

健康情報活性化のための公共データ開放とビッグデータ活用

이원석교수 李オンシク教授
延世大学校工科大学



1. 公共データ開放と活用

가. 政府3.0と公共データ開放

2014年韓国に住む私たちは、いつでもどこでも手の中のスマートフォンで希望するすべての情報が活用できるようになった。これはこの20年来夢見て来た情報化社会の仕上げ段階であることを意味する。特に UN の電子政府大賞を 3回連続して受賞した我が国の公共分野はほとんどの分野で早いインターネットで繋がれた運営システムを利用して業務の効率性を極大化している。

このような運営システムには私たちの社会のすべての改変が膨大な量のデジタルデータで蓄積されている。最近政府が積極的に推進している政府3.0は開放、共有、消費及び協力作用という革新手段で透明な政府、有能な政府、そしてサービス政府を実現して需要者のオーダーメイド型サービスを提供して、働き口と新成長動力を新たにつくることで国民すべてが幸せな大韓民国を作るという概念である。

ITの発展を顧みればハードウェアからソフトウェアに移した重心が、ソフトウェアからデータに移っている。したがって現在我々の公共分野で縮尺された多様で膨大な量のデータは我が国の未来新成長動力のための重要な資源であることを皆が共感している。ただデータという新しい IT 発展の軸をどのように展開させなければならないかに対する回答である。

このような状況は ITを善導する我が国がデータ時代に進入するファーストムーバ(FIRST MOVER)として当然経なければならない段階である。

この 30余年間、国家次元の情報化推進政策を通じて成熟した情報化社会を成した我が国で、政府 3.0の成功は一言でいえばデータの開放、データの共有、データを通じるコミュニケーション、そしてデータによる協力作用を通じて、効果的で創意的な方式で経済発展の原動力を捜すことができるように走っている。したがってデータに基盤を置いた透明な政府、有能な政府そしてサービス政府への変貌は短期間に可能ではなく、長期間に社会全般の基質と役目を変化させながら段階的に進行される課業である。

公共データとは、個別の公共機関が自分の固有した公共業務を情報システムを活用して遂行しながら縮尺されたデータを意味し、該当の業務に関する詳細な情報を盛っている。去る 1年間政府 3.0の初期革新戦略で積極的に推進している公共データの開放は、今まで公共機関で閉鎖的に使いつつ、個別的な情報公開請求制度を通じて接近していた公共データを、特別な事由や根拠がない限りすべて開放するという政策である。

すなわち個別公共機関が持っている公共データの開放は、その機関が該当業務を遂行した内訳

をすべての国民に詳しく知らせることで、機関の業務遂行透明性を進めるので、すべての公共機関の公共データ開放は政府3.0の透明な政府を実現して国家政策の透明性を、データを使用して確実にするという意味である。

18世紀イギリスの産業革命をきっかけに分業化されたプロセスを通じて効率的な大量生産体系を構築しながら産業化社会が始まった。今日の情報社会ではすべての分野でこのような分業化プロセスをインターネットとコンピューターを活用してより効率的に業務を遂行している。

公共機関も国家次元の功績な事を垂直分業化によって遂行している。つまり、個々の公共機関の公共データには、当該機関が分業して担当している公共業務やサービスの内容を詳細に記している。したがって種々の公共機関の公共データを意味的に結合すれば個別機関の業務担当者は把握できなかつた新しくて有用な業務関連知識を見出すことができる。

このような公共データを連携して活用することで、さらに新しくて有用な公共業務関連知識を見つけて実際業務に積極的に反映し、新しい次元の公共サービスを開発することで、政府3.0の有能な政府を実現することができる。同じく公共データと多様な民間データとの連携を通じて、公共と民間分野間の新しい知識を見つけことにより、公共機関が提供する公共サービスよりさらにアップグレードされた新しい次元の対国民オーダーメイド型サービスを提供する民間企業が現われることにもなる。

