３．データ・AI利活用における留意事項

東京大学 数理・情報教育研究センター

2020年5月11日

概要

- データやAIは，強力な道具であるだけに，使い方を誤ると人間や社会に大きなダメージを与えるおそれがあります．この節では，データやAIを使うにあたり最低限気をつけるべきことについて学びます．

- データやAIにまつわる基本的な倫理，合意事項について学びます．

- データを守ること，およびそれが破られて起こった事例について学びます．

目次

3-1. データ・AIを扱う上での留意事項

概要.

- 本節では，現代科学・技術が社会に対して果たすべき役割について考え，特にデータやデータサイエンス・AIを利活用する際に求められるモラルや倫理について理解することを目指します.
- いくつかの具体的事案を見ながら，データ駆動型社会における脅威（リスク）についても学びます.
- さらに，それらの流れを踏まえて，個人情報保護法やEU一般データ保護規則（GDPR）など，最先端の，データを取り巻く考え方や指針・法についても学びます.

3-1-1. ELSI：すべての科学・技術に関する普遍的考え方

“ELSI” = “Ethical (倫理的), Legal (法的), Social (社会的)
Implications (含意) あるいは Issues (事柄)”
= 「（科学における）倫理的・法的・社会的含意（事柄）」
= 「（科学技術を開発・展開した結果）
起こりうる倫理的問題・法的問題・社会的問題は何か」

1980年代に生命科学分野で，科学技術のみならずその社会的責任を考える必要を認め，提唱された概念です.
いまでは，あらゆる科学分野で必要とされています.

（参考；復習）従来の研究倫理の概念

従前より研究倫理として定着していたこと：

- 不公正・不適切な研究行為・発表の禁止：
  捏造・改竄・剽窃
  不適切な著者表示（ギフトオーサー，ゴーストオーサー）
  不適切なプレゼンテーション（チャンピオンデータ等）
- ヒト・生物に対する不適切な研究の禁止（→研究倫理審査）
- 違法な研究の禁止
- 利益相反

これらは**科学研究単体としての「健全性」「妥当性」.**
ELSI が考えるのは，より積極的な，
「社会の構成要素としての科学研究のあり方」

なぜELSIの考え方が必要なのか

[事例１：ヒトゲノムプロジェクト] 1990年頃，アメリカ

- ヒトゲノムの解読は，研究者や医療行為のみならず，すべての人，社会全体に影響が及ぶ．
  （遺伝情報＝個人情報の保護，遺伝情報差別の防止，…）
- そのため，研究者や医師・患者だけでなく，広い意味で社会がどこまでを「受容できるか」を議論する必要がある．

[事例２：原子力発電] 2011年，日本

- 東日本大震災による福島第一原子力発電所事故．
  日本でも科学技術・研究者に対する信頼が低下．
- 原子力発電技術はどこまで「受容できるか」の議論が必要．

社会は，科学技術をどこまで・どのように受け容れられるのか
科学は，社会にどこまで責任を持つのか
これらを議論する必要があります．

科学の責任：古典的な見方・より広い見方

【考え方１】科学は「直接意図した結果」にだけ責任を持つ.
→「狭い」「古典的」な見方
→想定外の事態まで責任は取れない
→科学研究は「善・悪」から切り離された客観的存在

【考え方２】「予想外だが引き起こされてしまった結果」にも責任を持つ.
→「より広い」「フォージの」見方
→過去の悲劇を振り返り，合理的に予想できる結果には責任を
→客観的事実の研究で含意がないから許されるとは言えない

（参考）ジョン・フォージ（オーストラリアの科学哲学者）

軽々にどちらとは言えません.
ただ，前ページの例を見ると，
「古典的な見方」だけでよいと言い切れないのは確かです.

