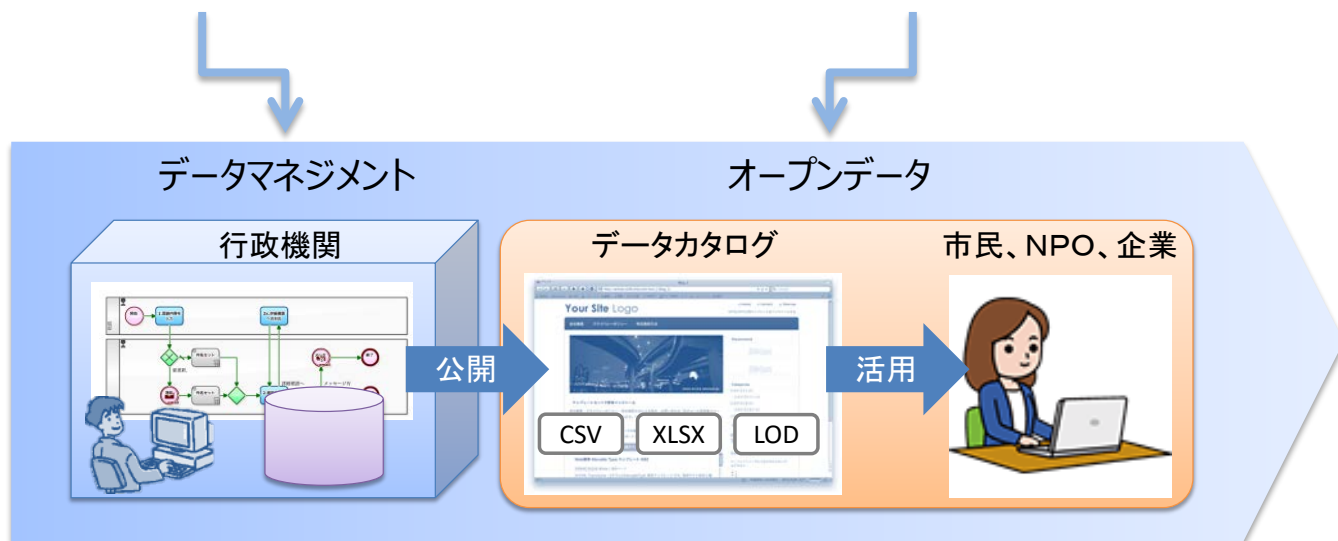
1-2. 用語解説

■ データマネジメント(DM)とは：

- データを利活用可能な状態に維持し、継続的に改善していく活動
- 具体的には、データの企画、設計、運用、利活用などデータのライフサイクルにわたりマネジメント活動を実践すること
- ビッグデータ、AI、IoT等のデータ利用の進展に伴い、関係団体※を中心に研究や普及活動が活発に進められている

■ オープンデータとは：

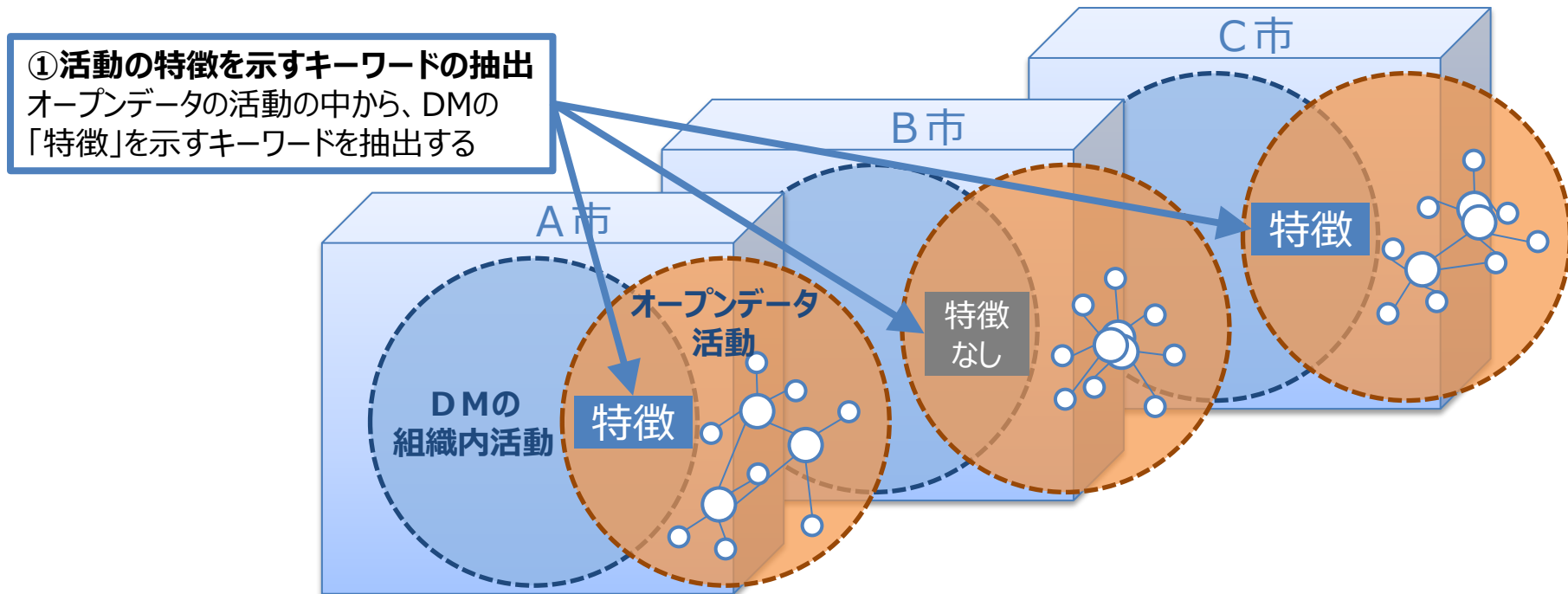
- 行政機関が保有するデータを再利用可能な形で公開する取組み
- 総務省の定義：「機械判読に適したデータ形式で、二次利用が可能な利用ルールで公開されたデータ」であり「人手を多くかけずにデータの二次利用を可能とするもの」
- 諸外国をはじめ我が国政府・自治体でも近年積極的に取組みが進められている



