

企業間データ流通における、 データ提供者に生じるリスクの定量化手法の提案

片山翔子^{†1} 寺内敦^{†1} 山崎育生^{†1}

概要: IoT では様々なデータを分析し経営等に活かす事業価値創造の取り組みが重要であるが、近年は自社データのみでなく他社データも活用することで更なる価値創造が可能になると期待されている。その営みには多くの企業が所有データを提供することが不可欠だが、実際には個々の企業ではデータを提供する意欲があっても断念することが多い。その要因として、データ提供に関する種々の決定（相手企業の選定や価格の決定など）のための客観的な評価指標が不十分であることが考えられる。

本研究では、そのような評価指標の 1 つとしてデータ提供者に生じるリスクに着目し、これを定量化する手法を提案する。複数の実際のデータ提供事例において、提案手法により算出したリスクとデータ提供価格の間に相関があることを明らかにし、本手法がデータ提供に関する決定の 1 つの評価軸として有効である可能性を示した。

キーワード: データ提供, データ流通, リスク定量化

A Proposal of Quantification Method for Data Providers' Risks in Data Sharing between Several Companies

SHOKO KATAYAMA^{†1} ATSUSHI TERAUCHI^{†1}
IKUO YAMASAKI^{†1}

Abstract: A main objective of IoT is to create value by analyzing data for business. In the future, it can be expected that further value can be achieved by utilizing not only a company's own data but also that of the others. Therefore, data sharing between companies become of vital importance. However, most of the companies tend to hesitate to share data with other companies especially when there are a competitive relation sometimes being afraid of the rumor damage or misuse of the shared data.

In this paper, the method for quantification the data sharing risk called "loss-risk-index" for data provider is proposed. The index for actual cases of data provision is calculated as an evaluation for the loss-risk-index in this case. In addition, the data provision prices are compared. Results show that the loss-risk-index tend to correlate with the data provision price in actual cases. The proposed method is proved to be effective for making decision related to price and be qualified as one evaluation axis to determine how the data should be provided.

Keywords: data provision, data exchange, quantification the risk

1. はじめに

あらゆるものがインターネットに繋がり通信をする Internet of Things (IoT) の市場が急速に拡大している。2020 年にはネットワークに接続する機器は 500 億台になり、これらの機器から発生するデータ量は 40 ゼタ (40 兆ギガ) バイトになると予測され、市場規模は約 200 兆円になると見込まれている [1][2]。更に新たなセンサの開発などによりデータとして計測できる情報が今後も増え続けると予想されるため、企業や個人が保有するデータ量はますます増加することが見込まれる。

IoT ではデータを収集することではなく、そのデータを分析し自社の売上向上・コスト削減などに活かすといった価値創造の取り組みが重要である。例えば、ある飲料メーカーでは、自動販売機に取り付けたセンサから消費者のアイトラッキングのデータ (商品を購入する際に消費者がど

こを見て商品を確認しているのかを表すデータ) を取り、そのデータ分析の結果を元に商品配列を改め、売り上げを増加させた。具体的には、これまで飲料業界では注力商品を左上に陳列することが常識とされていたが、左下に陳列することで売り上げが増加することを明らかにした [3]。このように、IoT により各企業はさまざまなデータを分析し価値を生み出してきた。

上記で述べたデータ活用の例は自社で収集したデータを自社で利用する活用形態であるが、近年は異業種同業種を問わず他社の所有する IoT データの活用に注目が集まっている。例としては、自動車会社が持つ個人の運転データを用いて、その個人が交通事故を引き起こすリスクを計算し、それを個人の保険料に反映させる「テレマティクス保険」が挙げられる。テレマティクス保険は、各個人に合わせた料金でサービスが提供できるという点が特徴で、注目

^{†1} NTT 未来ねっと研究所
NTT Network Innovation laboratories

を集めている。このように、他社データの活用は、自社のデータだけでは出来なかった新たな価値創造が可能であり、今後さらに需要が高まると予想される。

本研究では、従来個々の企業に閉じて活用されていたデータを広く流通させ、他社データを活用することによる業務改善や新しいビジネスチャンスの創出などの価値創造を促進させることを目的とする。

2. 企業間データ流通

2.1 企業間データ流通の現状

他社のデータを活用するには企業間でデータを相互に流通させることが必要であるが、このような企業を跨るデータ活用を促進する動きは、内閣官房や経済産業省など国家レベルでの検討会や民間企業によるデータ流通のプラットフォーム提供など、近年活発化している。現在サービス化している具体的なデータ流通の方法は以下に示す4つのモデルに大別される(図1)[4]。

