

“多様なデータ資源と消費を結びつける 情報共有システム構築に向けた基本的な規約 Ver1.0” 研究の意図について

規約が実現しようとしているもの

この度、情報環境研究会がリリースする「多様なデータ資源と消費を結びつける情報共有システム構築に向けた基本的な規約 Ver1.0」(以後、規約と表記します)は、インターネット等のネットワーク上に散在するデータ資源を的確に参照し、無駄なく利用することを実現する情報共有環境を実現するための規約としてまとめたものです。

インターネットの普及に伴い、ネットワーク上にはありとあらゆる種類の情報が存在するようになり、それらを多くの参加者が利用しています。ネットワーク上で公開、提供されているデータ資源は多種多様に渡っており、巨大なデータベースとして準えることもあります。このようなインターネット上に自然発生的に多数の参加者の持ち寄りによって構築されたデータ資源の集積は、データ資源の所有者・提供者の善意(意図)によって公開されており、多くの場合は公開・提供までで所有者・提供者の目的は達成されていると考えられます。一方、それらのデータを利用する側は、“公開されたデータ資源”として消費する者の論理によって利用しており、利用する側で恣意的に解釈し、場合によっては再加工するなど、適正に或いは有効にデータ資源として利用出来ているとは言い難いです。

ネットワーク上で多くの参加者によって提供されているデータ資源とそれらの消費は、それぞれが互いの自由意志で行っているだけであり、データベースや情報共有システムに準えることがあったとしても、実際の有り様を一面から観察した場合にデータベースや情報共有システムに準えることが出来るだけのことであり、データ資源の提供と消費が結び付いたシステムとは言い難いのが実情でしょう。

当研究会が規約の中で直接、間接にふれている情報共有システムとは、ある情報が必要となった情報利用者が、その目的とする情報を所有するサービスやシステムを発見し、粒度が整えられた情報として参照するものです。つまり、インターネット上に存在する多様なデータ資源とそれらを必要としている情報の利用者とは結びつけて、それぞれの論理・目的に適合する情報共有環境を目指すためのものです。情報利用者が参照した情報を加工・編集し、再生産された情報が再び情報共有システム上で共有されることを規約の中では考えており、一方的なデータ資源提供や資源の消費を目的にせず、それぞれの立場で利用可能な柔軟な構成を目指しています。多様性のある情報共有システムにするには、データ資源の提供側や消費側のどちらか一方に基準を置いた環境・システム構築ではなく、データ資源発見や交換のための手続きや、交換・流通するデータの記述にまで柔軟性と汎用性が求められることとなります。当研究会が規約として公開する文書では、このようなシステム・環境を構築するために必要な考え方や解決すべき事柄を整理したものです。そのため、従来の情報システム構築では触れられることが少なかった、データ資源の再利用、再生産に伴い、参照されたデータとそれらから再生産されたデータの間には発生する関連性を示すデータや、派生するデータの信頼性確保等についても触れることとなりました。

このような情報環境の有り様は決して目新しいものではないでしょう。但し、ここで解決したいと考える問題点について理解していただくことによって、当研究会が目指す情報共有システムについてご

理解いただけるものと考えます。

根源的な問題意識

情報環境研究会は、2006年度に内閣府が実施した「電子化に対応した経済社会統計のあり方の調査研究」にその基礎を置いています。この研究では、統計データを調査研究の対象とし、情報所有者が保有する最小粒度のデータを「マイクロデータ」と定義したうえで、このデータをネットワーク経由で参照することで、情報利用者が正確なデータを迅速に入手し、様々な表章を可能とするシステムの構築を提言しています。

その後、情報環境研究会の前身である「マイクロデータ研究会」は、内閣府の調査研究を基礎に様々な検討を行い、マイクロデータを「システム間でデータを流通させるために共通化された規格を有するデータ」と再定義し、これに基づき今日まで研究を重ねてきました。

また、同時期に行われていた政府の電子行政の取り組みと連動して、統計業務だけではなく、申請届出業務への応用も検討してきました。

こうした過程で明らかになった課題は、時間経過による情報の変化等の理由が存在するにしても、殆どの場合は情報所有者の一つのデータ資源から発信されたデータが、情報利用者に受け取られることにより新しいデータとして存在してしまうということです。本来一つのデータ資源であったデータが異なるデータとして数多く存在し、同じか異なるかの判別も厳密には行われなまま、実質的には異なるデータとしてそれぞれが存在し流通し、その過程で粒度にもばらつきが生じているように思われます。一つのデータ資源から発信されたデータが複数の実体として存在することの原因には、手続きや様式、組織によってそれぞれの独自の呼び方や書き方を要求することで、データ名や取り扱い単位の違いが多数のインスタンスを生むことになり、それぞれが異なるデータとして管理されているのは容易に想像できます。

つまり、データを入力する際には、そのデータを利用する側の利用目的に即した入力方法や形式が要求されるので、それらは同じデータであっても大本の情報所有者から提供された時点で別なデータとして扱われることになります。

また、利用者に提供したデータのインスタンスは、その流通の過程では複写や転送が繰り返すことも想像に難くありません。特に電子データであれば“原本”という考えを適応すること困難であり、その反面として複製の容易さを特徴としてあげることができます。データが作業仲間を流通する過程は、データが繰り返し複製されること同様な状態であり、大本のデータ所有者とは無関係なデータとしてデータ増殖し続けることを意味します。あまり意識されることもなく流通過程で複製され続けたデータは、増殖を繰り返し、所有者には一切の管理も検証も出来ないものとなっていることも問題のように思われました。