一歩進んでこのような改変は民間分野間のデータを結合し、活性化を通じた多様なデータを流通する産業を新しくして、データ関連民間企業の活性化及び関連分野の多くの新しい働き口を創出することになる。すなわち公共機関が民間企業で見つけた新しい公共業務関連知識や新しく開発した対国民オーダーメイド型公共サービスを積極的に収容することで、政府3.0が提示する新しい次元のサービス政府を実現するようになる。

4. 公共データ開放現況

公共データ戦略委員会によれば 2014年 8月時点で開放された公共データ数は全9,354で、前年度対比 96% 増加、公共データと繋がれたオープンAPI数は全 649で前年比 28%が増加した。また公共データを活用したウェブ開発事例は 2013年の42件から 2014年は318件と 6.5倍に増加した。

実際データ活用程度を示す開放されたデータセットのダウンロード件数は 2013年 4,108件から 2014年 8月 35,862件と 7.7倍増加し、オープンAPIで提供した月平均トラフィック件数は 2013年の133百万件から 2014年 220百万件に 65.4% 増加した。

去る 1年間の公共データ開放政策は量的に多くの成果を得たが、何より重要な成果は公共データが当該公共機関の所有物であるという過去の認識から、政府3.0の透明な政府を作るために国民に開放することが目的であるとの認識の改変である。

これから量的な側面だけではなく質的な側面で、公共データ開放の形態を改善しながら政府 3.0の透明な政府、有能な政府、そしてサービス政府を実現するための努力を持続的に進行しなければならない。

2. 保健医療ビッグデータ開放と活用

ガ. 政府 3.0とビッグデータ

ビッグデータは単純にデータの量が多いことを意味するだけでなく、既存データに比べてあまりにも膨大で現在までの方法や道具では回収、保存、検索、分析、視覚化がしにくい異型データの集合であることに特徴がある。ビッグデータは膨大な規模(VOLUME)、早い処理速度(VELOCITY)、多様な形態(VARIETY)、新しい価値(VALUE)の'4V'にその特徴が整理できる。

より具体的には表1(略)のように大規模、現実、時系列性そして結合性が特性である。ビッグデータの様々な定義が議論されており、概念自体も進化している状況で、断片的にビッグデータを定義するよりも、複数の視点からビッグデータを理解することが重要であり、ビッグデータ自体ではなく、これを活用して、新しいインサイト(INSIGHT)を創出する高度な分析の重要性がさらに強調されている。

SNSなどソーシャルメディア(SOCIAL MEDIA)の急成長に伴い、個人情報や消費パターン、位置情報などが含まれている貴重なデータが毎分、膨大に生成されており、物事インターネット(INTERNET OF THINGS、IOT)などが活性化しIT機器も多くのデータを直接生成し始めた。このように、新たに生成され通過する膨大な量のデータ、つまりビッグデータの多くは、非構造化データであり、それが全データ量の約80%に達している。

従来の方法では、これらのビッグデータ分析にはかなりの費用と時間がかかり、分析方法が制限されるために、ビッグデータに応じた新たな分析方法と技術が登場するようになった。ますます複雑になり多くの危険が発生する最近の社会・経済的環境で特にリアルタイムビッグデータ処理ができるようにする大規模データ回収、管理、処理、分析技術の発展は政府と企業が市場により効果的に対応することができる洞察力と価値を提供してくれ始めた。

実際にGOOGLE、フェイスブック、アマゾン、ヤフーなどのソーシャルメディアやインターネット企業等がビッグデータ基盤のソーシャル分析効果を立証しながら海外主要国家の政府次元でビッグデータを活用した公共サービス運営をはかっている。民間企業もビッグデータを活用して収益性の向上と処理の合理化はもちろん、ビッグデータ基盤の新しいビジネスモデル発掘を推進しながらビッグデータ市場が本格的に成長している。

現在までの情報システムは、私たちの社会の情報をあらかじめ決まった定型化された枠組みで保存して処理し、その量が膨大な場合には標本調査に基づく統計的な分析だけが可能であった。一方、ビッグデータ技術はいかに膨大な量のデータでも、その大きさに関わらずすべての種類のデータを対象に迅速な分析が可能という点で社会のすべての改変を極めて細密な単位で詳細な分析ができるようになったことを意味する。

