



# クラウドとデータ流通・処理

日本マイクロソフト株式会社

# 様々なデータストア・処理プラットフォーム



On-Premises  
*Infrastructure Platform*



IaaS  
*Infrastructure Platform*



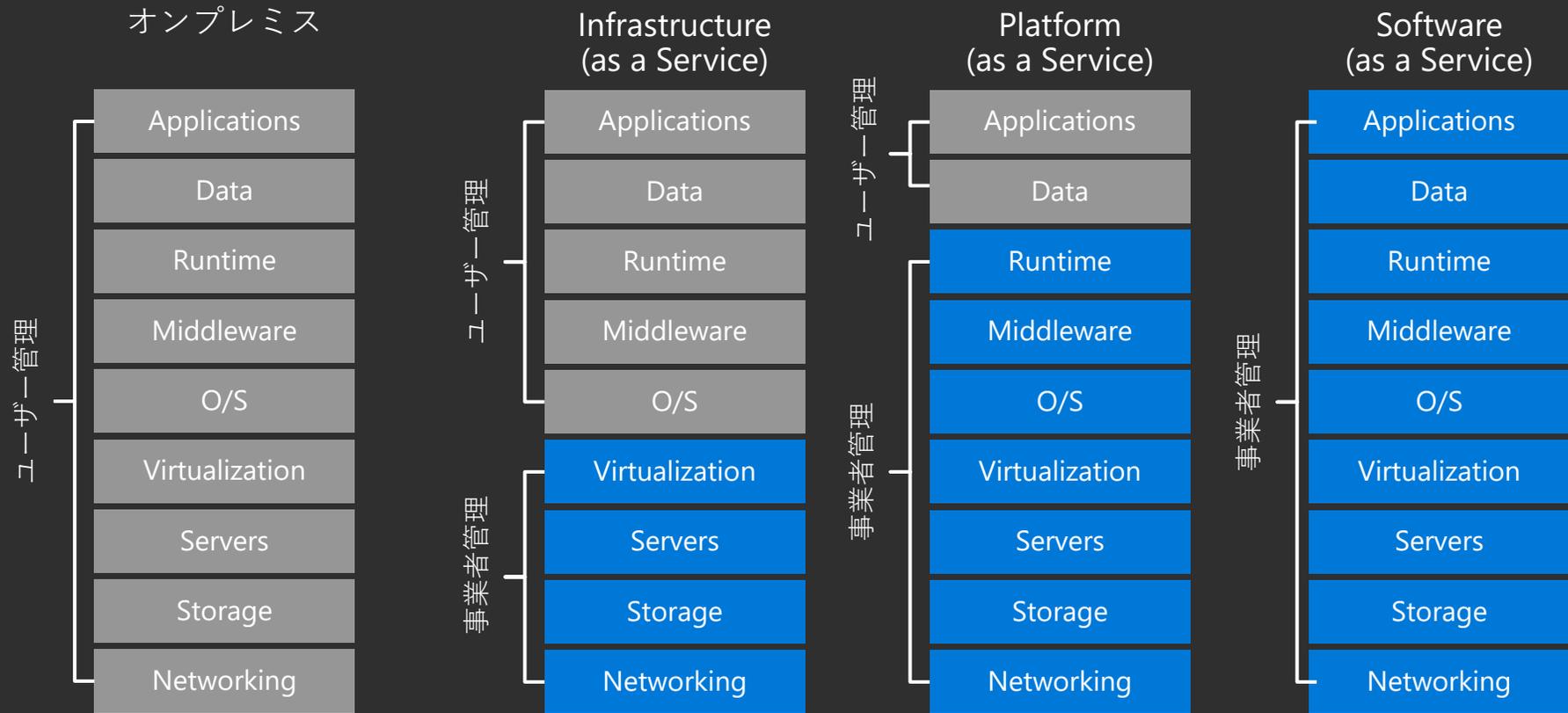
Containers  
*Managed  
Container Platform*