※一般社団法人 日本データマネジメント・コンソーシアムなど

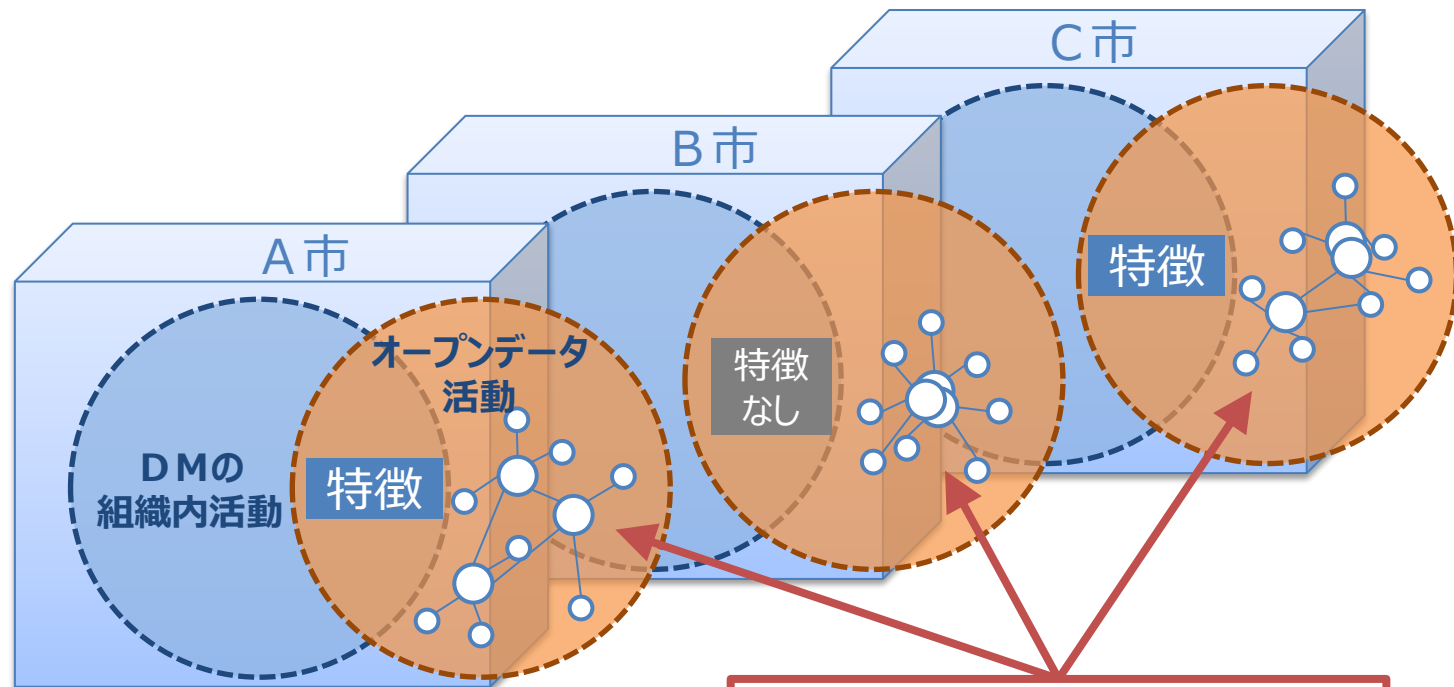
1-3. アプローチの概要(1/3)

- ① 活動の特徴を示すキーワードの抽出
- ② 組織内活動の関係性の可視化
- ③ 可視化結果の妥当性の検証



1-3. アプローチの概要(2/3)

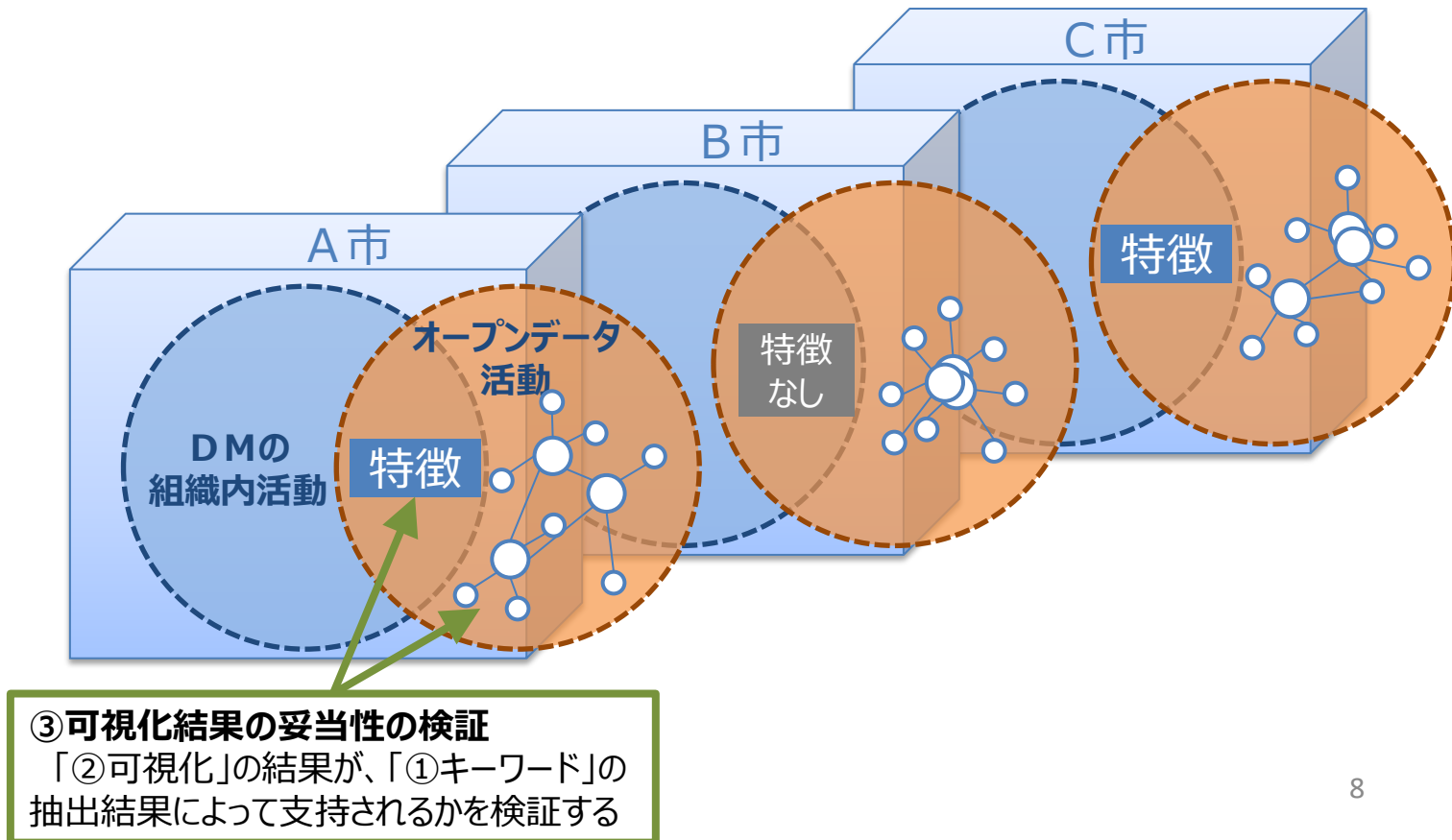
- ① 活動の特徴を示すキーワードの抽出
- ② 組織内活動の関係性の可視化
- ③ 可視化結果の妥当性の検証



②組織内活動の関係性の可視化
オープンデータに関する活動を「多軸的」に行っている組織を特定する

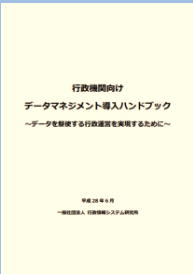

1-3. アプローチの概要(3/3)

- ① 活動の特徴を示すキーワードの抽出
- ② 組織内活動の関係性の可視化
- ③ 可視化結果の妥当性の検証




1-4. 使用データ

以下のドキュメントのテキストデータを使用

DMの活動を記述した文書	オープンデータの活動を記述した文書
<p data-bbox="131 468 962 564">「行政機関向けデータマネジメント導入ハンドブック」(一般社団法人行政情報システム研究所)</p> <div data-bbox="575 578 975 935"></div> <ul data-bbox="131 978 962 1320" style="list-style-type: none">• 汎用的なフレームワーク※をベースに、行政機関向けにアレンジしたデータマネジメントの望ましい取組みを提示• 本研究では具体的な活動を示した第4章および第5章のみを使用(11,500字)• 選定理由: 行政機関におけるデータマネジメントの活動を説明した唯一の資料	<p data-bbox="1006 468 1773 564">「e-Kansaiレポート2016」(一般財団法人関西情報センター)</p> <div data-bbox="994 578 1394 935"></div> <ul data-bbox="1006 978 1837 1320" style="list-style-type: none">• オープンデータに関する取組みについて、様々な組織・団体をインタビューして分析• 本研究では、関係性が薄いと見られた事例を除外した後の残り8自治体の事例を使用(各事例とも4,000字前後)• 選定理由: 行政内部の活動まで踏み込んで、具体的に説明している