流通過程での複製が行われたデータに、流通の過程でデータに対する改変や修正が発生しても、その事実が発信元となった所有者にそのことが伝えられることは希ですし、所有者以外からそのデータを受け取った利用者にとっては、そのデータを最初に発信した人が作成、入力したデータであることしか出来ません。提供と利用を繰り返す裡に、情報は制御できないくらいに多重な存在となり、原点の提供者(所有者)からは完全に切り離されたものになっています。流通過程で誤りや改変が発生していたとしても、それを発見することや訂正することは容易ではありません。

この時点で考えられた問題は以下のようにまとめることができます。

1. データ入力とデータ利用が固定化されている。
 - 利用者が定める仕様に従って、所有者はデータを作成しなければならない。
 - 求める入力書式、送信手続きに従ってそれぞれに作成を行わなければならない。
 - ➔ 同じ内容のデータであっても、利用者が求めに応じて繰り返し入力、作成を行い

ます。その結果、定義と粒度が微妙に異なる大量のデータが生成されることとなります。

2. 手続きの数だけ情報がデータされ、結果として多数の「同じ(と信じられている)データ」が存在することとなる。
 - 所有者(発信者)が作成するデータは一つであったとしても、データの流通過程に存在するそれぞれの「業務手続」に従って、それぞれの担当者のコンピュータに繰り返し複製されることとなる。
 - データが個人情報などの機密情報の場合、手続きと関係者が増える都度、漏洩のリスクが増加する。
 - データは、作成者の管理下になく、どこから漏洩したかの確認ができない。(個人情報保護の課題)
3. データの正しさ(真正性)の確認が、繰り返し必要となる。
 - データ受け取りの時点でデータ所有者の認証を行った場合、次の過程では、「データ受け取り実施者」の検証に加えて、もともとのデータ所有者の検証が必要となる。
 - ➔ リレーを繰り返す毎に検証を増やさないと全体の正しさが確認出来ません。
 - 順次手続きが進むつど、検証の手続きが多段階となり、複雑となる。

素朴な解決案

最初に意識した上記の問題を整理すると次のように考えることが出来ます。

多くの所有者が持っているデータは、様々なシステムや業務等でそれぞれに inputs を求められることとなります。これはそれぞれのシステムが効率的に処理を行うために自らに最適なデータ inputs を求めるからです。データを所有する側からは複数の異なる形状の投入口があるように見え、それぞれに適合するように投入するデータの形状を整えて投入しています。逆に、データを利用する側が所有者の持つデータに合わせて複数の投入口を備える場合もありそうです。

データ資源と消費を結ぶ流通過程やそこを流れるデータを固定的(合目的で融通の利かないもの)としてしまうと、一つの資源から流通過程毎のデータを作成することとなります。

そこで、もしデータから合目的性を排除し、可用性の高いデータとして提供することが可能であり、データ処理のシステムが目的に合致したデータを求めるのではなく、目的に合致するようにデータを参照する(取り出す)ことが可能となれば、このようないたずらに複雑で硬直した情報処理環境から離れ、柔軟で発展性に富む情報共有環境が出来るのではないかと、以下のような素朴な解決方法を考えました。

1. データ入力とデータ利用を切り離す。
 - データの作成・入力は、指定された手続きに従って情報を作成する(ここまでは、従来と同じ)。
 - 一旦、作成された情報は、他の用途にも利用可能な汎用的なデータとして整えられる。
 - ➔ 情報の再利用を想定
2. 利用者は、必要な情報をその所有者が管理するサイト等から閲覧・参照する。
 - 利用者は、所有者から直接入手した汎用的なデータを、自らの目的に応じて編集する。

- 所有者は、提供している情報の利用者を認証し、閲覧権限を確認する。必要な場合に、いつ、だれがアクセスしたかの記録を保持する。
 - 所有者は自らの判断と責任のもとに、情報を開示する相手を選択できる。
 - ➔ 自分の情報は自分で守ることが可能になります
3. 情報の正しさは、所有者自らが客観的証跡をもって保証する。
- 所有者は、正当な利用者を確認の上、自らが真正性を保証する情報を利用者へ開示する。
 - 情報の真正性を検証する機関が必要となる。
 - ➔ あるいは真正性を検証する手続き。
 - 第三者機関が証明、あるいは検証によって所有者による「否認 (repudiation)」も回避する。
 - すなわち、所有者から提供されるデータは、「いつ」、「だれに」によって作成された情報（「どの」）であるかが保証されている

この解決案を現実のものとするには、一元的な方策を提示することで実現出来るものではなく、何を行おうとしているのか、何を解決しようとしているのかを明確にし、考え方としての共通認識の上に立ってより現実的な解決方法を構築する必要があると考えました。

そのため、まずはこの考え方を情報システムにより適合する形に整理し、実装方法ではなく、どのように考え、何を実現使用しているのかを「多様なデータ資源と消費を結びつける情報共有システム構築に向けた基本的な規約 Ver1.0」としてまとめることとしました。

今後の課題

規約で書き表した情報共有システムとしての考え方や方向性は、十分に検証されたものと断言することは出来ません。また、規約の考えや思想を使用できる状況についても十分な想定がされたとは言えません。規約として考え方や方向性をより明確に、現実即した意見や見識を取り込むのはもちろんのこと、現実のシステムとして実装を伴う検証も必要とされています。

そのためには、多くの方の参加を求め、規約の検証や拡大の作業を続けるとともに、規約の成果を実際のシステムとして利用して社会的な理解を得ることが必要です。