

本文書は、日本企業の対中投資の参考に供するために、シティユーワ法律事務所（以下「当事務所」）が作成し、PDF ファイル形式で公開したものです。本文書に関し発生する著作権は当事務所に帰属しますが、ヘッダーを含め本文書の内容及び PDF ファイルのデータを改変せずに配布又は印刷される場合には、当事務所の承諾は不要です。それ以外の場合には事前に当事務所にご相談下さい。

付属文書

データ分野常用名詞解釈（第一弾）

（データ分野名詞解釈起草専門家チームにより 2024年12月30日発布）

1. **データ**とは、電子的又はその他の方式による、情報に対する一切の記録をいう。データは、視点の違いによって、生データ、派生データ、データ資源、データプロダクト及びサービス、データ資産、データ要素等と呼ばれる。
2. **生データ**とは、最初に生成され、又は発生源から収集された、加工処理を経していないデータをいう。
3. **データ資源**とは、価値創造のポテンシャルを有するデータの総称をいい、通常は、電子化された形式にて記録及び保存され、機械判読可能であり、かつ、社会化再利用に供することができるデータの集合を指す。
4. **データ要素**とは、生産経営活動に投じられ、かつ、価値創造に関与するデータ資源をいう。
5. **データプロダクト及びサービス**とは、データ加工を踏まえて形成される、特定のニーズを満たすことができるデータ加工品及びデータサービスをいう。
6. **データ資産**とは、特定の主体が適法に保有又は支配する、貨幣的測定を行うことができ、かつ、経済的利益又は社会的便益をもたらすことができるデータ資源をいう。
7. **データ要素の市場化配分**とは、市場メカニズムを通じ、データという新しいタイプの生産要素を配分していくことをいい、よりオープンで安全かつ高効率なデータ流通環境を確立し、データ要素の価値を絶えず引き出していくことを目的としている。
8. **データ処理**には、データの収集、保存、使用、加工、伝送、提供、公開等が含まれる。
9. **データ処理者**とは、データ処理活動において、処理目的及び処理方式を自主的に決定する個人又は組織をいう。
10. **受託データ処理者**とは、他人の委託を受けてデータを処理する個人又は組織をいう。
11. **データ流通**とは、別々の主体の間においてデータが行き来する過程をいい、データのオープン化、共有、取引、交換等が含まれる。
12. **データ取引**とは、データの供給者と需要者との間で行われる、特定の形態のデータを目的物とし、貨幣又はその他の等価物を対価とした取引行為をいう。
13. **データガバナンス**とは、データの品質、セキュリティ及びコンプライアンスを向上させ、データの有効利用を後押しする過程をいい、組織データガバナンス、業界データガバナンス、社会データガバナンス等が含まれる。
14. **データセキュリティ**とは、必要な措置を講じることを通じて、データが有効な保護及び適法な利用の状態にあり、並びに安全な状態の持続を保障する能力を具備しているよう確保することをいう。
15. **公共データ**とは、各級の党・政府機関及び企業・事業単位の法による職務履行又は公共