あらためてELSIの３要素を見ると

- 倫理的問題（Ethical）：
  その科学（技術）研究がどのような倫理的問題をはらむか
  そもそも各科学領域における「倫理」とはなにか
  既存の倫理学の体系に収まるのか

- 法的問題（Legal）：
  その科学（技術）はどの法の枠組で捉えられるか
  現行法で不十分な場合，どのような指針・法の形成が必要か
  （関連して，どのような行政規制が必要か）

- 社会的問題（Social）：
  その科学（技術）は社会に受容されるか
  どのような形であれば受容されるのか
  メリットとデメリットのトレードオフ：社会の受容ラインはどこか

「より広い見方」に向けて，３つの観点に分けて，
科学と社会の関係を議論する場を提案しています.

データサイエンス・AIにおけるELSIとは

本教材でこれから議論していくこと：

- データサイエンス・AIの倫理とはなにか
- その法整備はどうなっているのか
- どのようなデータサイエンス・AIが社会に受容されるのか

3-1-2. データの倫理

まず考えるべきこと：

- データ取り扱いの健全性：　禁止事項（既出）
  捏造（ないデータを作り出すこと）
  改竄（実データを曲げて書き換えること）
  剽窃・盗用（データを不正に使い回すこと）

- データの保護（情報セキュリティ的な）
  →後述：第3-2節

- 個人情報とプライバシーの問題
  →後述：本節後半

データ取り扱いの健全性１：捏造

ないデータを作り出すこと.

実例：高温超伝導事件（米，2001年頃）

- ある若手物理学研究者が起こした事件.
- それまでの（有機物）高温超伝導記録（-240°C）を大幅に覆し，-156°Cを報告.　（参考：安価な液体窒素が-196°C）
- これを含み，Nature誌7本，Science誌9本等の論文を発表
- しかし該当する実験装置は発見されず，捏造が発覚した.
- 論文は撤回され，当該研究員は解雇された.

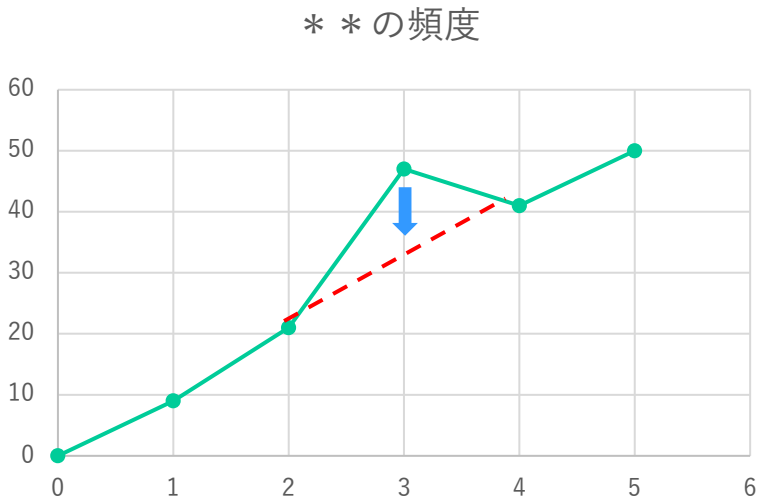
実例：STAP細胞事件（日本，2012年～2014年頃）

- ある若手生命科学研究者が起こした事件.
- iPS細胞より簡易な「初期化細胞」を作れる，とした.
- 査読（投稿論文を匿名の専門家が審査すること）で厳しい意見もついたが，結果的にNature誌に採択.
- しかしその後追試等でSTAP細胞は再現できず，調査の結果，捏造（や改竄・盗用）が発覚した.
- 結局，STAP細胞は既知の細胞の混入であることが分かった.論文は撤回された.

データ取り扱いの健全性２：改竄

実データを曲げて書き換えること.

例：実験データの一部を，つい曲げてしまいたくなる.



「このデータさえなければ
まっすぐなのに・・・」

実例：

- 鉄鋼素材品質検査データ改竄（日本，2017，製鉄会社）
- 免震装置データ改竄（日本，2018，産業装置製造会社）
- 電気泳動画像データ改竄（日本，2012，大学（生物系））

データ取り扱いの健全性３：盗用

データを不正に使い回すこと.

他人のデータの盗用は論外，自分のデータの「使い回し」も危険.