1. データホルダー型

自社のデータ(生データ、統計データに関わらず)を外部に提供、販売するモデルである。例えばNTT docomo[5]はスマートフォンから得られる位置情報と属性を合わせた統計データを販売している。具体的には、日本全国の1時間ごとの人口分布を24時間365日把握することができ、性別・年齢層・居住地域等毎の人口構成を知ることができる。

2. データアグリゲーター型

複数のデータ提供者から集めた情報を分析し、その結果を販売するモデルである。例えばPeek you[6]では、60以上のソーシャルメディア、ニュース、ブログなどを分析し、企業向けに消費者の傾向等の情報を提供している。

3. データマーケットプレイス型

データの提供者と利用者の間に入り、データの販売の仲介を行うモデルである。例えばEvery sense.inc[7]は、IoTセンサーデータを持つ企業と利用したい企業の間を取り持つ立場で、データの売買を仲介するプラットフォームを提供している。このプラットフォームの特徴は、データの抽象化とデータのフォーマットを合わせる役割を持つことである。

4. パーソナルデータストア型

3と同様にデータ提供者と利用者間の仲介を行うが、違いは本モデルではプライバシー・リスクのあるパーソナルデータを扱う点である。例えばDatacoup[8]は、企業等からのパーソナルデータ提供の依頼に対し、Webサイト上で個人に許可をとる。許可が取れた後にパーソナルデータが保存されている企業とアカウント連携を行い、データを取得し、匿名化などの加工を行って利用者に提供する。

2.2 企業間データ流通における課題

2.1で示した通り、企業間でのデータ流通は社会全体の期待も高く具体的な取り組みも行われる一方で、データを所有する企業単位では、データの提供に意欲があるものの実際にデータを外部に提供するに至っていない事例も多く見られる[4]。

1つ目の事例としては、都内約2万台のタクシーのプローブデータ(位置情報や加速度情報)を保有するX社がある。X社が保有するデータと、時間帯毎の通行可能な道路等の交通情報を合わせて解析することで、渋滞を考慮した最適ルート案内や都市計画の立案への活用が期待できる。X社には複数の企業からデータ提供の要望があり、同社はデータ提供を検討したが、プライバシー・リスクへの対応や、データ提供の方法に関していくつかの懸念事項があり、実際の提供には至っていない。また、2つ目の事例として、自社のナビゲーションアプリの利用者の移動履歴データを保有するY社でも、マーケティングへの応用を想定してデータの販売を考えているが、同じく提供には至っていない。このように、個々の企業においてはデータを外部に提供することは現状ハードルが高く、データが十分に流通していない。

2.1で挙げた4つのデータ流通モデルに共通して、一般に企業が自社データを他社に提供する際には、いくらか提供するか、どのようなデータ項目にするか、提供する手段など、多くの提供条件を決定する必要がある。これらの提供条件は、適切な評価軸に基づいて決定されるべきであるが(図2)、通常データ提供に伴う提供条件および評価軸は多岐に渡る上、データ提供に有効かつ客観的な評価軸も少なく、データ提供の実現を困難にしていた。

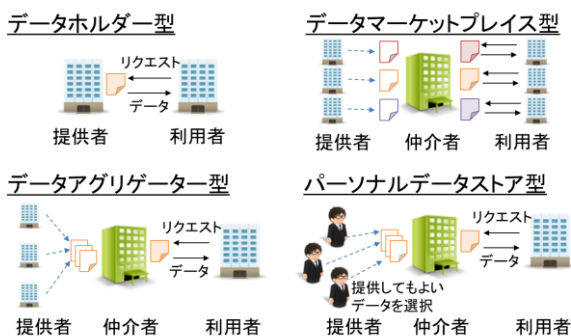


図1. 企業間データ流通4つのモデル

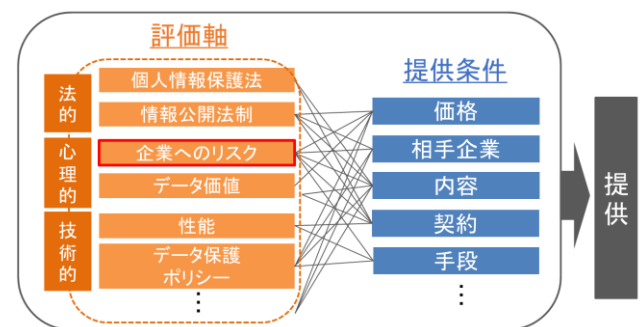


図2. データ提供のプロセス

データ提供における課題を法的、心理的、技術的課題の3つに分類した報告[9]があるが、提供条件決定のための評価軸にも同様に法的、心理的、技術的側面があると捉えることができ、例えば表1の各課題が解決されることにより提供条件の決定が行いやすくなり、データ提供が促進されると考えられる。