PaaS/Serverless  
*Application Platform*



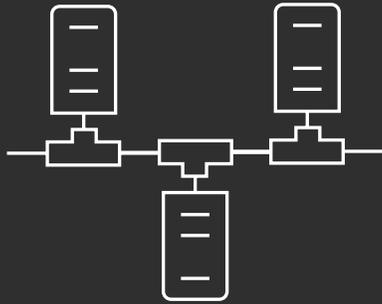
SaaS



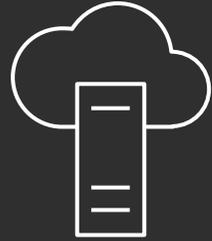
ユーザー管理

クラウド事業者管理

# クラウドで使用するサービス群（例）



異種混在の実行  
環境



ハイブリッド

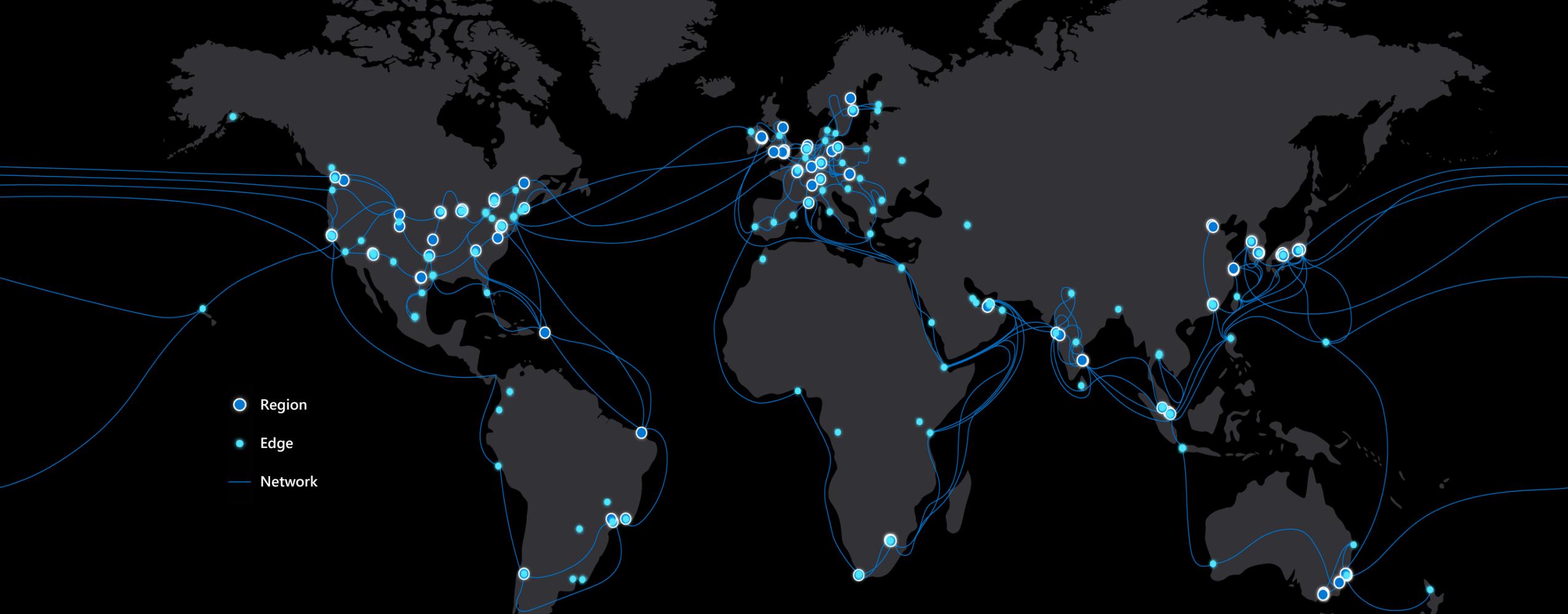


セキュリティ &  
マネージメント



監視及びコスト管理