例：38本の論文で同一データ使い回し（日本，2010年，大学（医学系））
　　→「二重投稿（self-plagiarism）」に該当する

【参考】実際に起こる剽窃・盗用には，むしろ，
- アイデアの盗用
- 文章の剽窃　（自分の文章の自己剽窃を含む）

が多いです.
図表等を断りなく借用することも「剽窃」にあたります.
→ 適切な引用表示が必要です.

～余談～　個人のレベルでも

現状は「個人情報」「個人データ」にかかるリテラシーが低く，
情報暴露が盛んに起きているように思われます．

例：SNSに家族の画像を無断で載せるケース．
　　→例えば子の画像の場合，子が望んでいるかどうか不明．
　　（後述「忘れられる権利」が確立するまで消せない．）

例：SNSに友人の画像を無断で載せるケース
　　→食事・飲み会の画像など．日時・場所にかかる個人情報を暴露．

例：仮名でSNS活動している人にかかる個人情報を書き込むケース
　　→うっかり友人の本名・勤務先等を含めた投稿をする．

すべて，個人情報の暴露であり侵害です．気をつけましょう．

個人情報とプライバシー

個人情報とは：

- 個人ID（個人を特定できるデータ）
- その他の個人の情報・データ

プライバシーとは：

- 上記の個人情報・データを含み，それを「守ること」自体
- 個人情報の中で，特に「機微データ」「要配慮個人情報」とその保護

個人情報とは何か

- 個人ID（個人を特定できるデータ）
  - 名前，住所
  - マイナンバー，学籍番号，職員番号
  - 免許証番号，パスポート番号，など

- その他の個人の情報・データ
  - 関連属性：本籍，人種，年齢，家族構成，勤務先，所得など
  - 健康情報（病歴）
  - 成績・評価データ：試験，勤務評価など
  - 文化的側面：宗教，性的嗜好など
  - 信用履歴（クレジットカード）
  - メールアドレス
  - 購買履歴
  - 機械等の操作履歴（ウェブ閲覧をふくむ）
  - オンライン識別子（IPアドレス，端末識別紙など），など

比較的新しい

注意：この「個人情報」が誰のものか，は実は曖昧になりえます．
→あとで改めて議論します．

プライバシーとは何か

- 個人情報・データそのもの，に加えて，
  それらを保護する行為，をもふくみます．

- また，個人情報・データのうち，
  　特に注意を要するデータ：「機微データ」「要配慮個人情報」
  を特に指すこともあります．

  前項の個人情報・データの中では，例えば以下のものです．
  - 人種
  - 健康情報（病歴）
  - 文化的側面：宗教，性的嗜好など

3-1-3. データサイエンス・AIで起こりうる論点

データサイエンス，あるいはAIを用いる際，次のような問題が起こりえます.

- 統計的差別
- データバイアス・アルゴリズムバイアス
- 個人情報の暴露，プライバシーの侵害

統計的差別

統計的処理が妥当であり，その処理結果を用いる人が（偏見なく）合理的に判断している場合でも，結果として差別・不平等が肯定され，継続されうること.

典型的な場合：
雇用主が，採用段階において，個々の応募者が属する属性グループごとの「統計的平均値（あるいは平均的描像）」に基づいて能力を推測し，採用の判断をする場合.

例：「女性」は「短期勤続」（長く勤めないこと）の傾向がある
→そうしなくてよい社会になっていないことが原因
例：「黒人」は「生産能力」が低い傾向がある
→歴史的に平等な機会が与えられてこなかったことが原因
例：「外国人」は「日本語能力」が低い傾向がある
→日本語教育支援，および業務の多言語化がないことが原因

どれも，これまでのデータを統計的に処理すれば（相関関係としては）正しくても，「現状」は既に行われた不公平・差別の結果でもあり，直ちにそれに基づいて判断することは差別の延長になります.