表1. データ提供における課題例

法的課題	データ所有権の帰属の決まりなどのグレーゾーンが存在すること等
心理的課題	データの使われ方に対する不安や、自社への影響の不安があること等
技術的課題	匿名加工の方法や暗号化方法が確立されていないこと等

これまで、法的、技術的課題に対しての取り組みは多く行われており、個人情報保護法の改定や、データの匿名加工技術の提案などがその例として挙げられる。

これに対し、心理的課題への取り組みは少なかった上、上記で述べたX社、Y社の事例[4]から、データ提供に意欲がある企業が最終的にデータ提供を断念した原因にはこの心理的課題が大きな影響を与えていることが分かった。特に、多くの事例に共通する心理的課題として「データ提供に伴う企業への影響」への懸念が挙げられていたため、これを評価軸として設定することが重要であると考えた。そこで、本研究では「データ提供に伴う企業への影響」を定量化することで、心理的課題の解決に取り組む。

3. 損失リスク指数の提案

3.1 アプローチ

企業への影響とはさまざまなものが考えられるが、ここでは特に、企業活動で重要な要素である経済的価値の観点からデータ提供に伴う企業への影響の定量化を行う。経済的価値とは大きく利益と損失リスクから成るため、経済的価値の観点で企業への影響を定量化することで利益と損失リスクのバランスを考慮してデータ提供条件の決定することが可能となる。

本研究では、利益はデータ提供価格により定量化できるとし、もう一方の損失リスクの定量化を試みた。

3.2 損失リスク指数の定義

本研究で検討する企業間のデータ流通では、データ提供者、データに関連のあるデータ提供者の顧客(個人/企業)、データを受け取るデータ利用者の3者によるモデルを想定し、データ提供者に生じる損失リスクを定量化した(図3)。

損失リスクの定量化にあたり、本研究では損失の積極度の観点からリスクを3つに分類して算出した。積極的損失とは支払いが生じること、消極的損失とは得られるはずの利益が得られなくなることであり、損失の積極度が高い順

に、損害賠償を顧客に支払うリスクをR1、秘密露呈により競争力が低下するリスクをR2、顧客離脱により逸失利益が生じるリスクをR3とし、それぞれを算出しその合計を損失リスク指数とした。各損失リスクR1~R3は、事象発生確率Pと損失規模Nの積として算出した。

データ提供におけるリスクに関連する従来研究としては、特に個人情報においてプライバシーを守るための匿名化技術などが挙げられ、データのみに着目した検討が多く見られる[10]。一方で、実際のデータ提供においては、データ自体の性質のみでなくデータを流通させる相手(データ利用者)や自社(データ提供者)との関係性なども考慮しなければならないと考えた。ここでは、データ自体に関する変数をC1、利用者に関する変数をC2、提供者に関する変数をC3とし、これらを変数として損失リスクを定量化することを提案した。各変数C1~C3は、企業のセキュリティ危険度や情報の区分法[11]など関連する既存の指標を活用した。変数C1~C3、損失リスクR1~R3と損失リスク指数Rの関係を図4に示す。

