広島文化学園短期大学紀要，Vol. 54，9−18，2021

データサイエンス教育に関するいくつかの提言

——データサイエンスとマイ・ティーチング・ポートフォリオの対比から——

古　川　博　仁*

Some Suggestions of the Data Processing

—— From the Contrast between the Data Science and My Teaching Portfolio

Hirohito FURUKAWA

Key words：AI　Artificial Intelligence，データサイエンス　Data Science，データサイエンティスト　Data Sicientist，
マイ・ティーチング・ポートフォリオ　My Teaching Portforio，ビッグデータ　Big Data，
自己点検　Self-Inspection，第三者評価　Third Party Eevaluation，PDCA　Plan-Do-Check-Action

1. は　じ　め　に

近い将来，職種の中で「事務職がAIに奪われる」という風評を耳にした。著者（私）は大学で教えているインテリア論の中で人間工学に触れている手前，19世紀の半ばの産業革命後，「大量生産で人間が機械に使われていた時代」から，20世紀後半のコンピュータの出現で制御技術が発達し，「人間にとって使い勝手のいい機械操作」が実現していることを強調してきたが，ここ数年，AI技術が急速に発展し，TVのCMで自動運転などを見るにつけ，今度はAIが人間にとって代わる時代が来た感が強まった。まさにAI社会の到来である。今度は「人間の仕事をAIが奪う時代」かと思いきや，平成31年２月，政府は「人間中心のAI社会原則」[1]を打ち出した。その基本理念は「人間の尊厳が尊重される社会」，「多様な背景を持つ人々が多様な幸せを追求できる社会」，「持続性のある社会」であり，これは情報社会（Society 4.0）に続いてAI，IoT（Internet of Things），ロボット等を駆使した社会の創生（Society 5.0）に向けての取り組みである。このような社会変革を「AI-Readyな社会」と表記し，社会全体がAIによる便益を最大限に享受するための変革の必要性と人間がAIの恩恵を実感できる社会の在り方を，「人」，「社会システム」，「産業構造」，「イノベーションシステム」，「ガバナンス」の５つの観点で打ち出している。この実現には各ステークホルダーが留意すべき２つの基本原則，「AI社会原則」，「AI開発利用原則」が重要であるとされている。

著者は，教育現場にいる手前，「AI社会原則（７原則）」，すなわち（1）人間中心の原則，（2）教育・リテラシー原則，（3）プライバシー確保の原則，（4）セキュリティ確保の原則，（5）公正競争確保の原則，（6）公平性，説明責任及び透明性の原則，（7）イノベーションの原則に関して，特に（2）教育・リテラシーの原則に関心を置くものである。ここでは，データサイエンスの素養を育成することがリテラシー教育の課題であることが掲げられている。その背景には，我が国のデータサイエンス教育が欧州や北米などのOECD諸国に比べて遅れていることが指摘されている[2,3]。これに対する取り組みとして「AI戦略　2019」が示され，小中高，および大学・高専等にレベル別の人材育成目標が掲げられた。そこでは，デジタル社会の基礎知識である「数理・データサイエンス・AI」に関する知識・技能をすべての国民が育み，2025年までに社会のあらゆる分野でこれらの基礎力を習得した人材が活躍することを目指しての具体目標が打ち出されている。まさに我が国は，データサイエンス教育の黎明期を迎えたと言える。

データサイエンスという術語はPeter Naurによる著書「Concise Survey of Computer Methods」（1974）の中で使われている。確たる定義がなされている訳ではないが，「A basic principle of data science is this: The data representation must be chosen with due regard to the transformation to be achieved and the data processing tools avail-

* 広島文化学園短期大学コミュニティ生活学科

able. This stresses the importance of concern for the characteristics of the data processing tools.」とある。

データサイエンスという学問領域は広範囲に及ぶ。この学問領域には大きく次の６つの関連が内蔵されていると著者は見なすことにする。すなわち，(1)．コンピュータサイエンス[5] との関連（アルゴリズム，プログラミング，計算可能性など），(2)．個人情報等倫理との関連（著作権，セキュリティ，プライバシー保護など），(3)．AI との関連（機械学習），(4)．IoT を含むビッグデータとの関連（インターネットから得られたデータの解析），(5)．データマイニングとの関連（統計解析，データベースなど），(6)．文系・理系の専門分野との関連（プレゼンテーション能力，コミュニケーション能力，ビジネススキル，実務能力を基調とした IT スキルなど）である。

これ等の関連を踏まえてデータサイエンスを一言で言うならば，「大規模なデータセット（データベース）をサイエンスして実用的な洞察を引き出すこと」であると言うことができる。