産業化社会で大衆の体格の統計値で既製服を作る場合、ビッグデータ技術はすべての個人の体格にジャストフィットするオーダーメイドの服が作れる技術である。

4. 公共ビッグデータ活用現況

2011年11月、大統領所属の国家情報化戦略委員会でビッグデータ活用の必要性が我が国で最初に提言されて以来、2012年には公共ビッグデータ活用マスタープランが企画され、2013年から政府3.0までにはほとんどすべての部処と地方自治体でビッグデータの活用が積極的に試みされている。安全行政府と未来創造科学部は表2と表3(これは略しているが末尾に資料添付)のような多様な公共エリアでビッグデータ技術を試験的に適用している。

国内民間部分もビッグデータを活用したサービスモデル活発に試みている。クレジットカード会社は顧客取り引き、ソーシャルメディア、加盟店などで発生したビッグデータを利用して加盟店と顧客消費パターンなどを分析してリアルタイムマーケティングに活用する方を模索している。

ロッセマート、ロッセデパート、ロッセホームショッピングなどを通じて顧客の購買情報を手軽に分析することができるロッセカードの場合、'ビッグデータマーケティング'を通じてクーポンを受けた顧客が実際商品を購入する割合である'マーケティング反応率'が既存対比大きく増加したことで確認された。相加平均 43.7%だったロッセマートのマーケティング反応率は 65.9%。ロッセデパートは 30.8%から 36.1%で、ロッセホームショッピングは 11%から 15.2%で大きく上昇した。このうち、おもちゃの売上げは前年対比 7倍も増加した。

三星カードは'Mポケット'サービスを通じて 250万加盟店に対する決済データを分析して売上高と使用者決済件数を基準で加盟店利用形模様を分析している。分析結果を基盤で顧客に地域別、年令別、性別、時間帯別週末/平日など条件によって状況にあう検索ができるように提供している。

SKカードはビッグデータを活用して顧客オーダーメイド型イベントとクーポンを提供する'ゲットモア(GET MORE)'サービスを成功させている。

加盟店と係わるイベント情報と顧客の決済データ分析を通じる個別顧客のカード使用形模様によって顧客オーダーメイド型ショッピング情報を提供するのが特徴である。

2013年にビッグデータ試験事業に進行されたソウル市の深夜バス路線最適化モデルは民間ビッグデータである KTの移動通信データとの結合を通じて公共業務の効果性を進めたモデルで 2014年多くの地方自治体で実施されており、カード会社の決裁内訳ビッグデータを結合して地方自治体の祝祭マーケティングの正確さを進めている。SKテレコムはティマップ(T-MAP)ナビゲーションを通じてコールタクシー、油類運搬車、高速バスなどに GPSを装着して回収される膨大なリアルタイム交信情報をリアルタイムに分析して目的地までの最短距離と正確な到着時間を提供している。ビッグデータ基盤で地図と流動人口、業種別/月別売上げなど 10余種の情報を分析して商圈分析サービスを提供している。

4. 保健医療データ開放と活用事例

保健医療分野のデータは多くの個人情報や個人別診療情報を含んでおり、ビッグデータ分析の活用性は非常に高いがデータの開放や活用が制限的である。健康保険審査評価院は患者個人情報を除いて疾病別、特定医療行為別、治療材料別などに集計された診療情報と病院情報など保有した各種保健診療情報を APIを通じて民間に開放しており、健康保険審査評価院が保有した年間 200億件(49.5TB)に達する保健医療ビッグデータを活用した多様なサービス開発と産学研究が可能な保健医療ビッグデータセンターを運営している。

食品医薬品安全庁は健康機能食品、検査不向き食品、国内訴訟医薬品など 85のデータを公開している。国民健康保険公団では 5,000万名のヘルスサービス加入者の中で 100万人に対する 2002年から 2010年までの診療内訳、検診結果、居住地及び保険料、療養機関情報等を個人識別が不可能な形態で抽出した全国健康情報標本データベースを研究用に提供しており、国立保健研究院では中央人体資源銀行を運営して全国 17病院で遺伝担体疫学調査などに確保した一般人 32万名、病院で確保した疾病群 20万名の疫学情報、血液、組織、DNA など情報を収集して提供している。