データバイアス

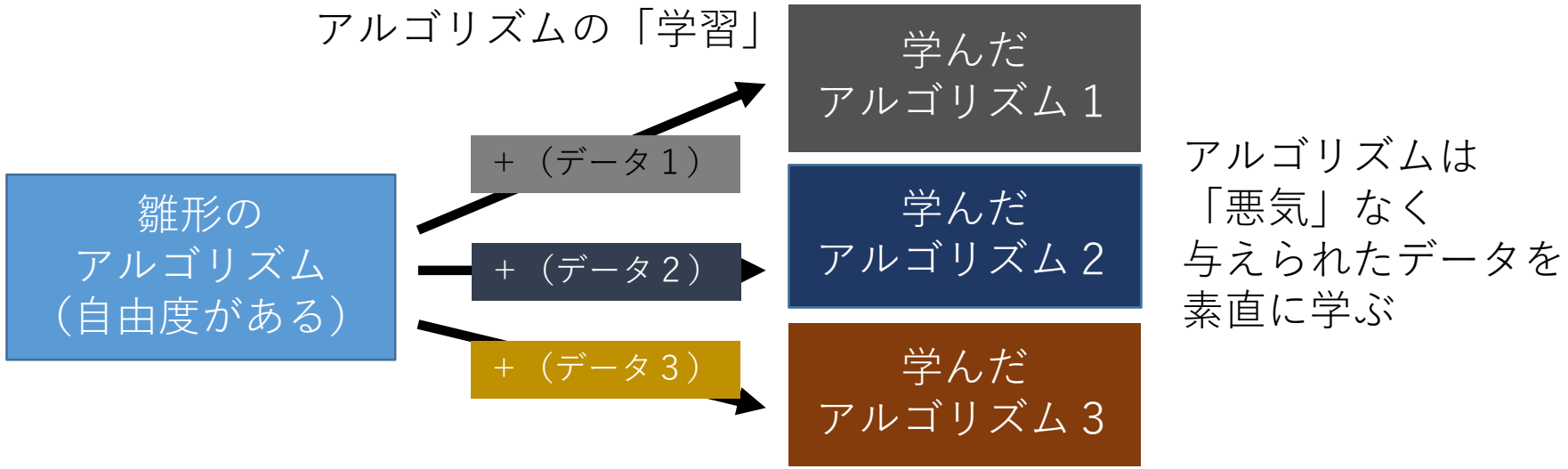
統計処理，あるいはデータサイエンス的処理を行う際扱うデータに，
そもそも　バイアス（bias)=偏り　があること，
あるいは，それに起因して起こる（よくない）こと．

例：　ある病気の治療方法について研究するとき，
特定の病院の患者のデータだけを集める
→　患者の地域性，富裕層・貧困層の偏り

例：　国内世論を推し量るのに，twitter 上の呟きから抽出する
→　そもそも twitter をやっている人の層，
その中でも頻繁に呟く人の層，の偏り

アルゴリズムバイアス

機械学習により，アルゴリズムにデータから学習させたとき，
データにバイアス（偏り）があったがゆえに，
学習結果のアルゴリズムにもバイアスが生じてしまうこと．



データ・アルゴリズムバイアスの実例

AmazonのAI人事アルゴリズム（米，2014年頃）

- Amazonは2014年頃から，新規採用に関して，AIによる自動書類審査の導入を始めた．大量の応募を迅速に処理可能．
- このAIアルゴリズムは，過去10年間の雇用パターンで「学習」したもので，アルゴリズムで応募書類をランク付けする．
- だが2015年頃，すぐに，Amazonはこのアルゴリズムが「性別に対して中立的」でないことに気がついた．
  「過去10年間の雇用パターン」は男性雇用に偏っていて，アルゴリズムはそれを素直に学習してしまっていた．
- アルゴリズムは“women’s” など，「女性」に関するキーワードにペナルティを科し，一方で男性が用いがちな表現に高い点を与えていた
- Amazonは，結局，このプロジェクトは停止した．

3-1-4. 社会的合意の形成に向けて

この節では，以下について見ていきます.