# グローバルに稼働するデータセンター



66 Azure regions

175k+ miles of fiber + subsea cables

180+ edge sites

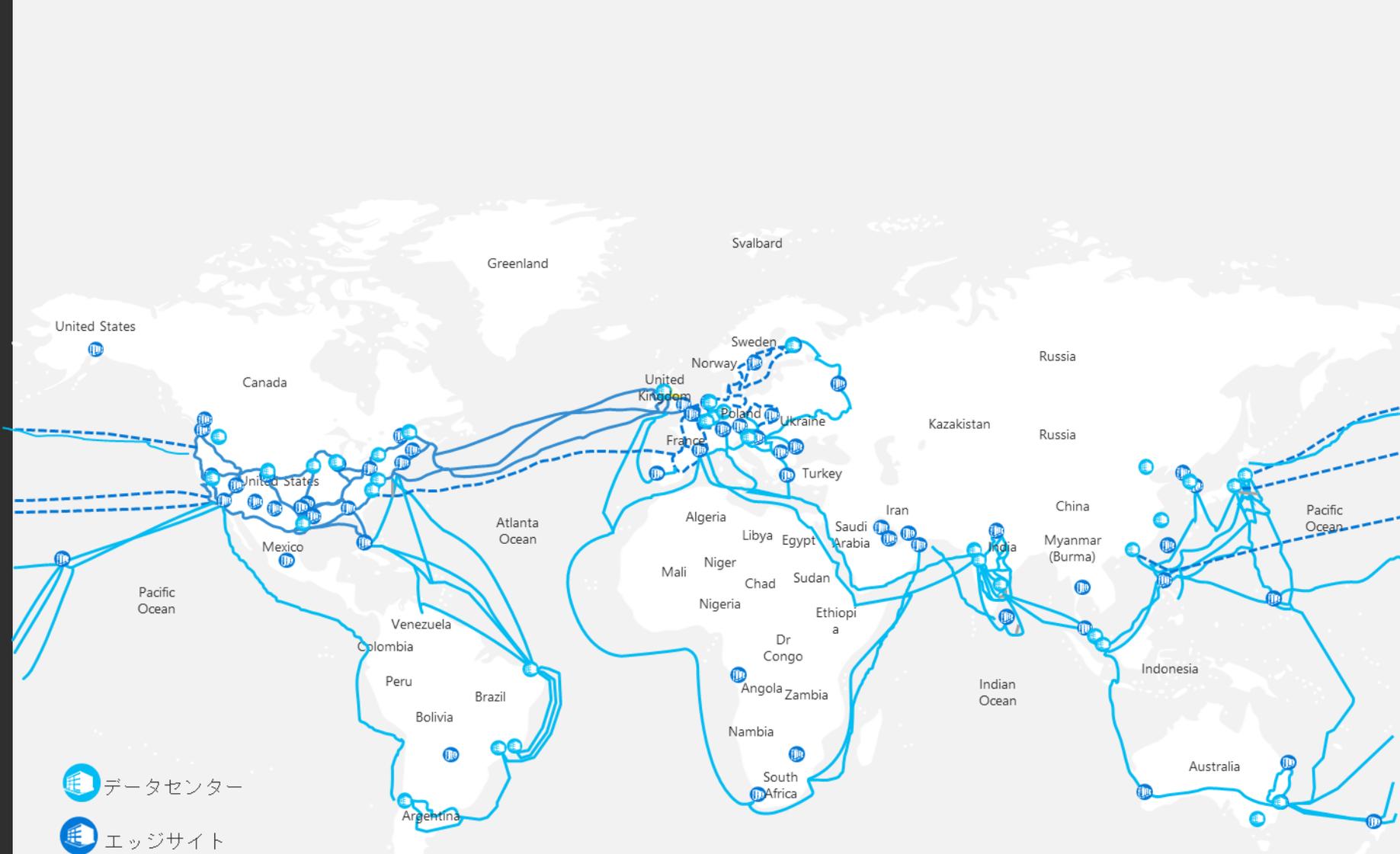
500+ network partners

20k+ peering connections

# 世界中の リソースを 一体的に運用

## 世界最大規模の 民間ネットワーク網

- 8,000+ ISP セッション
- 130+ エッジサイト
- 44 専用線拠点
- 53,000Kmにおよぶ光網
- SDN (SWAN, OLS)