データサイエンスを専門とする職業人をデータサイエンティストと言うが，これから AI 技術革新が進展するにつれてその人材育成の需要は激増し，最先端を行くアメリカに於いてでさえ「2016年度25万人不足している[4]」（p.51）と言われている。

企業で働くデータサイエンティスト達の仕事内容は，その企業のビジネスコンセプトの理解，その目的を達成するために必要なデータセットの設定（データベースの構築），適切なデータ収集と分析，データ分析から洞察されるビジネスモデルの構築（AI を導入）とその評価方法，評価を踏まえた結果の査定と改善計画など多岐にわたる。これ等はプロジェクトで行われるが，データサイエンティスト達が主として関わる仕事は，ビジネスモデルを想定して得られたデータとその分析・評価を何度も繰り返すことによってビジネスモデルを修正し，企業の意思決定に役立つ洞察を得ることにある。

前述のデータサイエンスに内蔵され６つの関連とは，データサイエンティストの人材育成のために必要な素養を，教育・リテラシーの範疇で捉えた学問領域であると理解していただきたい。

本報告は，データサイエンス教育と著者がこれまでに教育現場で教えてきたデータに関わる授業科目（マイ・ティーチング・ポートフォリオ）・研究・大学業務に絞ってこれとの対比を俯瞰することで，今後どのようなデータサイエンス教育が可能なのかを著者なりに模索し，教育・リテラシーの範疇でのデータサイエンス教育に関して提言を行うものである。

2. データサイエンス教育に関する模索といくつかの提言

教育・リテラシーの範疇でデータサイエンス教育を模索する上で，1984年～2021年の38年間，著者が携わってきた授業科目（最長で10年程度）・研究・大学での業務を，次の A ～ E の５つのフェーズで俯瞰してみたい。

A. データサイエンスとマイ・ティーチング・ポートフォリオとの対比

データサイエンスが内蔵する６つの関連とマイ・ティチング・ポートフォリオとの対比を次に示す。

(1)． コンピュータサイエンスとの関連

情報処理概（オペレーティングシステム，ハードウェアーとソフトウェアー，ネットワークの仕組みと暗号化と復号化など）

アルゴリズム（フローチャート，データ構造，ソート，探索木）

プログラミング（C 言語３級程度，Java３級程度，Java スクリプト）

Web デザイン（タグの使い方（３級程度）とホームページへの仮想アップ）

情報活用演習Ⅰ（コンピュータの歴史，コンピュータの仕組み，ネットワークの仕組み，クラウド，コンピュータウィルスとセキュリティ，共通鍵など）

(2)． 個人情報等倫理との関連

情報活用演習Ⅱ（個人情報保護，知的財産と著作権など）

(3)． AI との関連（機械学習）

計測と制御（パソコンの仕組み，制御，誤差逆伝播法（３層のパーセプトロンによるニューラルネット解析））

(4)． IoT を含むビッグデータとの関連

情報ネットワーク論（ネットワークの仕組みなど）

インターネットビジネス（ビジネスモデルの紹介など）

卒業研究（呉 OCN　空き家問題の現状と課題）

(5)． データマイニングとの関連

データ解析（ピアソン統計とフィッシャー統計，ノンパラメトリックなど）

数理統計学（確率と統計の基礎，ピアソン統計とフィッシャー統計）

統計学演習（ピアソン統計演習とフィッシャー統計演習）

多変量解析入門（重回帰分析と数量化Ⅰ類など）

データベース検定講座（関連データベース３級程度）

情報数学（ブール代数，情報エントロピー，線形代数，線形計画，ゲーム理論など）

(6)． 文系・理系の専門分野との関連

OA 演習Ⅰ（ワープロ検定３級程度）

OA 演習Ⅱ（表計算検定３級程度）

コンピュータ演習Ⅰ（タッチタイピングとビジネス文書の作成（３級程度））

コンピュータ演習Ⅱ（表計算検定３級程度）
パソコン活用演習（パワーポイントによるプレゼンテーション）
数値シミュレーションエン演習（ニュートン法，掃き出し法，リプシッツ不動点定理など）