- データサイエンス時代の諸概念
  - 忘れられる権利
  - 説明に基づく同意
  - オプトイン・オプトアウト
- 形成されつつある合意の例
  - GDPR（欧州一般データ保護規則）
  - 人間中心のAI社会原則
- データサイエンスやAIの責任は誰が負うのか

諸概念１：忘れられる権利

インターネット時代のデジタルデータに関して，削除・アクセス遮断によりプライバシーが保護されることを求める権利.

[考えられるようになったきっかけ]
- デジタルデータがネット上に出回ると，消す手段がない
- 検索エンジンに登録された情報も，消す手段がない（たとえ犯罪歴のような情報であったとしても）
- 従って，通常のアナログ情報と異なり，永遠に「忘れられる」ことがない

[現状]
- EUなど一部の国では法制度化済みです.
- 日本では「インターネット法」に関係して議論が始まり，検討中.
- 「プライバシー保護」と「表現の自由・知る権利」の両立がなかなか難しいのも議論が長引く一因です.
- 法整備だけでは不十分で，技術的に消せるようになる必要があります.

諸概念２：説明に基づく同意

個人情報・データの提供を求める際は，「説明に基づく同意（informed consent)」に基づいて，提供を求める必要があります．

- どのようなデータを提供してもらうのか
- 何に使うか，いつまで使うか
- 誰と共有するか
- データを提供することのメリット・デメリットは何か

（参考）インターネット，あるいはデータサイエンスの黎明期は，様々なデータが「勝手に集まって」いましたが，いまはそのような考え方は許されなくなっています．
　例：ウェブの閲覧履歴

（参考）「説明に基づく同意」に基づいて集めたデータを，当初宣言した目的以外の目的に使用する場合は，同意をとり直す必要があります．

注意：研究に使う画像・データを友人，研究室仲間からもらう場合も同様です．必ず「説明に基づく同意」書をとります（→研究倫理審査）．

諸概念３：オプトイン・オプトアウト

何らかのサービス・手続き等に「参加することを希望する」，
あるいは反対に「参加しないことを希望する」ことを表明してもらう手続き.

- 「オプトイン」(opt in)　（英単語の“opt”＝「～を望む」）
  参加することを希望することを言います.
- 「オプトアウト」（opt out)
  参加しないことを希望する（すでに参加している状態から抜ける，あるいは黙っていると参加してしまうことを拒否する）ことを言います.

例：JR東日本による，SUICA利用データの販売

- SUICAサービスに加入した最初の状態では，各利用者のSUICA利用データが，「匿名化（→後述）」を施した状態で，外部企業に販売されます.
- これを希望しない利用者は，JR東日本のホームページから「オプトアウト」手続きをとると，販売データから取り除いてもらえます.

形成中の合意１：GDPR

“EU General Data Protection Regulation (GDPR)”
＝「欧州一般データ保護規則」

- EU内28カ国でバラバラであった個人情報保護関連規則を一元化したものです.
- それと同時に，ここまで述べてきたような最先端のデータ保護の考え方を具現化したものでもあり，最先端の考え方を表したものでもあります.
- 本教材と特に関係するところ：
  - 個人情報・データは誰のものか？
    第3-1-2節で挙げたものはすべて「データ主体」（＝そのデータを発生させた自然人）のもの，とみなされます.
  - 説明に基づく同意は，目的・期間を必要な範囲に限り，一般人に分かりやすく，いつでも同意を撤回できることが求められます.
  - 手続きは「オプトアウト」ではなく，「オプトイン」で設計することが求められます.
  - AI等により機械だけで不利な判定が出る場合は，異議申し立ての権利を確保することが求められます.

形成中の合意２：人間中心のAI社会原則

内閣府 統合イノベーション戦略推進会議 決定（2019年3月29日）

- 人間中心の原則
- プライバシー確保の原則
- セキュリティ確保の原則
- 公平性・説明責任・透明性の原則，など

いずれも，本教材で議論している内容を，国レベルで方針づけたもの.