- サービス提供の過程において生成されるデータをいう。
16. **デジタル産業化**とは、モバイル通信、人工知能等のデジタル技術がデジタルプロダクト・デジタルサービスに転化し、データが資源・要素に転化して、デジタル新産業・新業態・新モデルが形成される過程をいう。
 17. **産業デジタル化**とは、従来型の農業、工業、サービス業等の産業が、デジタル技術の応用、データの取得・融合及びデータ資源価値の発掘を通じて業務運営効率をアップさせ、生産経営コストを引き下げ、ひいては思考・認知を再構築し、組織管理モデルを丸ごと作り変え、生産運営フローをシステムティックに変革し、全要素の生産性を絶えず向上させていく過程をいう。
 18. **デジタル経済の質の高い発展**とは、新たな質の生産力の育成加速をめぐって、データ要素の市場化配分改革を主軸としながら、データ基礎制度及びデジタルインフラの協調的な完全化、デジタル技術と実体経済の深い融合の全面的な推進、並びにデジタル経済のガバナンス能力及び国際協力水準の持続的な引上げを通じ、より強く・より良く・より大きくという目標を実現させていくデジタル経済発展の新段階をいう。
 19. **デジタル消費**とは、デジタル技術又はアプリケーションサポートによって形成される消費活動及び消費方式をいい、デジタル化・インテリジェント化された技術、プロダクト及びサービスに対する消費も、消費内容・消費ルート・消費環境のデジタル化及びインテリジェント化も含まれるほか、オンラインとオフラインが深く融合した消費の新態様も含んでいる。
 20. **インダストリアルインターネット**とは、デジタル技術及びデータ要素を利用して全産業チェーンのデータ統合を推し進め、産業のデジタル化・ネットワーク化・インテリジェント化発展に力を与え、業務プロセス、組織機構、生産方式等の再編変革を後押しして、産業チェーンにおける川上・川下の協調的モデルチェンジ、オンライン・オフラインの融合発展、並びに全産業のコスト削減・効率アップ及び質の高い発展を実現させ、ひいては新たな産業連携、資源配分及び価値創造の体系を作り出すものいう。
 21. **都市全域デジタル・トランスフォーメーション**とは、都市において、データの統合及び開発利用の全面的な深化を主軸としながら、デジタル技術及び制度革新ツールを総合的に利用して、技術アーキテクチャの一新、都市管理フローの変革及び産業と都市の深い融合を実現させ、デジタル・トランスフォーメーションの全領域における効率アップ、サポート能力の全方位における増強、トランスフォーメーションに係るエコシステムの全過程における最適化を促進する、都市の質の高い発展新モデルをいう。
 22. **「東数西算」プロジェクト**は、東部地区の経済活動で生成されたデータ及びニーズを西部地区に移して計算及び処理を行うもので、データセンターに対しレイアウト、ネットワーク、電力、エネルギー消費、コンピューティングパワー、データ等の方面において統一的計画を行う重大プロジェクトである。例えば、AIモデルの訓練・推論、機械学習等の業務シーンで、「東数西算」の方式を通じ、東部の業務を、風力・太陽光・水力発電が豊富な西部のエリアに移転させ、東部と西部の協調的発展を実現することができる。「東数西算」プロジェクトの建設を急ピッチで推し進めることは、有効にデータ要素のイノベーションを活性化させ、デジタル産業化及び産業デジタル化の歩みを速め、新技術・新産業・新業態・新モデルを生み出し、経済の質の高い発展をサポートすることになる。

23. **高速データ網**とは、データ流通利用シーンに向けて、ネットワーク仮想化、ソフトウェア定義ネットワーク（SDN）等の技術を活用し、帯域可変型で安全かつ信頼性があり、高効率伝送のデータ伝送サービスを提供するものいう。
24. **全国一体化コンピューティングパワー網**とは、情報ネットワーク技術を介して、全国範囲内における各種コンピューティングパワー資源の高比率・大規模な一体化ディスプレイ・運用を促進するデジタルインフラをいう。「東数西算」プロジェクトのバージョン 2.0 として、集約化・一体化・協働化・価値化という 4 つの典型的特徴を有している。
25. **メタデータ**は、特定のデータを定義及び記述するデータであり、データの構造、特徴及び関係性に関する情報を提供し、データの組織化、検索、理解及び管理に役立つものである。
26. **構造化データ**とは、データ表示形式の 1 種をいい、この形式では、データエレメントが集まってできた全ての記録の構造がどれも一致し、かつ、関係モデルを使用して有効な記述をすることができる。
27. **半構造化データ**とは、関係データベース又はその他のデータテーブルの形式で関連付けられているデータモデル・構造に適合しないが、関連タグを含んでおり、それを用いてセマンティック要素を分離し、並びに記録及びフィールドに対し階層化が行われているというデータ化構造形式の 1 種をいう。
28. **非構造化データ**とは、事前定義モデルを有しない、又は事前定義方式で組織化されていないデータをいう。
29. **データ解析**とは、特定の技術及び方法を通じ、データに対して整理、研究、推論及び要約総括を行い、データの中から有用な情報を取り出し、規則性を見つけ出し、及び結論を形成する過程をいう。
30. **データマイニング**は、データ解析の手段の 1 種であり、統計分析、機械学習、パターン認識、エキスパートシステム等の技術を通じ、データ中に隠れている情報又は価値を発掘する過程である。
31. **データ可視化**とは、チャート、グラフィック、地図等の図形化手段を通じてデータ中に含まれる有用な情報を明瞭かつ有効に表現し、もってデータ使用者がデータをより良く理解及び分析することができるようにすることをいう。
32. **データウェアハウス**とは、データプレパレーションの後における、データの永続的な保存に用いられるデータベースをいう。
33. **データレイク**とは、高度にスケーラブルなデータストレージ・アーキテクチャの 1 種で、大量の生データ及び派生データ（これらのデータは、各種ソースに由来し、かつ、異なるフォーマットで存在することが可能であり、構造化、半構造化及び非構造化データが含まれる。）の保存用に特化したものをいう。
34. **レイクハウス**とは、新しいタイプのオープンなストレージ・アーキテクチャの 1 種で、データウェアハウスとデータレイクとを統合し、データウェアハウスの高い性能及び管理能力をデータレイクの柔軟性と融合させたものをいう。最下層では複数のデータ型の併存に対応してデータ間の相互共有の実現を可能にし、上層では統一的にカプセル化したインターフェイスを通じてアクセスを行うことができ、リアルタイムのクエリ及び分析に同時に対応可能となっている。