DCおよびネットワークサイトは、全てを網羅したものではありません。

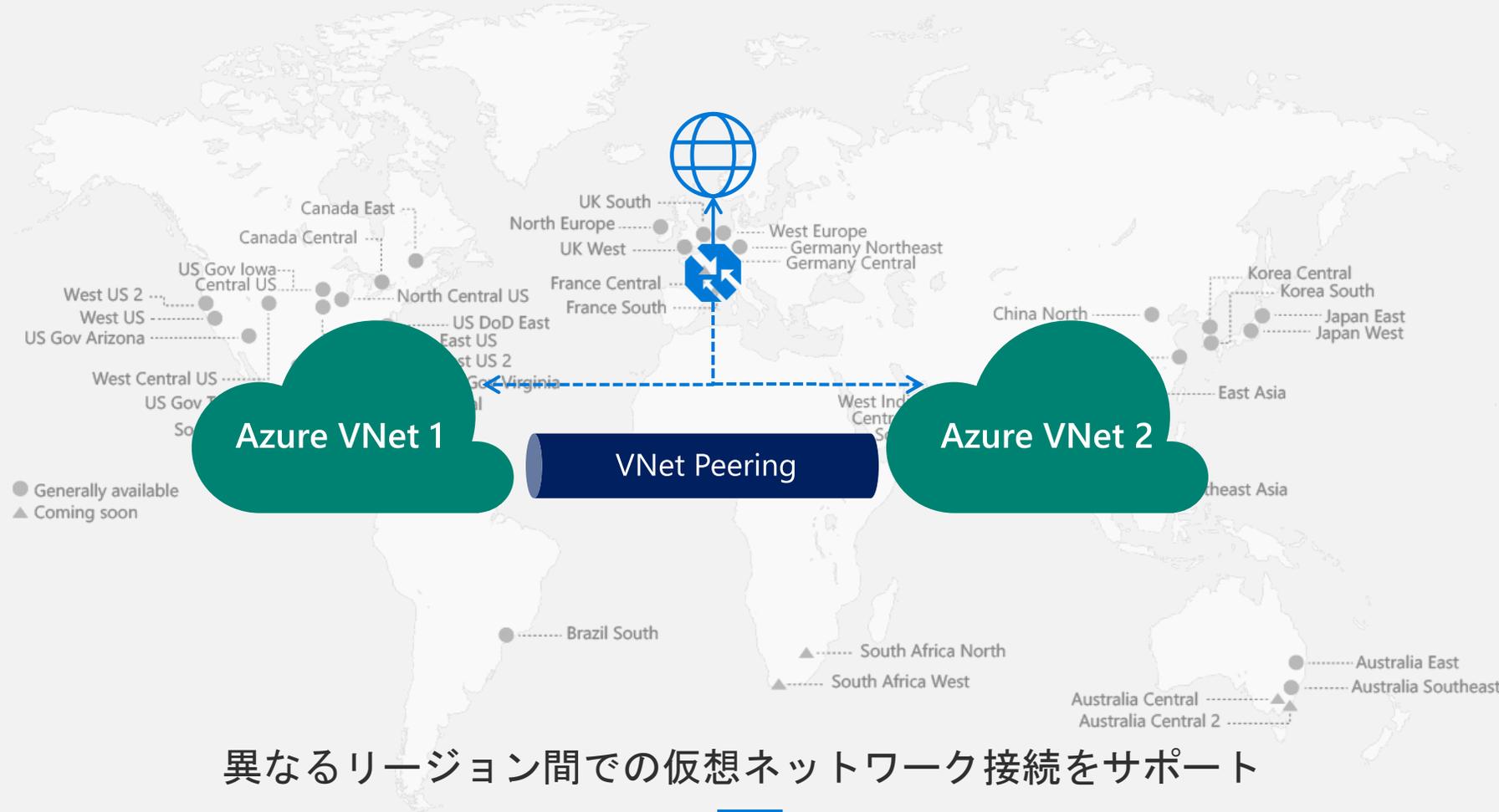
# グローバル接続



From on-premises



Within Azure



異なるリージョン間での仮想ネットワーク接続をサポート



国・地域をまたいだ環境の構築が可能

# コンフィデンシャル コンピューティング

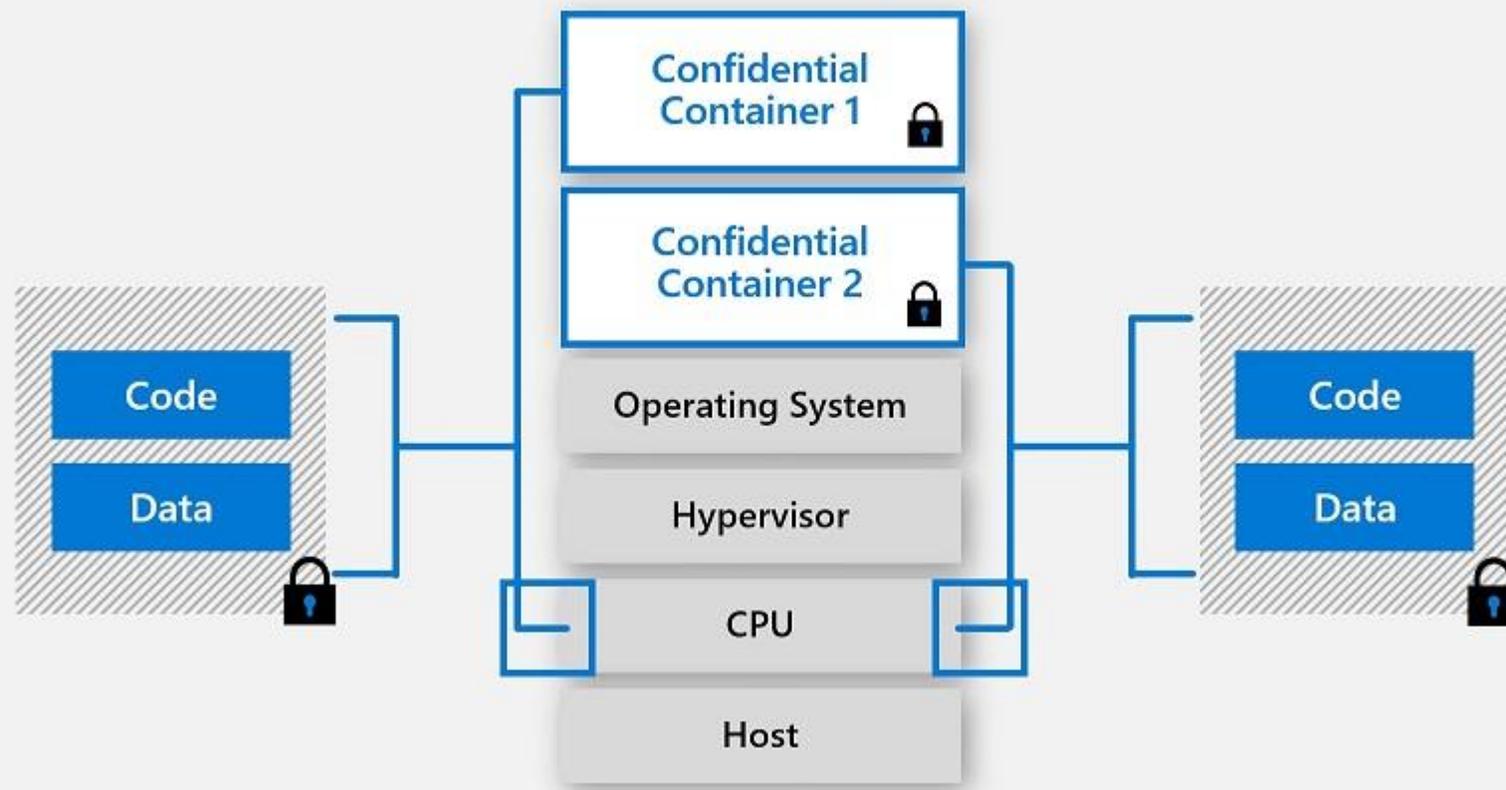


BYOKと暗号化



Within Azure

## AKS Confidential Compute Node



# 国際化するデータとオーバーライド及び知財法の違い



AI・機械学習によりデータが持つ価値がこれまでと大きく異なっている。



把握が困難になるデータと処理の所在地（国等）



偏りのある言語資源

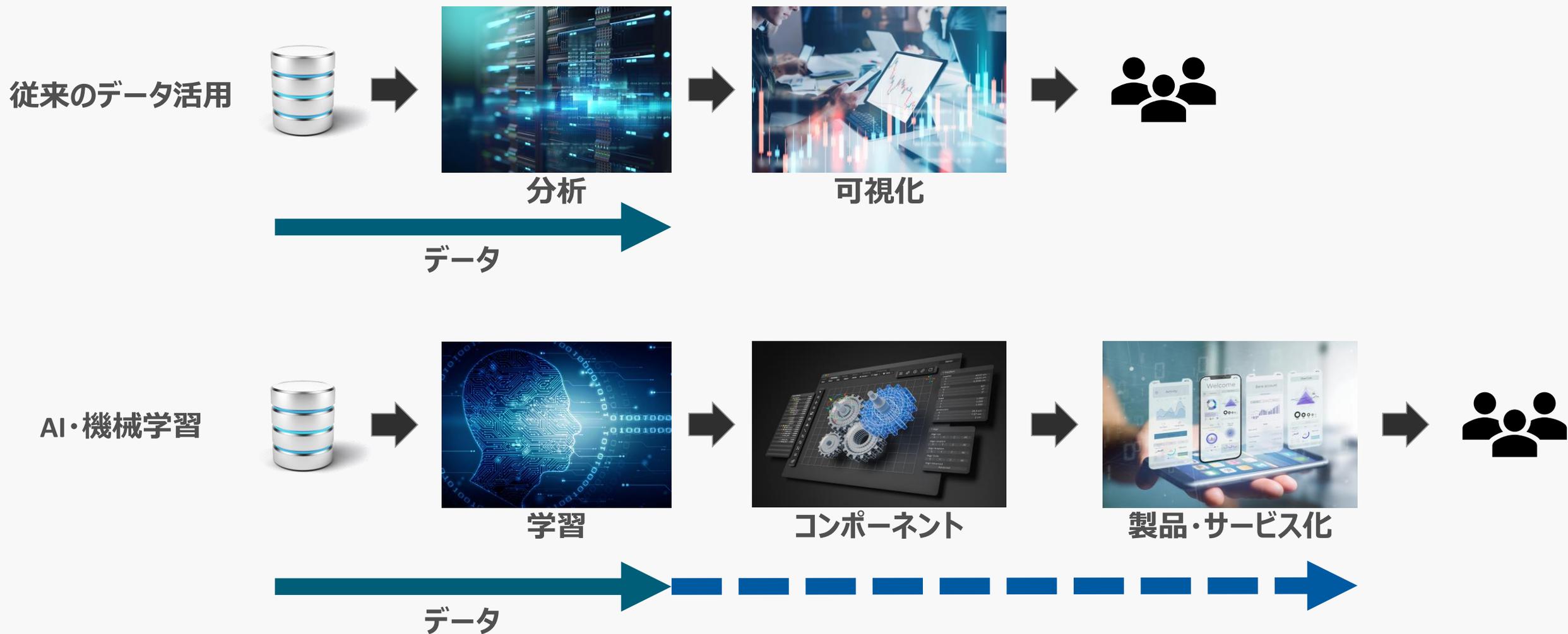


データ処理の自動化と利用規約と責任

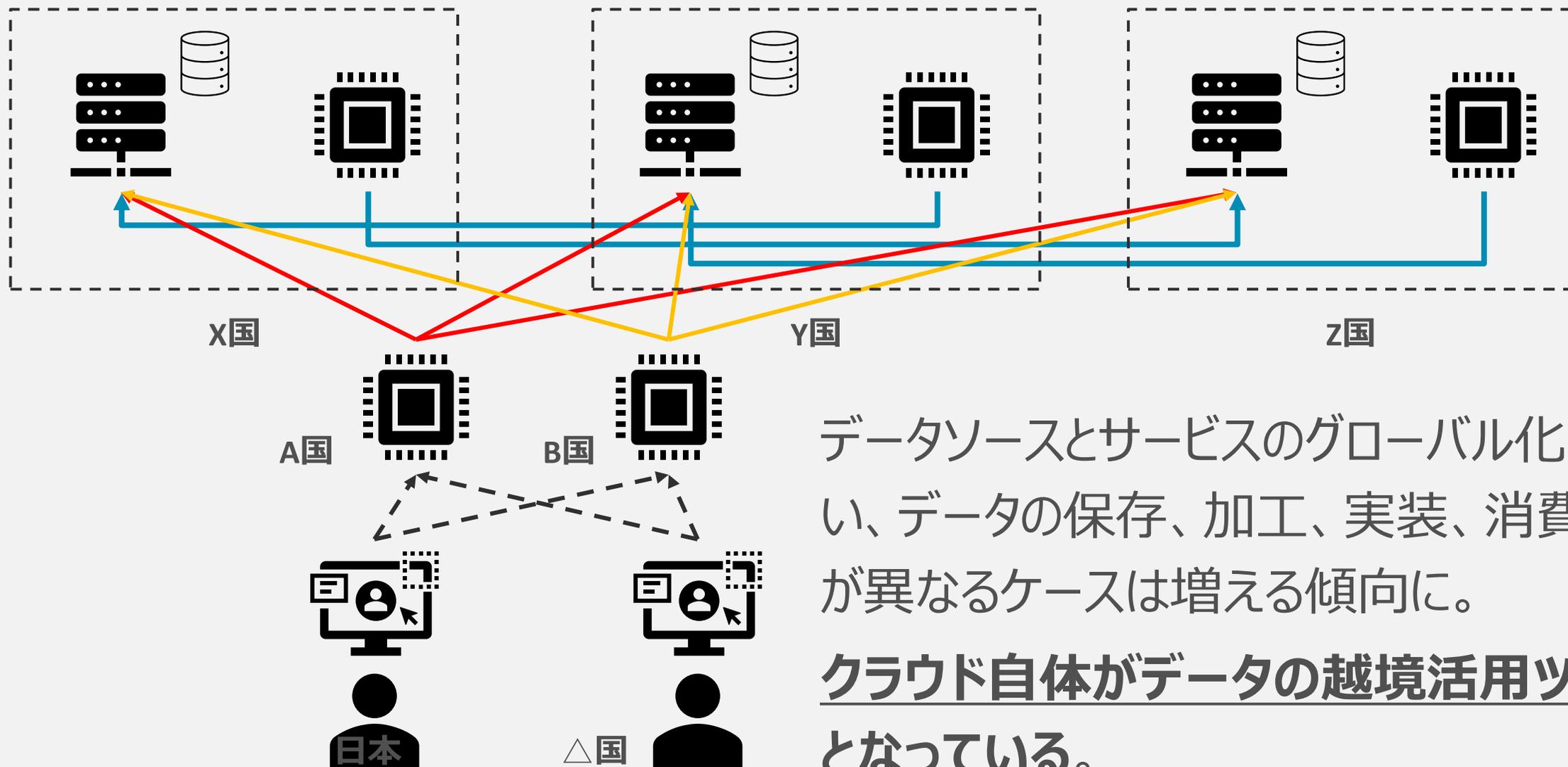