類似のものは，他にも多数提案されています：

- 人工知能学会
  倫理指針（2017年2月28日）
  機械学習と公平性に関する声明（2019年12月10日）
- 人間中心の機械学習（https://www.fatml.org/）
- 総務省 AIネットワーク社会推進会議 報告書（各年）
- IEEE Ethically Aligned Design, first edition (2019年3月)，など

データサイエンスやAIの責任は誰が負うのか

データサイエンスやAIは，万能でも完全でもありません.

- 思ったよりできないこと，反対に，人間が想像する以上にできてしまうことがある（→匿名性が剥がれ機微データが流出する事故）
- 深層学習等，いわゆる「ブラックボックス」で，科学的に明解な説明ができていない技術もある.

責任者が分かりにくい：開発者，販売者，購入者，….
どのひとりも完全なる責任者にはなりにくいし，完全なる無関係者とも言えません.

それでは，技術の提供者は，何をどこまで行うべきでしょうか？

- Accountability（アカウンタビリティ）：技術や商品そのものに加え，誰が責任を負うのかまで説明できること.
（ただしこれは現状では難しいです.）
- Trust（トラスト）：（アカウンタビリティを必ずしも完全に果たせない場合に）過去の類似例を示して，現状の技術や商品の妥当性，公平性，正当性に納得してもらう方法.

3-2. データを守る上での留意事項

概要.

- この節では，前節までに述べた個人情報・データを中心に，
  それらをどのように守るかについて理解します.

- より具体的には，下記について学びます.
  - ✓ データの守り方
  - ✓ 悪意のある攻撃と既に起こった事例

3-2-1. データの守り方

守り方１：情報管理三原則

外部の脅威から守りながら，情報（データ）をうまく活用するための
３つの原則を「情報管理三原則（情報セキュリティ三原則）」と呼びます.

- 機密性
  情報（データ）に，正当な権限を持つ者だけがアクセスできること.
  このためには情報の適切な分類，アクセス権限の設定，情報の暗号化
  などが必要です.
- 完全性・整合性
  情報が完全であり，誤りのない状態であること.
  不正なアクセスの検知，誤り検出などの技術が必要です.
- 可用性
  （正当な権限を持った者が）情報に適切にアクセスできること.
  可用性を阻害する攻撃（DoS攻撃など）への防御,
  ハードウェア障害のための冗長保存性などが必要です.

データの守り方２：匿名化（1/2）

データから「個人ID」（→第3-1-2節）を適切に削除して，
データを扱いながらも当該個人の情報であることが分からないように処理することを「匿名化」と言います．

これまでよく行われてきた匿名化：

1. 連結可能匿名化：仮IDを発行して個人IDを隠す方法．

No.1, 身長170cm, 体重60kg；	No.1＝山田太郎
No.2, 身長160cm, 体重45kg；	No.2＝佐藤花子．．．

↑「連結表」と呼ばれる

データ処理は左の表（連結可能匿名化を施したデータの表）を用い，
個人特定が必要なときに限り，別途厳重保存した「連結表」を使う．

2. 連結不可能匿名化：そもそも連結表を持たない方法．
連結表が存在すると危ないほどの機微データの場合，
連結表を持たない，あるいはさらに仮IDすら持たないデータのみを扱う．

しかしながら，これで「本当に個人特定は不可能なのか」には慎重な議論が必要で，上記の匿名化の考え方は，2015年の個人情報保護法改正で破棄されています．

データの守り方２：匿名化（2/2）

匿名化は本当に可能でしょうか？

- 匿名化が剥がれた例：AOL検索履歴公開案件
  - 2006年8月4日，AOL（米インターネットサービス）は，65万人の3ヶ月に渡る検索履歴（2,000万キーワード）を研究目的で公開した．
  - データには仮IDしか付けられていなかったが，検索データから，一部ユーザが特定され問題になった．
  - AOLは8月7日にデータを取り下げたが，すでにネット上に広まり手遅れであった．
  - 同9月，訴訟が起こされた．
- 社会学者 Sweeneyの研究（2000年）：
  ZIPコード（郵便番号）＋生年月日＋性別 で，アメリカ人の87%が数学的にひとりまで絞り込み可能