以上（1）～（6）は著者が担当した授業科目を羅列したに過ぎない。1994年～2021年の27年間，情報関連の科目を振り返ってみて言えることは，情報関連の科目を羅列したところで，1974年に示された「データサイエンスの原理」に基づくカリキュラムの将来の方向性を俯瞰できる訳でもないと言うことである。1980年代ごろから各大学に情報関連の学科が急激に新設されたが，2004年ごろから陰りが見えはじめ，著者が勤務していた経営情報学科は募集停止に追い込まれた。2005年ごろから各大学に福祉関連の学科が新設されるにつけ，情報関連の授業科目はその姿を次第に消していったように思われる。著者が担当した授業科目は，大学教育に於いては情報技術革新のその時代に流行したものに過ぎないと言える。これでデーサイエンティストを育成する教育システムであるとはとてもいえない。

著者にとって，これがデータサイエンスに近い研究であると見なせるのは，2014年に学生の卒業研究で呉市の地域活性化研究で取り組んだ「呉市の安心・安全な生活環境の確保」で，「呉市内の売物件データの分析から分かること[6]」（2014年３月）程度である。その経緯は昨年度公表した「インテリアと住居学に関わって―数学を手段とした取り組み―[13]」の中で述べたと通りである。

そこで分析した売物件データはインターネット上から取り込んだものであるが，呉地域から取り込んだそのソースデータは，インスタンス（行）が480件，属性（列）が33項目で公表されたデータセットである。それを因子分析をし，赤池の説明変数選択基準で重回帰分析を行うことを繰り返して有効な属性を選定するまでには，相当な時間を費やしたことを付言する（ここでは機械学習は行っていない）。

このデータ分析で得た知見は，次の通りである。

「世帯数と売物件データを地域事象に選んだときの地域分析により，呉市の地域特性を明らかにした。呉市は18地区からなる実質地域であり，本解析での主な成果は，これ等の地区を中心性と時系列的な世帯数変化量とを組み合わせることにより，４つに分類できたことである。中心性としては，時系列的な世帯数変化量との相関性から売物件エネルギーの残差を採用した。ここで，エネルギーとは重力モデルで算出された地区間の相互作用のことである。この解析手法により，売物件データによって呉市の地域特性を見出す方法を確立した。さらに，この解析手法を190の地域メッシュからなる形式地域に適用して，その成果を呉 OCN 会議で発表した。」（2014年３月）

学生による卒業研究とはいえ，呉市という自治体を対象に「空き家問題」を１年間掛けてインターネットからデータを取り込んで分析するという実に骨の折れる地域活性化研究に取り組んだ。許された時間内で結果を出すというのは，教育面で指導していくにはやや負担が大きすぎるのではという感が否めない。

B.（当時，データサイエンスは意識していない）著者の研究

データサイエンスとは科学である。科学とはどういうことなのか，中村正直訳「史学」（1879年）によれば「実験や観測に基づく経験的実証性と論理的推論に基づく体系的整合性」と言われている。

科学には２通りの手法がある。１つは演繹的方法（論理的推論に基づく体系的整合性），他の一つは帰納的方法（実験や観測に基づく経験的実証性）である。

演繹的方法とは，はじめに理論を打ち出して，それを演繹して思考実験・シミュレーションを行い，これを事実と照合して予測を行うもの。

帰納的方法とは，沢山の事実の中から規則性を抽出して理論を導き出し，原因と結果の関係を明らかにするもの。

端的に言えば「同じ条件で同じ実験を行えば，誰でも同じ結果が得られる」，この普遍性がサイエンスである。逆に，結果を見ればその原因が予測できるものを扱う。原因と結果の関係が見出せないデータ，改ざんされたデータ，風評などから得られたデータ，一部分だけが漏洩されたデータなど，悪質なデータは扱えないことを熟知することであるとも言える。

データサイエンスは，科学である以上はこの２面性を備えているのであって，データサイエンティストはその２面性の素養を培ってきた者ということになる。具体的には，コンピュータサイエンスにおいてデータサイエンティストは，理論をもとにアルゴリズムを見出してそれをプログラミングし，演繹的に数値シミュレーションで事実と照合する素養，またフィールドワークを行い統計解析により帰納的に現実を洞察する素養の両方を兼ね備えた者ということになる。

この点を踏まえてデータサイエンスと1998年～2005年の８年間，著者が取り組んだ研究とを対比してみる。演繹的な研究に関してはこの節で，帰納的なフィールドワークに関しては次節で俯瞰する。

◎演繹的な研究に携わって

1）．多変量解析

「直交射影行列を用いた多変量解析（データ解析の数学的理論）[7]」

ここでは「統計解析で用いる基本統計量を，要素がすべて１の列ベクトルで生成された直交射影行列 P，Q をデータに作用する変換で示し，これを PQ 変換と名付けて多変量解析に応用した。この変換により回帰分析，主