※「データマネジメント概説書(JDMC版)」および「DAMA-DMBOK データ管理知識体系ガイド」をさす




2. 先行研究および 使用ツール

2. 先行研究および使用ツール

- テキストマイニングを用いた組織分析には、以下のような先行研究があるが、組織内の活動そのものにフォーカスして、特徴の可視化を試みたものは見当たらない
 - テキストマイニングを用いた組織分析の主な先行研究

文献	研究概要
新藤晴臣, 起業家のパースペクティブによる戦略形成へ影響関する考察, 2014	起業家が持つ戦略的パースペクティブと、それが企業戦略にどのような形で反映されるかを分析
疋田眞也・萩原克幸・鶴岡信治, 組織研究におけるテキストマイニングを用いた系統的分析法, 2012	経営組織で用いられる言葉の質的な違いを分類し、意味分析につながる系統的な方法を研究
斎藤進也・稲葉光行, テキストマイニングによる非営利ネットワークの組織化プロセスの可視化, 2010	非営利ネットワークの組織化プロセスの可視化や自発的に形成された組織の動態把握の手法を研究
上野山達哉, 「組織らしさ」のテキストマイニング, 2007	組織の独自性や価値規範にかんする個人の知覚・認知プロセスを分析
喜田昌樹, アサヒの組織革新の認知的研究-有価証券報告書のテキストマイニング, 2006	企業の組織革新と組織内での概念に関する認知の変化を分析

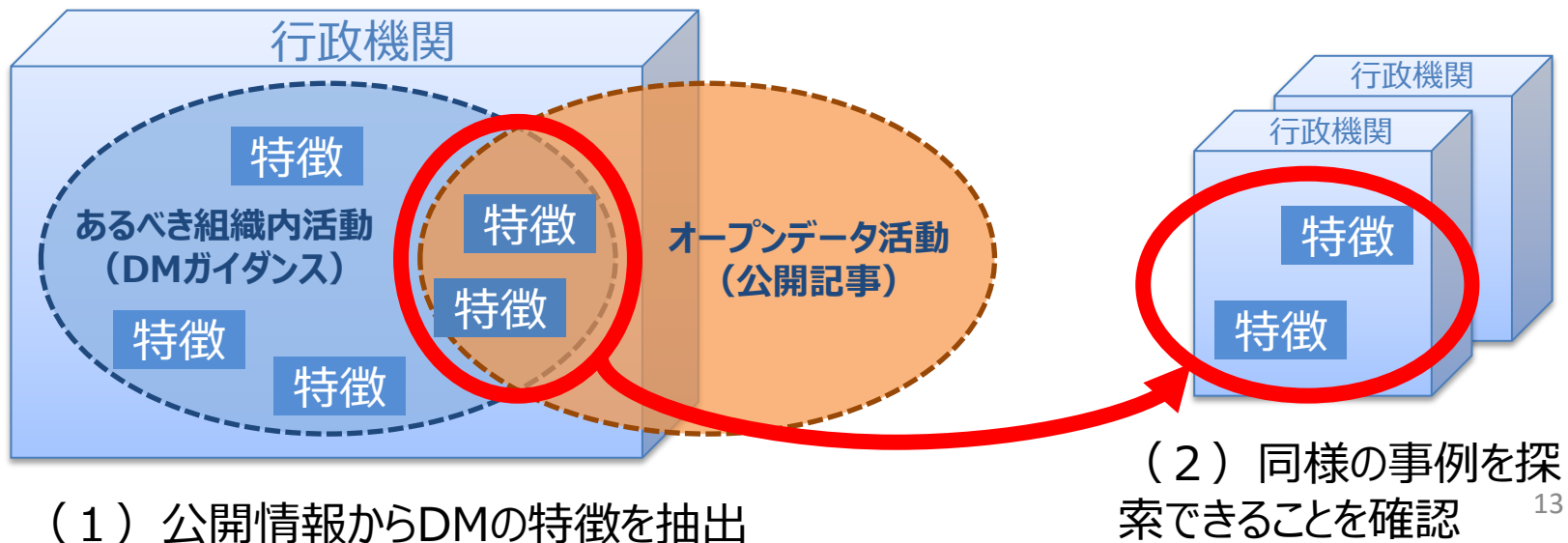
- テキストマイニングのツールとして以下を使用
 - Text Mining Studio 5.2.1
(選定理由) 形態素解析から共起ネットワークのグラフ作成まで一連の処理を直感的な操作でシームレスに実施可能であり、詳細条件の設定・変更も容易



3. 活動の特徴を示す キーワードの抽出

3-1. 実施概要

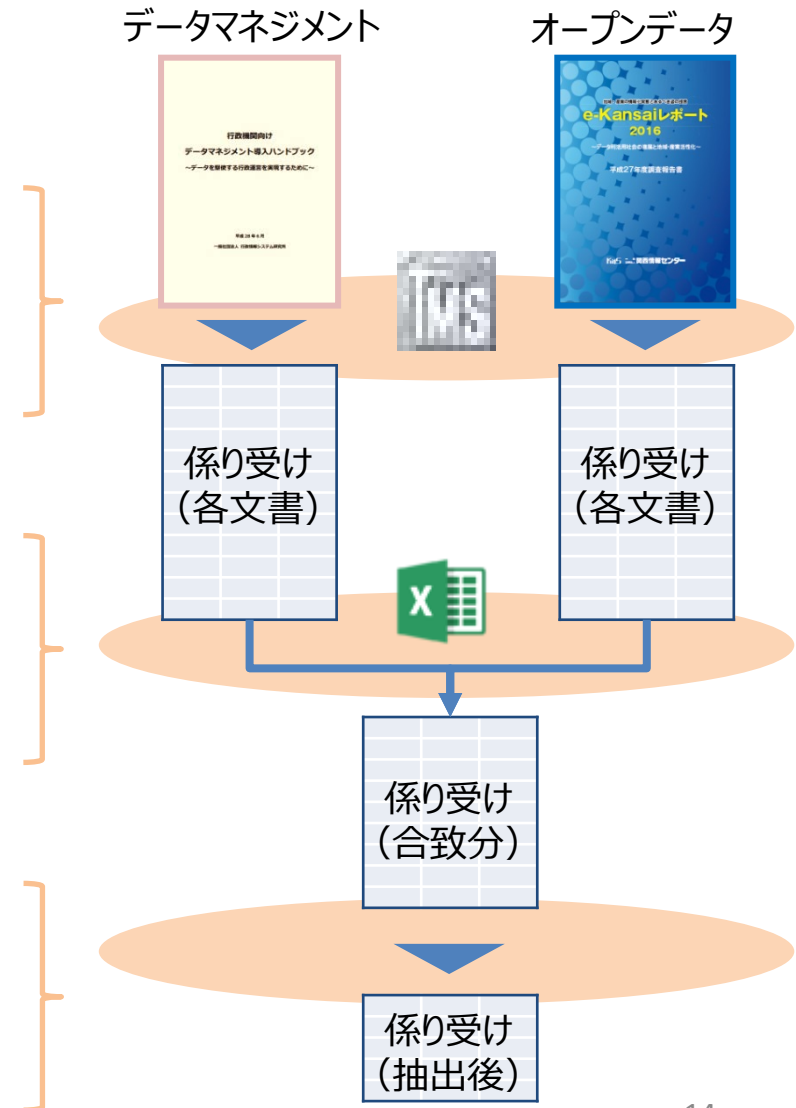
- (A) キーワードの抽出： DMの活動を解説したガイダンスと、「オープンデータ」の活動を紹介した公開記事を突き合わせることで、DMの活動の特徴を示し、かつ、公開記事に含まれているキーワード（係り受け）を抽出する
- (B) キーワードの検証： 上記により抽出したキーワードで同様の事例をWeb検索し、有効性を確認する