国民健康保険公団の国民の健康検診資料にある個人の視力検査結果を連繋活用して運転免許証の発給プロセスを簡素化している。メデーベンチャーズ、メディラテは、健康保険審査評価院のビッグデータを利用して使用者に従って地域別、恩恵別、距離別などリアルタイム更新のオーダーメイド型情報による最適病院の選択を行っている。また、実際の治療を受けたお客様のみレビューを書くことができるシステムで虚偽の作成の可能性を遮断し、治療を受ける治療費のうち

非給与項目について最大10%までの報酬にして返し、消費者に医療費を節約できるようにして獲得されたポイントは、様々な商品と交換が可能である。

2013年未来創造科学部のビッグデータ試験事業で推進された国民健康注意予報供覧サービス事業は国民健康保険公団の診療内訳データとツイッターデータを結合して疾病危険を予測する課題を遂行した。医薬品安全性早期警報サービス事業ではネイバーのブログやカフェに上って来る医薬品副作用に関連する内容を集めて分析し、医薬品の副作用かどうかを早期に知らせるサービスを試験的に実装した。

また保健医療ビッグデータ活用サービス試験事業では次のソーシャルデータとソウルアサン病院のデータを活用してインフルエンザ流行予測、心臓不整脈予測、麻薬類監視、そして病院病床割当最適化サービスを試験的に実現した。2014年にも健康保険審査評価院の診療情報を基盤で疾病予測とオーダーメイド型病院情報を提供する医薬品と診療情報を統合して医薬品副作用の原因分析を糾明する事業を進行している。

3. 保健診療情報活性化方案

現在まで国民の診療歴情報は保健医療行政のみのために管理されてきたが、これからは国民ひとりひとりのためのオーダーメイド型保健医療サービスを実現する未来核心資源に活用されると予想されるので新たな用途に適合するように保健医療データの生産と管理プロセスと開放範囲を早急に整備する必要がある。政府3.0の透明な政府のための公共データ開放は政府の行政プロセスを明確に見せてくれるデータを対象にする。

一方、有能な政府やサービス政府のために公共機関や民間企業のビッグデータ活用はその区域と視標がより広くて多様なことがある。宝石の原石に例えられるビッグデータは、産業的民間活用性が非常に高いので現在の公共データ開放方式で誰も活用することができるように開放することは、国内で生産される主要資源を外国企業が何らの条件なしに採鉱して行くのと同じようになる。

また、公共ビッグデータの活用性をより向上させるためには、すべてに公開することは困難で敏感な個人情報や、個々の医師の診療行為情報を公益次元に限った範囲と方法で活用しなければならない。このような観点で見れば、保健医療分野で公共データ開放とビッグデータ活用には二つの命題をより明確に区分しなければならないと考える。

ビッグデータの最大の長所は分業化された情報化社会の互いに相違する分野で生産されたデータの結合で、その分野の専門家も把握できなかった新しい事実が見つかることができるということである。

公共エリアの医療行政データと民間病院の医療の臨床データ、そして人間誘電体および蛋白質の研究データなどが一つに接続されて個別に組み立てられ、リアルタイムビッグデータ分析が可能となるなら、人類の文明に多大な波及効果を発生させるきっかけになると考える。このような特性を最大限活用するためには保健医療分野のビッグデータ生産を担当するコントロールタワー (CONTROL TOWER) と保健医療分野に相応しい新しい活用方案を創意的に独立したシンクタンク (THINK TANK) が必要である。

各機関で運営されている個別システムから保健医療分野の融合的な観点で活用性が高い主要ビッグデータを掘り出して、ビッグデータ活用に対応しいビッグデータ品質指標を設定して新しく流入するビッグデータを常時監視するシステムを備えなければならない。制度的な問題と機関の間の利害関係によってデータの回収と融合が妨げられないようにしなければならない。ビッグデ