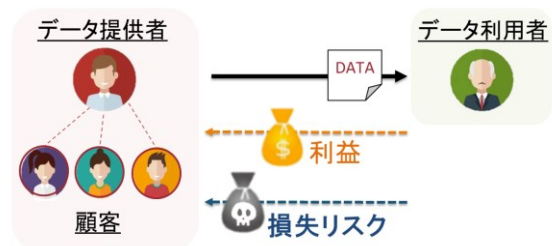


図3. データ流通のプレイヤー

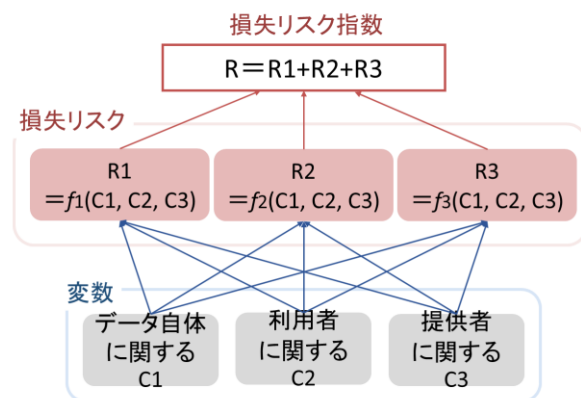


図4. 企業間データ流通におけるリスクモデル

各損失リスクR1~R3の定義を以下に示す。変数C1~C3の定義は表2の通りである。

▼損害賠償支払いのリスクR1

$$= P1 \times N1 = \{C21 \cdot C22\} \times \{C11 \cdot C32\}$$

損害賠償が発生する危険度(確率)と、想定される損害額の積で表す。損害賠償が発生する危険度(確率)は、顧客との契約違反が起こる危険度(確率) = データ利用者がデータを漏洩させるなどのインシデント発生を起こす危険

度(確率) ≡ データ利用者のセキュリティレベル C21 を用いる。また、提供者と利用者の業種の同異による危険度の差 C22 を乗ずる。想定される損害額は、データ種別により決定される額 C11 と顧客規模 C32 の積により表す。

▼秘密露呈による競争力低下のリスク R2

$$= P2 \times N2$$

$$= \left\{ \left(\frac{1}{k_2 + 1} \cdot C12 + \frac{k_2}{k_2 + 1} \cdot C21 \cdot C12 \right) \cdot C22 \right\} \times \{C31\}$$

企業秘密の露呈が発生する危険度(確率)と、想定される損害額の積で表す。企業秘密の露呈が発生する危険度(確率)は、データを提供した時点で発生する場合、またはデータが漏洩等した時点で発生する場合に分けられ、それらを足し合わせたもので定義する。これらは、データ提供者にとってのデータの重要度 C12 と、データ利用者のセキュリティレベル C21 によって表す。想定される損害額 C31 は、有価証券報告書などの公開資料から、対象データに関わる事業の利益額を抜粋して用いる。式中の k_2 は、データを提供した時点で損害が生じる危険性と、データが漏洩した時点で損害が生じる危険性の比であり、後者が k_2 倍危険であることを示す係数であり、ここでは $k_2=2$ と仮定した。

▼顧客離脱による逸失利益発生リスク R3

$$= P3 \times N3$$

$$= \left\{ \left(\frac{1}{k_3 + 1} \cdot C11 + \frac{k_3}{k_3 + 1} \cdot C21 \cdot C11 \right) \cdot C22 \right\} \times \{C31\}$$

顧客離脱により逸失利益が発生する危険度(確率)と、想定される損害額の積で表す(R3の求め方はR2と類似している)。データ提供者の顧客にとってのデータの重要度 C11 と、データ利用者のセキュリティレベル C21 によって表す。式中の k_3 は、データを提供した時点で損害が生じる危険性(提供の事実に対して嫌悪感を示すなど)と、データが漏洩した時点で損害が生じる危険性の比であり、後者が k_3 倍危険であることを示す係数であり、ここでは $k_3=5$ と仮定した。

表2. 損失リスク指数の変数一覧

変数	説明
データ 自体に 関するC1	データ種別C11
	データ提供者の顧客にとってのデータの重要度 想定損害賠償額の式を用いる[6]. = 基礎情報価値 [500] × 情報危険度 [Max(10 ^{x-1} + 5 ^{y-1})] × 本人特定容易度 [6,3,1] × 情報漏洩元組織の社会的責任 [2,1] 尚、xおよびyはそれぞれ精神的苦痛レベルと経済的損失レベルをあらわしており、データ種別ごとにx,yが決まる。例えば、「氏名」は(x,y) = (1,1)であり、より機微な「年収」は(x,y) = (2,2)である。
	管理区分C12
	データ提供者にとってのデータの重要度 {その他, ..., 極秘} = {0.1~0.5(5段階)}

データ 利用者に 関するC2	利用者業種毎のセキュリティレベルC21
	データ利用者のセキュリティレベル[8] 具体的には、各企業が自社のセキュリティについて答えたアンケートを元にセキュリティ危険度を数値化したものであり、業種ごとの平均値が公開されている[12].
データ 提供者に 関するC3	業種の同異C22 異業種間の場合、同業種間よりリスクが0.8倍と仮定 {同, 異} = {1, 0.8}
	財務情報C31
	四半期の営業利益をxとして、=x/{ln(x) · 10 ⁶ }
	契約者規模C32 顧客数(データ発生に関わる契約者数)をxとして、=ln(x)/10 ²