3-2. (A)キーワードの抽出 (1/4)

【実施の流れ】

- ① 各文書を分かち書きし、係り受けを抽出する
- ② 両文書の係り受けを突き合わせる
- ③ 目的に合わない係り受けを除外する



3-2. (A)キーワードの抽出 (2/4)

①Text Mining Studioを用いて両文書を分かち書きし、係り受けを抽出

データマネジメント



係り受け
取り組む 想定リスク
データ品質
受託者 明示
要件 明確化
留意 必要
ある+べきだ 姿
データ 意味

オープンデータ



係り受け : A市
オープンデータ活用
見る+できる 化
資金 調達
クラウドファンディング 調達
ICTプラットフォーム
おばあちゃん 編み物会社



係り受け : B市
オープンデータ 推進
開
データ
治体 連携
取り組む
影響



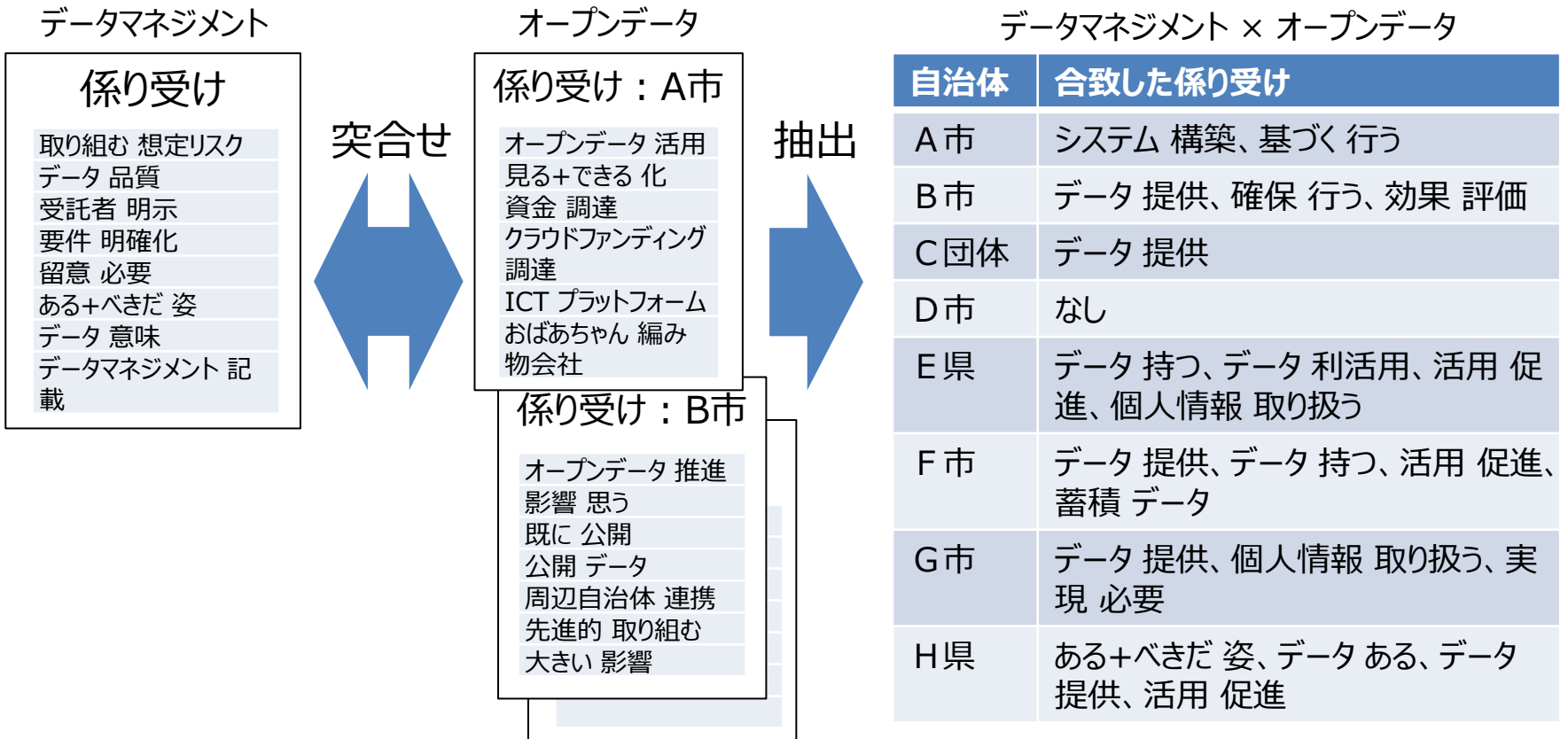
.....

係り受け
頻度解析※

※TMStudioの係り受け頻度解析を係り元品詞設定で「名詞 数」を除外して実行

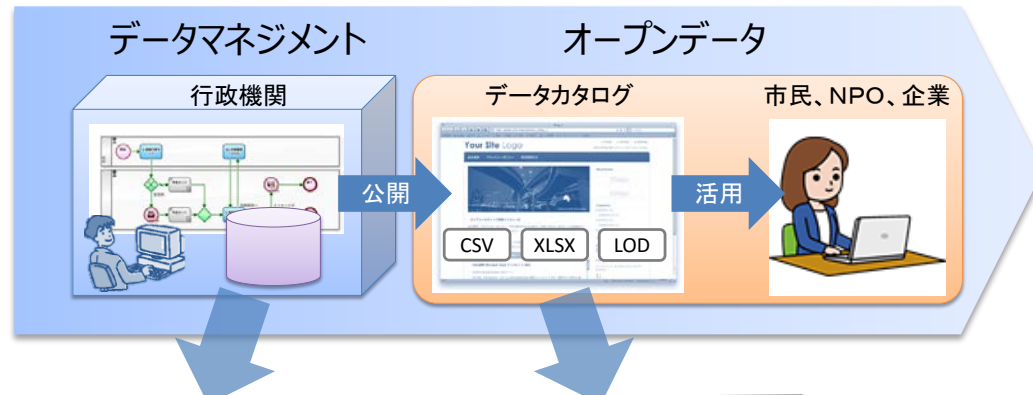
3-2. (A)キーワードの抽出 (3/4)

②両文書の係り受けを突き合わせ、合致した係り受けを抽出



3-2. (A)キーワードの抽出 (4/4)

③目的に合わない係り受けを除外し、キーワードを絞込み



	データ生成・管理 段階の活動	データ利活用 段階の活動	
能動的 活動	システム 構築 効果 評価	データ利活用 活用 促進	抽出された キーワード
受動的 活動	データ持つ データある 個人情報 取り扱う 蓄積 データ	データ提供	除外
一般的 行動	基づく 行う 実現 必要	確保 行う ある+べきだ 姿	