データは，組み合わせることで匿名性が剥がれ，機微データが漏れ出すことがあります．データサイエンスの進化に伴いこの危険性は上昇します．データは，使う人も，提供する人も，これを前提にする必要があります．

データの守り方３：暗号化とパスワード

- 暗号化：
  元のデータに対して，特別な処理を施して，そのままでは読めない特殊なデータに変換すること.
  元に戻すことを「復号化」と言います.
  暗号化には様々な種類があり，「安全性（強度）」，「処理速度」が異なります.
  電話やインターネットの通信などは，原則として暗号化され，通信内容が保護されるようになっています.

- パスワード：
  データやサービスにアクセス権限を持つ人間であることを証明するための文字列.
  銀行のキャッシュカードやクレジットカードのPINコードを始め，日常で幅広く用いられていますが，短いパスワード，単純なパスワードは簡単に見破られるため注意が必要です.

3-2-2. 悪意のある攻撃と既に起こった事例

データは，暗号化やパスワードで原則として保護されますが，
それでも，過失や悪意を持った人の攻撃で，データが漏洩したり，
それによってプライバシーが侵害されることがあり，注意が必要です．

以下，いくつか類例を挙げます．

- データの持ち出し，あるいは紛失による流出
- 攻撃による情報漏洩
- スパイウェア，マルウェアによる情報搾取

データの持ち出し・紛失による流出

厳重に保管しているはずのデータでも，内部の人間による（悪意のある）持ち出し，あるいは正当な権限者が持ち出した際の過失による紛失で，データは漏洩しえます.

実例（持ち出し）：
- 市議会に立候補した市職員が，職務上の地位を利用して市民個人情報（住所等）を持ち出し，選挙応援依頼書を発送（平塚市，2020年）
- 県庁で使用していたパソコンのハードディスクの処分を依頼された企業の職員が，それを持ち出してオークションサイトで売却；ハードディスクには各種納税記録なども入っていた（神奈川県，2019年）

実例（紛失による流出）：
- 学校等において，成績データなどが入ったUSBメモリ，パソコンなどを紛失（複数案件あり）
- メール誤送信によるメールアドレス流出（複数案件あり）

（注意）メールアドレスも個人情報です.
本人に断りなく他人に暴露してはいけません.
（複数人に同時にメールを発送する際注意）

攻撃による流出

内部の人間は気をつけていても，外部からの（しばしばインターネット経由での）攻撃により，情報は漏洩しえます．

実例：

- 化粧品会社の決済サーバーの脆弱性（セキュリティ上の「穴」が開いていること）を攻撃され，購入者のクレジットカード情報が流出（日本，2020年3月；他，類似案件多数あり）
- QRコード決済システムのひとつで不正利用が発覚，多数のユーザの電子マネーが不正に使用された．後に，認証システムの設計が甘く，攻撃者が容易に他人のアカウントを乗っ取れることが明らかになった．そのQRコード決済システムはサービスを終了．（日本，2019年）

スパイウェア，マルウェアによる情報搾取

通常のアプリ，あるいはシステムの一部のように見せかけてパソコン等に取り込ませ，ユーザに分からないように内部情報・データを外部送信するプログラムを「スパイウェア」と呼びます．
「マルウェア」(“malicious”＝「攻撃的な」から来た名称）とも呼びますが，この場合，「ウィルス」（パソコン等に入り込んで動作に障害を起こすもの）なども含み，悪意あるソフトウェア全体も指します．

実例：

- Androidスマホに入り込み写真・動画・音声記録・連絡先等の情報にアクセスして外部送信しうる“Exodus”が発見され，その後iOS版も発見された（世界的，2019年）※現在は対処済み
- 仮想通貨取引所が攻撃され，ある仮想通貨580億円相当（当時レートのもの）が盗まれた．取引所内のパソコンがマルウェアに汚染され，情報搾取・外部からの悪意を持ったアクセスを許した，とされている（日本，2018年）