4. 損失リスク指数の評価

4.1 評価方法

提案した損失リスク指数の算出方法を用いて、実事例においてそれぞれリスク指数を算出し、その値の解析により提案した手法を評価する。具体的には、過去の実事例における損失リスク指数とデータ提供価格との相関関係进行评估する。

評価には、損失リスク指数算出に必要な情報(データ種別/データ提供者及び利用者の業種/営業利益/提供者顧客の規模)およびデータ提供価格が公開されている事例を用いた。本稿では、5件の事例による評価・考察を行った。用いた事例は、位置情報や住所などの個人情報を含む事例4件(No.1~4)と、含まない1件(No.5)である。各事例の説明を表3に示す。

表3. 評価に用いた事例の説明

No.	提供者	説明
1	企業A	携帯電話契約者の位置データや属性データを含む情報を匿名・統計化した情報を販売。 まちづくりや防災計画、企業の出展計画などに役立つと考えられる。また移動需要予測技術の開発にも利用される。
2	企業B	鉄道利用者の電車の乗降履歴や属性データを含む情報を匿名・統計化した情報を販売。 B社は社内でも公益利用を目的としたデータ活用を行っており、例えば駅構内の乗り換え動線を調査し乗り換え時間を軽減するための対策等実施してきた。
3	企業C	企業Cはインターネットサービスを展開しており、C社に登録している各店舗向けに、顧客情報を販売。 氏名、住所、クレジットカード番号等と合わせて購入商品や商品問い合わせ履歴等の情報を取り扱う。

4	企業 D	米国の事例であり、メールアドレスに年齢・性別・配偶者の有無などの 30 以上の属性が関連付けられた情報を販売。 米国消費者の 80%分の情報を取り扱う。
5	企業 E	エリアマーケティング向けに店舗の情報を収集し販売。 店舗名、住所、電話番号、緯度・経度情報を基本情報とし、その他駐車場有無や売り場面積等の情報を取り扱う。 個人情報に含まれない。

4.2 結果

各事例において求めた R1~R3 及びその合計値 R と、データ提供価格を表 4 に示す。ここで、データ提供価格は各事例において代表的な値を 1 つ選択して用いた。具体的には、価格が桁を跨ぐことを考慮し、取りうる価格の最小値 x_{min} 、最大値 x_{max} を用いて、 $10^{average(\log(x_{min}), \log(x_{max}))}$ と定義した。これにより、単純な加重平均により代表的な値が過度に高くなることを防いだ。

ピアソンの積率相関係数を用いて、R 及び R1~R3 と、データ提供価格との相関調査を行ったところ、R とデータ提供価格の間には高い相関がみられた（相関係数：0.98）。各実事例における損失リスク指数 R とデータ提供価格との関係を図 5 に示す。

また、R1~R3 それぞれとデータ提供価格の間の相関係数はそれぞれ、R1 : 0.82, R2 : 0.85, R3 : 0.81 であった。これらの値はサンプル数 5 に対して 5% 有意水準である 0.87 未満であったため、R1~R3 と価格には弱い相関があるといえる。

表 4. 損失リスク指数算出結果一覧

事例(データ提供者)	R1	R2	R3	R	データ提供価格
1 企業 A	266.0	210.4	114.6	590.9	10,000,000
2 企業 B	257.6	0.07	0.04	257.8	5,000,000
3 企業 C	419.0	61.7	22.9	503.6	7,286,975
4 企業 D	32.0	0.01	0.00	32.06	2,121,320
5 企業 E	0.00	0.27	0.00	0.27	100,000