3-2. (B)キーワードの検証 (1/2)

- ①で得た係り受けのDMの活動との関係性を検証するため、キーワードとしてWeb検索(Google)を実施
- 検索の結果、DMに特徴的と思われる複数の活動が発見された

<検索式>

抽出キーワード

- オープンデータ AND 市 AND システム AND 構築
- オープンデータ AND 市 AND 効果 AND 評価

<検索結果※>

システム+構築	a市	b市	c市	d市	e県	f市	g市	h市	i市	j市
	-	○	-	-	-	○	-	○	-	-
効果+評価	k市	l県	m市	n市	o市	p市	q市	r市	s市	t市
	-	○	○	○	○	-	-	-	○	-
(参考) キーワードなし	u市	v市	w市	x市	y市	z市	a市	β市	γ市	Δ市
	-	-	-	-	-	-	-	-	-	-

DMに特徴的と思われる活動を発見 (○部分)

上記のキーワードなしで検索しても結果は得られず


※検索結果：自治体が情報ソースとなっている上位10件を抽出
(凡例) ○：データマネジメントに関する活動が行われている可能性が高い
-：データマネジメントに関する活動が行われているか不明

3-2. (B)キーワードの検証 (2/2)

- 一例として、検索結果には、上記キーワードを使わなければ発見できなかった以下のようなユニークな事例が発見され、その有効性を確認できた
 - DMの存在を示唆するユニークな事例

自治体・取組み	DMの存在を示唆する活動内容
浦安市：「浦安市におけるGIS及びオープンデータの取り組みについて」	GISのオープンデータ化と利活用を実現するため、「人材育成」「業務とGISの連携」などを実施
塩尻市：「地域オープンデータプラットフォーム、及び分析結果のデータの環流の仕組み作りと土砂災害防止への応用」	最先端の研究データや研究成果のオープンデータ化、さらに新たなオープンデータを循環的に生成するスキームを策定
草津市：「草津市のオープンデータのあり方に関する調査研究報告書」	手持ちのデータを公開するにとどまらず、不足しているデータ項目等を市民と協働で豊かにしていくこと等を提言

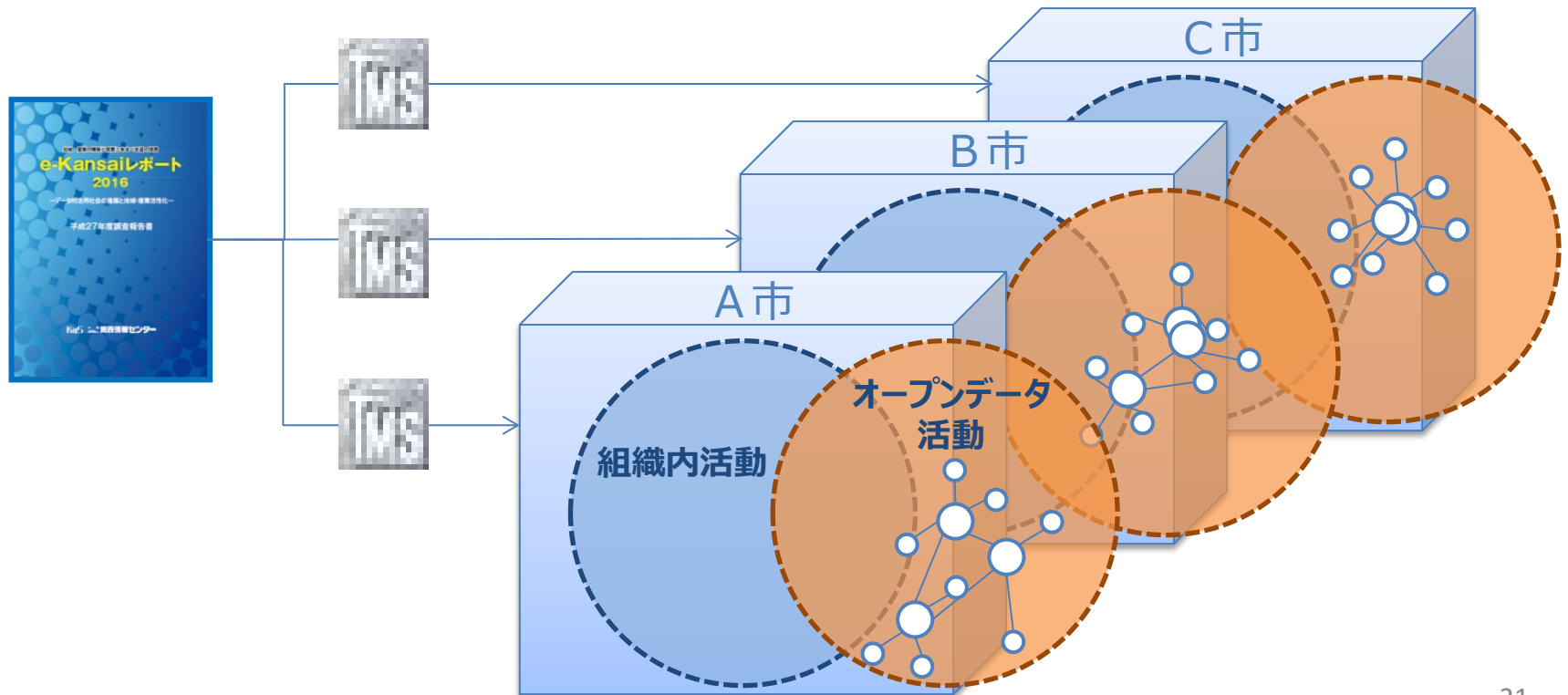
- 政府が発行した「オープンデータをはじめよう～地方公共団体のための最初の手引書～」の事例紹介の中でも、上記事例は発掘されていないものであった。