国毎に異なるデータに対する考え方（法律）

# データが持つ価値の変化



# 国・地域を跨るデータ処理とサービス



データソースとサービスのグローバル化に伴い、データの保存、加工、実装、消費国が異なるケースは増える傾向に。

クラウド自体がデータの越境活用ツール  
となっている。

# 各サービスに見る言語資源の偏り



## 登録コーパス数

- 英語 545
- 中国語 215
- 日本語 32

<https://www ldc.upenn.edu/>



## 登録対訳コーパス言語数

- 英語 48
- 中国語 18
- 日本語 13

<https://www.taus.net/>



ASPEC ( Asian Scientific Paper Excerpt Corpus )

- 自然言語 : ヘルスケアなど活用が進むデータの多くは自然言語
- 数値データ : GAIA-Xなどインダストリー中心の領域

# データ持ち出しにかかる規制と共有に関する制度の違い

- Plain Languageが貢献するAI・機械学習の活用
- US = Plain Writing Act \*E.O. 12866, E.O. 12988, and E.O. 13563.

## プレインランゲージ (Plain Language) とは

「速く」「効率的」「理解しやすい」伝達法

プレインランゲージ (プレインイングリッシュ、プレインジャパニーズなど) とは、「速く」読め、「効率的」にポイントを見つけることができ、そのポイントを「理解しやすい」、一言でいえば「読みやすい」文章を書くためのガイドを体系的にまとめた情報伝達術です。読みやすい文書は、読者のストレスを軽減します。イメージとしては、会話と同じリズムでスムーズに読め、内容がダイレクトに頭に入る次のような文章です。

- visually inviting まず読んでみようと思わせるレイアウトになっている
- logically organized 読者に必要な情報が論理的にわかりやすく整理されている
- understandable on the first reading 一読して理解できる
- in simple word 一般的に使用する平易な言葉で表現されている