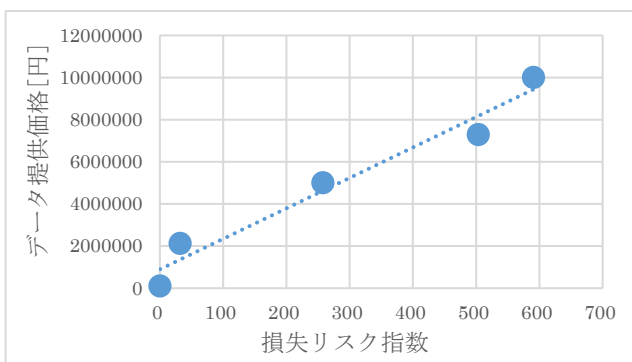


図 5. 損失リスク指数とデータ提供価格の関係

4.3 考察

4.2 の結果から、「損失リスク指数 R は価格に関連する意思決定の評価軸として有効である」という仮説が推測できる。

R が価格と相関する理由としては、例えば、評価に用いた 5 件の事例において価格決定時に実際に考慮されたパラメータと、R を導出するためのパラメータが類似しているためである可能性がある。価格決定の方法論として、通常の商品においては、その商品を作るためにかけたコストの観点や、市場性、需要供給のバランスの観点などにより価格決定が行われる。一方で情報の販売においては、販売後にも企業に影響が発生するなど不確定な作用（リスク）が大きいため、通常の商品販売における観点に加えてリスクを考慮することは自然であると考えられ、リスク R と価格の相関が高い原因であると推測できる。

また、R1~R3 も価格と弱い相関があるが、それらの合計である R はより強い相関を示していることから、価格決定においては R1~R3 を単独でなく総合的に評価することが重要であると推測できる。

一方で、上記の結果および考察は 5 件の事例によるものであるため、まず相関の有無を統計的に示すためには多くの実事例（必要サンプル数は、母数や許容誤差に依存する）を用いての検証が必要であり、今回の評価結果は予備的なものであると考えるべきである。今回、十分な事例数で評価が行えなかった理由は、R の算出に必要な情報である（主に）営業利益、もしくは評価に必要な情報であるデータ提供価格が公開されていない事例が多かったためである。これらの情報はデータ提供者であれば知りうる値であるため、例えばデータ取引の仲介をするような機関と協力してそれらの値を入手することで、多くの事例による損失リスク指数の評価が可能になると考えられる。今後は、まず多くの事例を用いた評価を行い損失リスク指数とデータ提供価格との相関関係が統計的に有意であることを確認する。両者の比例関係が有意であると示せた場合は、次に損失リスク指数とデータ提供価格との相関関係を正解値として、R 算出式の各パラメータを最適化するなどの調整を行うことで、提案する損失リスク指数の精度を上げていきたいと考えている。実際には、リスクとして考慮すべき要素が本提案のもので十分であるか等を検討し、式の調整と相関評価を繰り返し行うことで損失リスク指数の信頼度および精度を高めていくことを考えている。

また、データ提供者となる企業と検討を進めることで、実際の価格決定プロセスとの関連についても考察を行い、損失リスク指数が持つ意味や、データ提供条件の決定における効果を評価したい。

5. まとめと今後のスケジュール

本研究では、異種 IoT データ流通の促進を目的とし、データ提供時にデータ提供者に生じる損失リスクの定量化および評価を行った。

損失リスクを定量化することにより、データ提供時に決定すべき提供条件の決定に、リスクを反映できるようにすることを考えた。具体的には、データ提供により提供者が被る損失リスクを経済的観点から定量化することを提案し、算出の計算式を示した。提案した損失リスク指数がデータ提供価格と相関があることを明らかにし、データ提供に関する意思決定において有効な評価軸である可能性を示した。

今後は、提案する指数の改良、ツール化を進めながら、実ビジネスへの適用を目指す。

参考文献

- [1] Explosive Internet of Things Spending to Reach \$1.7 Trillion in 2020, According to IDC, 02 Jun 2015.
- [2] Peter Middleton, Peter Kjeldsen and Jim Tully, "Forecast: The Internet of Things, Worldwide, 2013," Gartner, November 18, 2013.
- [3] 総務省 平成27年版 情報通信白書
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/n_c254330.html
- [4] IoT推進コンソーシアム データ流通促進WG
<http://www.iotac.jp/wp-content/uploads/2016/01/資料4-ユースケースからみた論点の整理.pdf>
- [5] NTTdocomoモバイル空間統計
http://www.dcm-im.com/service/area_marketing/mobile_spatial_statistics/
- [6] PeekYou <http://www.peakyou.com/>
- [7] EverySense Inc <https://every-sense.com/>
- [8] Datacoup <https://datacoup.com/>
- [9] IoT推進コンソーシアム データ流通促進WG
http://www.iotac.jp/wp-content/uploads/2017/02/資料1-3_オムロン提出資料.pdf
- [10] 富士通研究所 PRESS RELEASE
<http://pr.fujitsu.com/jp/news/2016/07/19.html>
- [11] 2009年 情報セキュリティインシデントに関する調査報告書 第1.1版
- [12] 情報セキュリティ対策ベンチマーク
<https://www.ipa.go.jp/files/000048358.pdf>