4 . 組織内活動の関係性の可視化

4-1. 実施概要

- テキストマイニングを用いて、組織内の活動間の関係性から、活動全体の特徴を抽出し、DMの活動が存在すると思われる行政機関を推定する

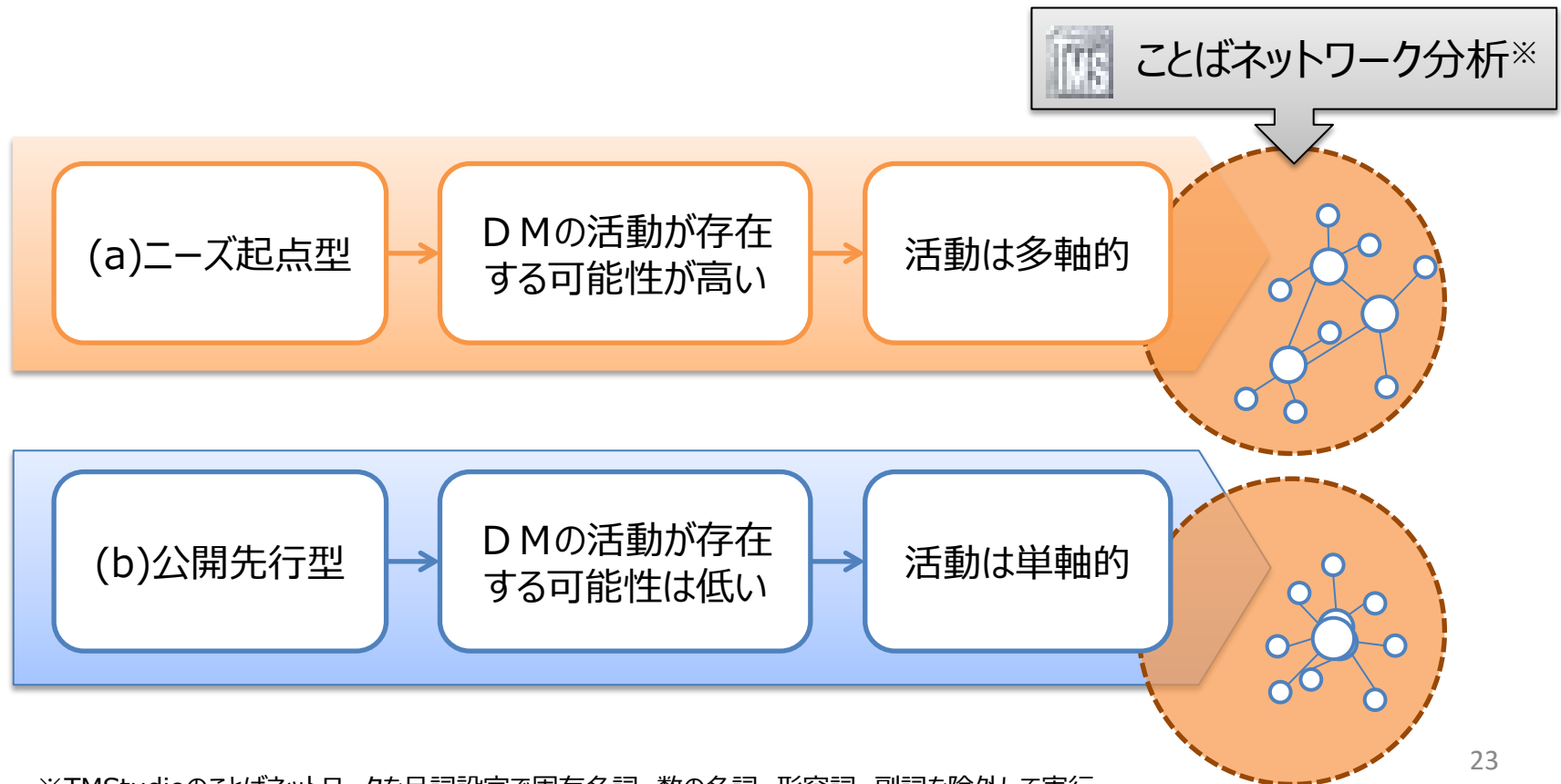


4-2. 活動の関係性の可視化(1/6)

- オープンデータのアプローチには、以下の2通りが存在すると考えられる
 - (a) **ニーズ起点型**：ニーズに基づき行政が主体的にデータの生成や利活用を考える。この場合、様々な取組みが複合して行われるため、**活動は複数の軸に分かれる（多軸的）** 傾向があると想定される
 - (b) **公開先行型**：今手持ちのデータの公開を優先し、ニーズや用途は、企業、NPO等の利用者に委ねる。この場合、オープンデータの公開という**単一の軸に活動が収斂する（単軸的）** 傾向があると想定される
- **DMの活動が存在する可能性が高いのは(a)の場合**

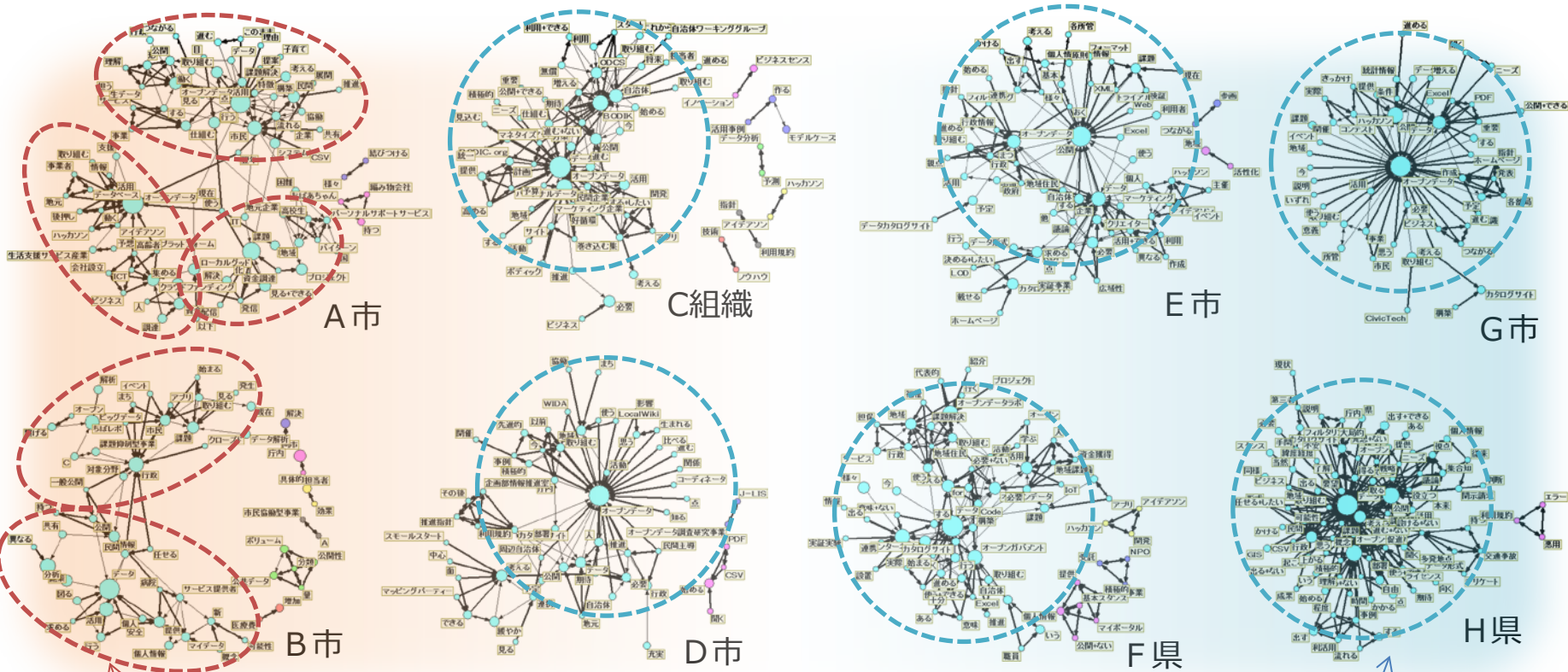
4-2. 活動の関係性の可視化(2/6)

- Text Mining Studioの「ことばネットワーク分析」機能を活用して、組織内の活動の関係性の違いを可視化、その機関が(a)/(b)どちらのタイプであるかを識別する



4-2. 活動の関係性の可視化(3/6)