# インターネットにおけるデータ活用と課題



サービス（データ）は  
どこでホストされているのか



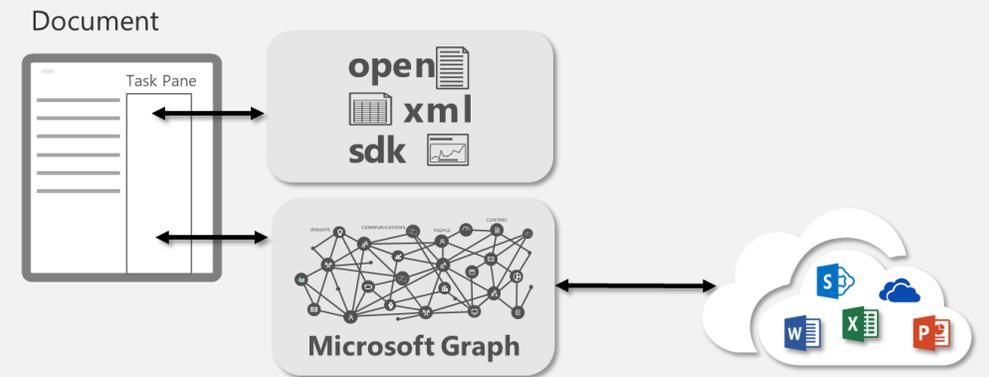
利用規約は  
マシンリーダブルなのか



利用規約の  
上書き（更新）問題

- インターネット上で収集を含むデータの利活用活発化しているが、「利用規約」（Terms of Use）を識別する合意された手法が存在しない。

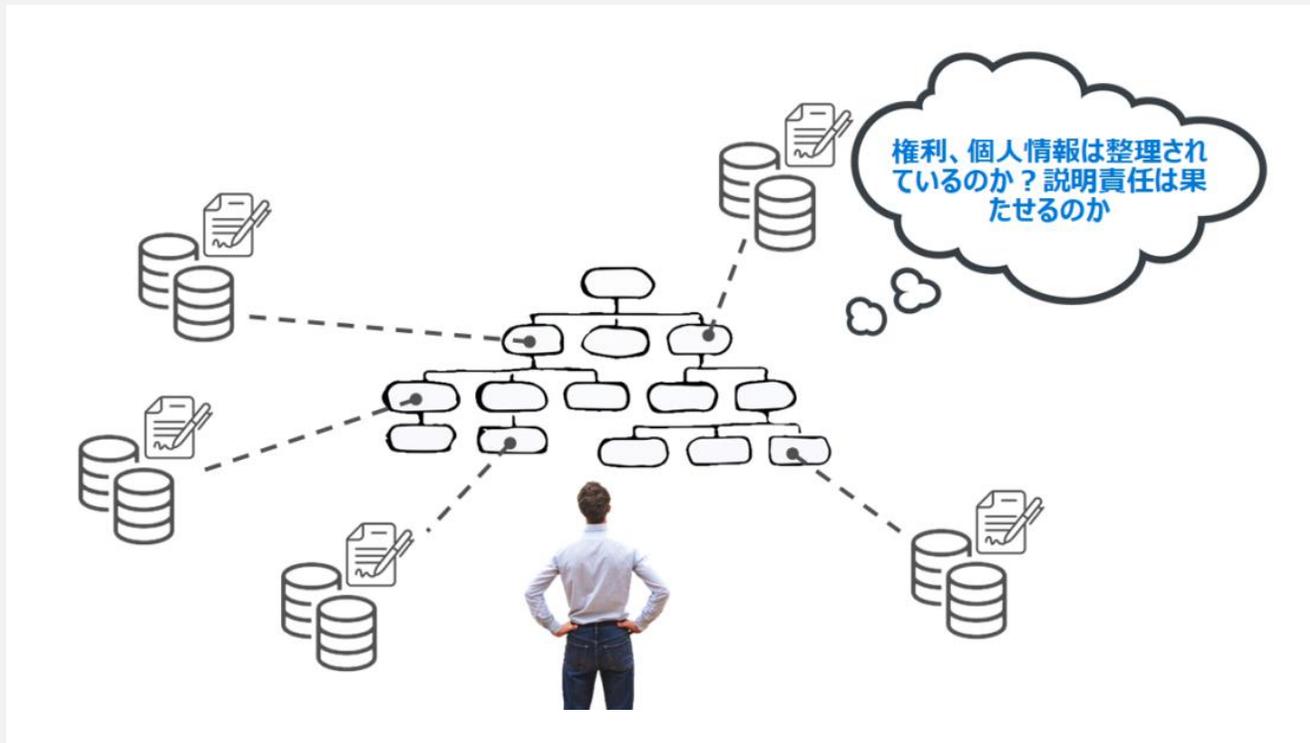
- 求められるHTMLを含む各種ドキュメントフォーマットの拡張と標準化



マイクロソフトでは、使用するフォーマットとのオープン化により、誰もが標準拡張可能にしている。

# 説明責任を果たせるデータとは

- ・ 情報システム及びその利用において、これまでデータの来歴管理は必要とされていない（保証・知的財産）
- ・ 研究・開発現場の多くでデータは属人的に管理されている。



**把握が困難なデータの来歴**

# 国により異なるデータの活用範囲

	英国	EU	スイス	日本
非営利・学術研究 目的	OK	OK	OK	OK
商業的 目的		OK (ただしオプトアウト可)	OK	OK
頒布・公衆送信				OK
違法収集コンテンツ				OK

# データ活用とデータ所在地のグローバル化

- データとサービスのグローバル化による様々な課題
  - ・ アクセス元とデータのレジデンシーが一致しないだけでなく、複数の国と地域に跨ることが珍しくなくなっている。
  - ・ データがサービスの一部として利用規約に基づいて提供されている場合がある。
  - ・ 国・地域により、著作権に対する考え方、法体系が異なる。
- 利用規約とドキュメントフォーマットの課題
  - ・ 利用規約等を合意された手段取得することができない。
  - ・ HTMLタグ、ドキュメントフォーマットの拡張と標準化。
- データの来歴管理が想定されていない
  - ・ 既存システムの多くがデータの来歴管理・記録を想定していない。