ータの逆機能を賢明に解決しながら主要ビッグデータを個人情報及び営業情報敏感などによってグレード化して、ビッグデータ管理主体を指定してコントロールタワーを中心に貴重なデータを資源化する保健医療ビッグデータの体系を確立しなければならない。

保健医療ビッグデータシンクタンクは保健医療ビッグデータを総体的に活用する新しくて多様な方法を持続的に捜して既存の保健医療行政業務に反映して対国民行政業務プロセスを持続的に好適化することによって政府 3.0の有能な政府が実現する。これと同時に貴重な資産である保健医療ビッグデータを活用する民間企業が新しいビジネスモデルに永続的に成長する時政府3.0のサービス政府が実現する。政府はこのように公共データを提供して民間企業を通じて国民に新しい働き口と新しい次元のオーダーメイド型保健医療サービスを提供することで政府3.0の窮極的な目標である国民幸福時代を達成することである。

4. おわりに

ビッグデータはインターネット技術以後に近づく未来社会への重要な変革を始める技術である。今後 20年以内に人類は 600億個のスマート機器を使いながら毎日 15ペタバイト(PB)の分量の個人別生活情報を生成することが予想される。ビッグデータ技術が善導する未来は社会が作動しながら発生したすべてのデータを詳細にモニタリングしながら個人の価値と大衆の価値を最適に調和させながら各自に与えられた状況に最も適合した合理的な方式で業務が処理されてサービスされる最適化社会になるだろう。

インターネット時代に経験したように、新技術によった急激な改変は社会各分野に新しくて多様なビジネスモデルを創出させてこれを通じて個人や企業そして国家が発展することができる大事な機会を提供する。なおかつ成熟した情報社会を具現した我が国で、個別公共機関の情報システムを通じて持続的に回収されているデータを国家次元で付加的な活用方法を捜すことは我が国の未来競争力を牛耳る核心力になるだろう。

政府 3.0の公共データ開放の努力はこのような観点の手始めでデータが中心になる未来時代を開く重要な出発点であると思われる。

30年以前から我が国の上位 1%の人材が医学分野に進出して IT 強国である我が国の情報システム環境で保健医療ビッグデータを生産している。したがって我が国の診療行政ビッグデータの正確さは非常に高いと予想される。

18世紀数学者コンドルセの「陪審員理論」によると、陪審員ひとりひとりが正解を当てる確率が1/2を超えると陪審員の数が多いほど正解を得る確率が1に収束するというものである。上位 1%の能力を取り揃えた我が国の医師ひとりひとりの診療行為が正確である確率は当然 1/2を超えるはずであり、このような医師の診療行為をすべてビッグデータに集積して分析すれば常に正解が得られるという意味である。

国家の優秀な人材が持続的に導入されてきた保健医療分野は、公共データ開放を越して我が国の未来核心成長動力を新たにつくる保健医療ビッグデータ活性化のための真摯な歩みを始めなければならない時である。

すべての資料がデジタルデータ化された情報社会において国民ひとりひとりの IT 機器活用度が世界最高水準である我が国は保健医療ビッグデータで始まる人類の未来社会進化を主導的に導いて行ける十分な潜在能力を備えている。

振り返って見れば今までの情報化は産業化の業務プロセスをそのままコンピューターに移植

する単純作業で川を下流に沿って運行してきた。ビッグデータを活用する未来最適化社会への移転は、まるでコロンブスが新大陸を尋ねる航海に比喻することができるだろう。

予想することができなかつた多様な形態の挑戦を肯定的な思考で賢明に乗り越えながら我々が IT 強国になるためにコンピューターの HW と SW に対する研究開発 (R&D) に多くの努力を競ったように、データが中心になる未来時代に相応しい新しい形態のデータ R&D 体系を作って実際データと HW 及び SW が融合される新しいデータ基盤先端道具を先導的に作らなければならない。

それと同時に民間と公共の多様なビッグデータを安全で活発に流通させることができる新しい観点のデータスタイルの保健医療行政体系とデータスタイルの法制度の用意が必須である。IT 強国であり毎年修学能力成績の上位 1% が医療人力になる我が国が、近い将来に人類の保健医療と生命産業に責任を負う国家になる日を期待している。