- 分析結果： (a)活動が多軸的なA市・B市と、(b)活動が単軸的なその他の自治体の違いが可視化された

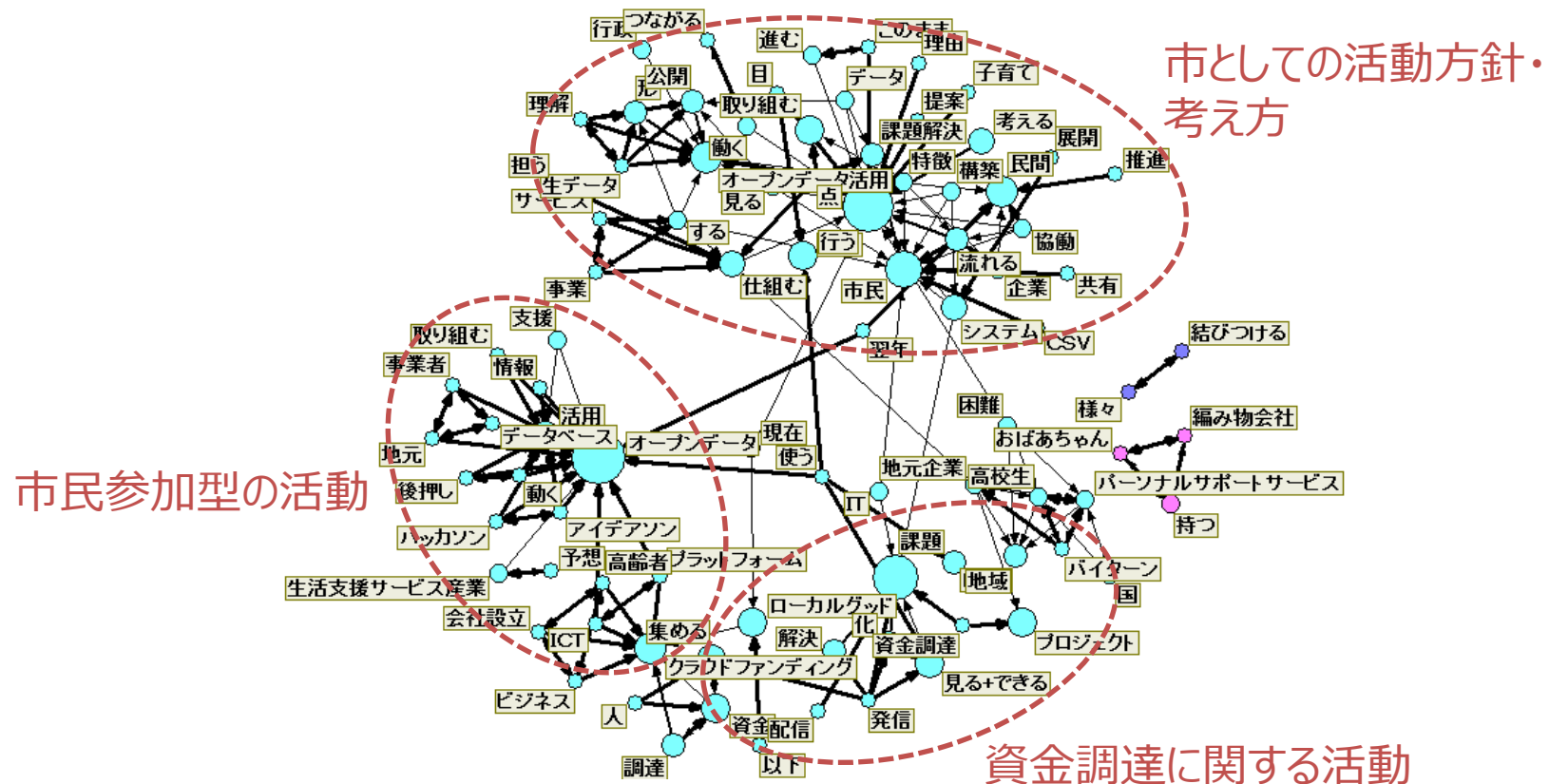


活動が多軸的に行われる
傾向→(a)

活動が一つの軸に
収斂する傾向→(b)

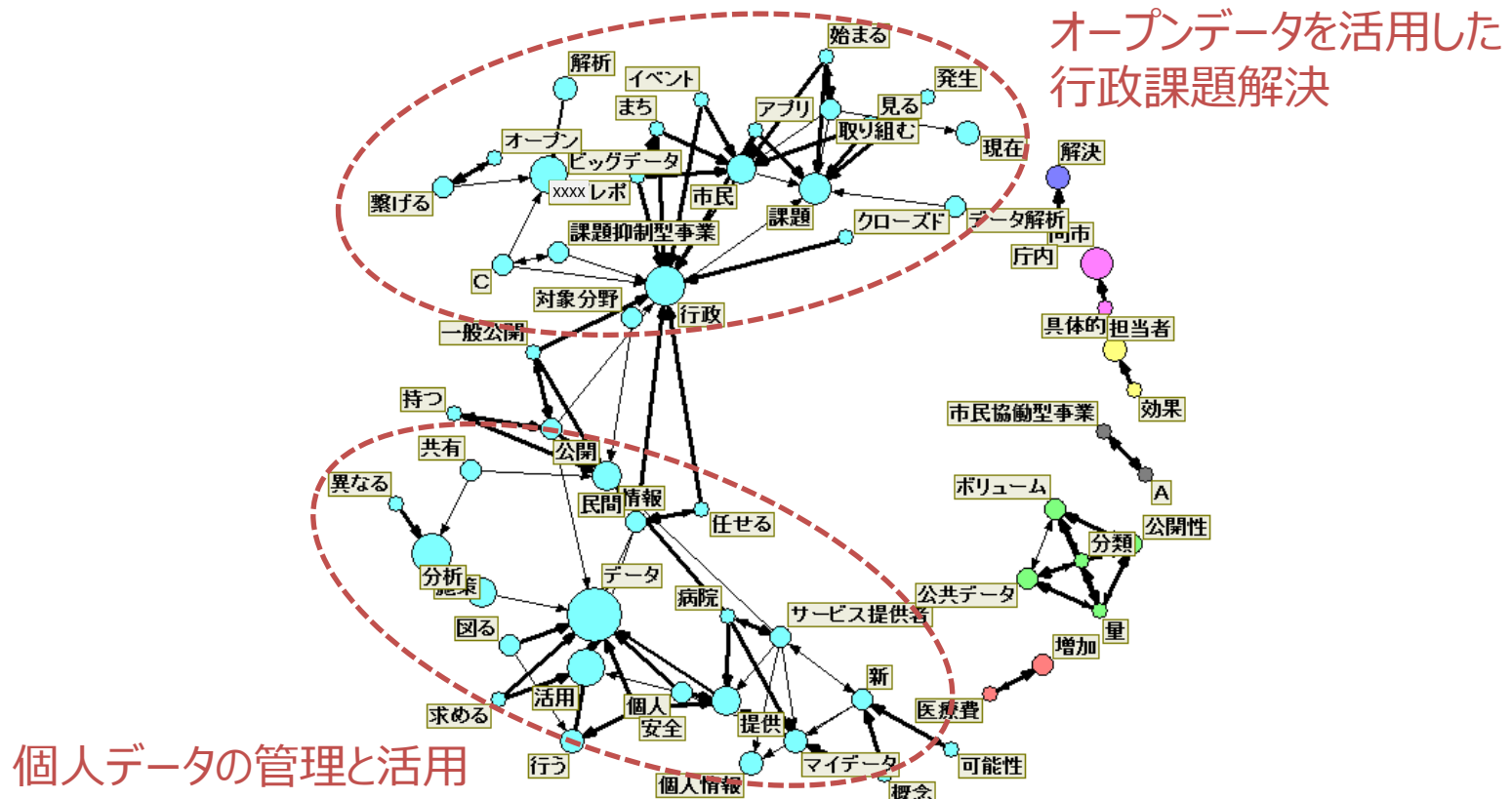
4-2. 活動の関係性の可視化(4/6)