技術の進歩と制度のギャップが広がり、且つビジネスが1つの国にとどまらない。グローバルにビジネスを展開する企業はTDMをはじめとする諸問題に対してどのような対応をすればよいのか等が課題。

# Appendix

# 著30条の4第2号

著作物は、次に掲げる場合その他の当該著作物に表現された思想又は感情を自ら享受し又は他人に享受させることを目的としない場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

<略>

二 **情報解析**（多数の著作物その他の大量の情報から、当該情報を構成する言語、音、映像その他の要素に係る情報を抽出し、比較、分類その他の解析を行うことをいう。…）の用に供する場合

# イギリス法 (2014改正)

1. The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—
  - (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work **for the sole purpose of research for a non-commercial purpose**, and
  - (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).
- (2) Where a copy of a work has been made under this section, copyright in the work is infringed if—
  - (a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or
  - (b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.
- (3) If a copy made under this section is subsequently dealt with—
  - (a) it is to be treated as an infringing copy for the purposes of that dealing, and
  - (b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.
- (4) In subsection (3) “dealt with” means sold or let for hire, or offered or exposed for sale or hire.
- (5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.”.

# 欧洲DSM指令 (2019)

Art.2 [Definitions] For the purposes of this Directive, the following definitions apply:

(1) 'research organisation' means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research:

(a) on a **not-for-profit basis** or by reinvesting all the profits in its scientific research; or

(b) pursuant to a public interest mission recognised by a Member State;

in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation;

(2) 'text and data mining' means any **automated** analytical technique aimed at analysing text and data **in digital form** in order to generate information which includes but is not limited to patterns, trends and correlations;

(3) 'cultural heritage institution' means a publicly accessible library or museum, an archive or a film or audio heritage institution;

# ドイツ法 (2021改正)

§60d [Text und Data Mining für Zwecke der **wissenschaftlichen Forschung**]

(1) Vervielfältigungen für Text und Data Mining (§44b Absatz 1 und 2 Satz 1) sind für Zwecke der wissenschaftlichen Forschung nach Maßgabe der nachfolgenden Bestimmungen zulässig.

(2) Zu Vervielfältigungen berechtigt sind Forschungsorganisationen. Forschungsorganisationen sind Hochschulen, Forschungsinstitute oder sonstige Einrichtungen, die wissenschaftliche Forschung betreiben, sofern sie

1. **nicht kommerzielle Zwecke** verfolgen,
2. sämtliche Gewinne in die wissenschaftliche Forschung reinvestieren oder
3. im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind.

Nicht nach Satz 1 berechtigt sind Forschungsorganisationen, die mit einem privaten Unternehmen zusammenarbeiten, das einen bestimmenden Einfluss auf die Forschungsorganisation und einen bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung hat.

(3) Zu Vervielfältigungen berechtigt sind ferner

1. Bibliotheken und Museen, sofern sie öffentlich zugänglich sind, sowie Archive und Einrichtungen im Bereich des Film- oder Tonerbes (Kulturerbe-Einrichtungen),
2. einzelne Forscher, sofern sie nicht kommerzielle Zwecke verfolgen.

(4) Berechtigte nach den Absätzen 2 und 3, die nicht kommerzielle Zwecke verfolgen, dürfen Vervielfältigungen nach Absatz 1 folgenden Personen öffentlich zugänglich machen:

1. einem bestimmt abgegrenzten Kreis von Personen für deren gemeinsame wissenschaftliche Forschung sowie
2. einzelnen Dritten zur Überprüfung der Qualität wissenschaftlicher Forschung.

Sobald die gemeinsame wissenschaftliche Forschung oder die Überprüfung der Qualität wissenschaftlicher Forschung abgeschlossen ist, ist die öffentliche Zugänglichmachung zu beenden.

(5) Berechtigte nach den Absätzen 2 und 3 Nummer 1 dürfen Vervielfältigungen nach Absatz 1 mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahren, solange sie für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind.

(6) Rechtsinhaber sind befugt, erforderliche Maßnahmen zu ergreifen, um zu verhindern, dass die Sicherheit und Integrität ihrer Netze und Datenbanken durch Vervielfältigungen nach Absatz 1 gefährdet werden.

# スイス法 (2019)

Art.24d [Verwendung von Werken zum Zweck der wissenschaftlichen Forschung]

1. Zum Zweck der **wissenschaftlichen Forschung** ist es zulässig, ein Werk zu vervielfältigen, wenn die Vervielfältigung durch die Anwendung eines technischen Verfahrens bedingt ist und zu den zu vervielfältigenden Werken ein **rechtmässiger Zugang** besteht.
2. Die im Rahmen dieses Artikels angefertigten Vervielfältigungen dürfen nach Abschluss der wissenschaftlichen Forschung zu Archivierungs- und Sicherungs- zwecken aufbewahrt werden.
3. Dieser Artikel gilt nicht für die Vervielfältigung von Computerprogrammen.