35. **プライバシー保護計算**とは、データ提供者は生データを明かさないとすることが担保されているという前提において、データに対し分析計算を行うという情報技術の1つであって、生成、保存、計算、応用、廃棄等データの流れ全過程の各段階において、データに関し「使えるが中身は見えない」ことを保障するものをいう。プライバシー保護計算でよく用いられる技術的ソリューションには、セキュアマルチパーティ計算、連合学習、TEE（信頼できる実行環境）、暗号化計算等がある。よく用いられる前提技術には、Garbled Circuit（秘匿回路）、紛失通信、秘密分散、準同型暗号等がある。
36. **セキュアマルチパーティ計算**とは、1つの分散型ネットワーク中において、複数の参加実体が秘密データをそれぞれ保有しながら、それらを入力値としてある関数についての計算を共同で完成させたいとしているところ、計算結果及び公開可能と見込まれる情報を除き、他の参加実体の入力情報は一切入手できないという要求が全ての参加実体に対して課されているものをいう。信頼できる第三者がいない場合に、複数参加者による共同での計算を安全に行うという問題について、主に研究されている。
37. **連合学習**とは、各自の生のプライベートデータがデータ当事者の定義する信頼可能な領域から出ないことが担保されているという前提において、プライバシーデータを保護した方式で中間計算結果を交換することにより、複数の参加者が、何らかの機械学習タスクを連携して完成させるという1種の手法をいう。
38. **TEE**とは、ハードウェアレベルでの隔離及びセキュアブートメカニズムに基づいて、セキュリティに係るセンシティブなアプリケーションに関連するデータ及びコードの機密性、完全性、真正性及び否認防止性を確保するという目標のために構築されたソフトウェア実行環境の1種をいう。
39. **暗号化計算**とは、暗号学、信頼できるハードウェア及びシステム・セキュリティ関連技術を総合的に利用することにより、計算過程のデータについて、使えるが中身は見えないという状態を実現し、計算結果については暗号化されたままにすることができ、もって組立てが複雑な組み合わせの計算に対応し、計算に係る経路全体についての保障を実現し、データの漏洩及び濫用を防止するものをいう。
40. **ブロックチェーン**は、分散型ネットワーク、暗号化技術、スマートコントラクト等複数の技術が組み合わさった新しいタイプのデータベース・ソフトウェアであり、多極分散化、コンセンサス・信頼性、改ざん不可能性、トレーサビリティ等の特性を有し、主としてデータ流通過程における信頼性及びセキュリティの問題の解決に用いられる。

（法令原文名称：データ領域常用名詞解釈（第一批））