- 【A市の例】 3つの軸で活動全体が構成されている
→ [(a)ニーズ起点型]-[DM活動が存在]-[多軸的]の可能性が高い



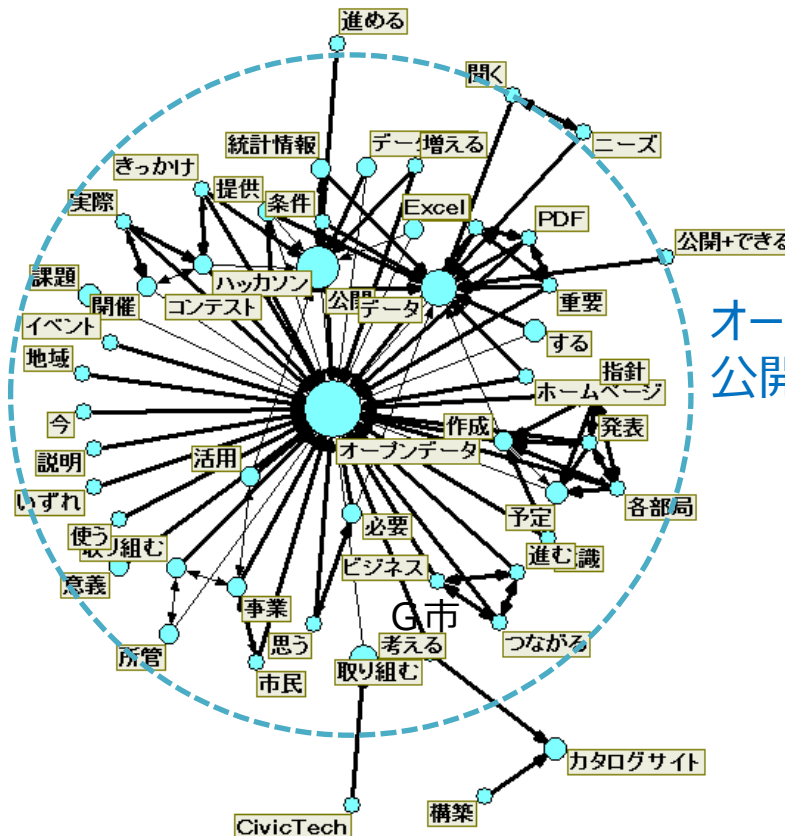
4-2. 活動の関係性の可視化(5/6)

- 【B市の例】 2つの軸で活動全体が構成されている
→ [(a)ニーズ起点型]-[DM活動が存在]-[多軸的]の可能性が高い



4-2. 活動の関係性の可視化(6/6)

- 【G市の例】 1つの軸に活動が収斂している
→[(b)公開先行型]-[DM活動は不在]-[単軸的]の可能性が高い



オープンデータの
公開・イベント・活用、、、

4-3. 可視化結果の妥当性の検証

- 前章「活動の特徴を示すキーワードの抽出」と本章「組織内活動の関係性の可視化」の分析より、DMの活動の存在が推定された自治体はどちらも同じ2市※であった

※これら2市（A市とB市）は、オープンデータに積極的な自治体として知られており、単なる保有データの公開にとどまらない取組みを行っている

	A市	B市	C組織	D市	E県	F市	G市	H県
DMの特徴キーワードの有/無 (第3章の分析)	あり [システム+構築]	あり [効果+評価]	なし	なし	なし	なし	なし	なし
組織活動の軸が複数/単数 (第4章の分析)	複数 (3つ)	複数 (2つ)	単数 (1つ)	単数 (1つ)	単数 (1つ)	単数 (1つ)	単数 (1つ)	単数 (1つ)

どちらの分析でも同じ結果が得られた

- 以上から、公開情報にテキストマイニングを適用することで、行政機関内の活動の関係性を可視化し、DMの活動が行われているかどうかを推定できることが確認できた



5. まとめ

5-1. 小括

- 本研究では、以下の手順により、テキストマイニングを用いて行政機関内でのデータマネジメント（DM）の活動の可視化を試みた
 - ① DMの活動の特徴を示しており、かつ、公開情報に含まれているキーワード（係り受け）を抽出。さらに、そのキーワードをWeb検索で検証し、有効性を確認した
 - ② 次に、テキストマイニングを用いて、DMが行われている可能性が高い組織と、そうでない組織の違いを可視化した
 - ③ 最後に、その可視化の結果の妥当性を、①のキーワード分析の結果と突き合わせることで検証した
- 以上の結果、公開情報にテキストマイニングを用いることで、行政機関内の活動の関係性を可視化し、DMの活動が行われているどうかを推定できることが確認できた

5-2. 研究成果のポイント

- ① 行政機関内のDMの活動に関して、他の方法では得られない情報を入手することができた
 - 特定の特徴を持つ活動を効率的に探索する手法として、他の組織活動への応用が期待できる

- ② 組織内でのDMの活動の有無による活動全体の関係性の違いを可視化して把握することができた
 - 組織内の特定の活動の状況を定量的に評価するための手法としての活用が期待できる

5-3. 本手法の限界と今後の課題

(本手法の限界)

- 文書間で比較を行う場合は、文章のレベルや分量がある程度揃っている必要がある。バラバラの文書間での比較は難しい（本研究で利用したテキストもそれぞれ同一機関が同一方針で作成していた）
- 活動の「特徴」を抽出するためには、活動の具体的な内容にまで踏み込んだ記述が必要（行政機関が公表している方針や計画といった、判で押したような公式文書ではほとんど特徴は抽出できなかった）

(今後の課題と可能性)

- 本研究で対象とした活動や文書の種類は限られたものであった。今後、他の種類の活動や文書にも本手法を適用し、どのような場合に適用可能性や有効性が高まるかを見極めていく必要がある
- 同手法に適した条件や適用方法が明らかになれば、雑誌の取材記事や情報システムの調達仕様書など様々な領域の文書にも応用できる可能性が出てくると考えられる。

引用文献

- 一般財団法人 関西情報センター, e-Kansaiレポート2016, 2016
<http://www.kiis.or.jp/research/e-Kansai/>
- 一般社団法人 行政情報システム研究所, 行政機関向けデータマネジメント導入ハンドブック, 2016
<http://www.iais.or.jp/ja/membersinfo/datamanagement2016/>
- 一般社団法人 日本データマネジメント・コンソーシアム, データマネジメント概説書 (J D M C 版) , 2015
- 株式会社 データ総研, DAMA-DMBOK データ管理知識体系ガイド (日本語版) ,2011
- 内閣官房 情報通信技術 (IT) 総合戦略室, オープンデータをはじめよう～ 地方公共団体のための最初の手引書 ～, 2016更新
http://www.kantei.go.jp/jp/singi/it2/densi/kettei/OpenData_Tebikisyo3.pdf
- 総務省行政管理局, データカタログサイト
<http://www.data.go.jp/>
- 浦安市,浦安市におけるGIS及びオープンデータの取り組みについて, 2016確認
<http://www.gsi.go.jp/common/000133677.pdf>
- 塩尻市,地域オープンデータプラットフォーム、及び分析結果のデータの環流の仕組み作りと土砂災害防止への応用, 2016
http://www.vled.or.jp/committee/20160121_siryu3-4.pdf
- 草津市, 草津市のオープンデータのあり方に関する調査研究報告書, 2016
<http://www.city.kusatsu.shiga.jp/shisei/kenkyu/chousakenkyu/kusatsumirai280706.html>