成分分析，正準相関分析などの理論的展開が数学的に見通しの良いものとなった。PQ 変換を，分散分析，判別分析，因子分析に適用してみたが顕著な成果があったとはいえない。」（1998年11月）

回帰分析，分散分析，正準相関分析などの統計解析にはプロジェクション（投影）という共通の概念がある。その一つが PQ 変換である。また，分散比についても共通の概念があり，統計解析の有意性を保証するものである。

2）．計算理論

「帰納的関数を用いた数値計算の基礎的理論（代数系の構造と計算可能性）―アルゴリズム―[8]」

ここでは「理想的計算機を仮定して代数系の基本的な問題を計算する場合，どのようなアルゴリズムが必要なのかを考察した。代数系の問題が計算可能か否かは，その問題を帰納的関数で表すことができるかどうかと同義であると定義し，ここでは帰納的関数を処理する理想的計算機上でのアルゴリズムが存在し，構造化プログラミング可能であることを前提とした。本論文の特徴は，有理数の g 進展開，連続関数の多項式近似をホーナー法で記述し，これは帰納的関数で表すことができることから計算可能であるとした点である。さらに，関数空間上での計算可能構造性についても考察した。」（2001年 7 月）

3）．確率微分方程式

「確率微分方程式による流体粒子の挙動（測度論的確率論と数値解析）―数値解析―[9]」

ここでは「乱流平板境界層内の流体粒子の流跡線を，２つの仮定を設けることによって確率微分方程式による数値シミュレーションで示した点である。１つ目の仮定は，流体粒子の位置を２次元ラグランジェ記述しこれに揺動を付加した。２つ目は，流速はあらかじめ求められた２次元速度場（オイラー記述）のものが採用できるとした。本論文の動機は，流れ場の支配方程式に揺動項が付加されるべきではないだろうかという著者の予見から始まる。考察として，フォッカー・プランク方程式からネルソンの加速度の定義を援用して流体粒子の運動方程式を導いた。得られた方程式は，あたかも NS 方程式に揺動項が付加された形であり，流れ場の支配方程式の構成において揺動散逸原理の必要性を示した。」（2002年 7 月）

4）．地域分析

「地域分析と数値シミュレーション（地域メッシュによる解析）―数値解析―[10]」

ここでは「数値シミュレーションとして地域メッシュ上で表章された地域事象（数値データ）の分析方法を示した。地域分布状況については，地域傾向面分布の解析式を著者の提案する最急降下法で求めた。また，地域分布の時系列的変化の解析式も同様に求めた。地域間相互作用については，重力モデルを適用してその地域ポテンシャルと地域エネルギーで表し，さらにこの作用を地域間の移動量に関係づけて地域の変化をとらえようと試みた。結果は，お互いの移動が相殺されて地域間に大きな変化は表面的に表れなかった。」（2003年 7 月）

5）．制御理論・学習機能

「制御に関する一考察（変分原理と最適化問題）[11]」

こでは「制御を最適化問題として捉えたものであり，この限りに於いては従来の方法と何ら変わりない。その根本原理は変分原理であり，本論文で採用した目標値達成への手法は最急降下法である。著者は，最急降下法に独自の勾配修正を付加してその収束性（収束時間の短縮）を試みた。また，２段階非線形制御（３層のパーセプトロンによるニューラルネット解析）を取り上げ，そのパラメータ同定問題を取り扱った。２段階の線形制御にその中間信号の処理としてシグモイド関数を導入したものを本論文では２段階非線形制御と捉え，そのパラメータの同定には誤差逆伝播法を適用した。また，パラメータの修正は最急降下法で行ったが，この方法にも著者の勾配修正を付加した方法を導入した。これらの具体的なプログラムは Java の文法に従って本論文中で公開した。著者の提案した勾配修正を付加した最急降下法の収束性は，通常の方法よりも若干の計算時間の短縮が認められた。また，この方法は非線形多変量解析にもそのまま役立つものであることを強調した。」（2003年 7 月）

6）．間欠カオス

「間欠カオスに関するデータ解析（複素フーリエ，ウエーブレット，非線形多変量の 3 つの解析）[12]」

ここでは「ベルヌーイ・シフト力学系で生じる間欠カオスについて，その時系列データの諸特性を複素フーリエ，ウエーブレット，非線形多変量（３層のパーセプトロンによるニューラルネット解析）の３つの解析手法により明らかにした。複素フーリエ解析では，時系列データ全体のパワースペクトル，自己相関関数を求め間欠カオスの特性を図示した。また，ガボール変換により時系列データの局所的な時間－周波数解析を行った。これに対して，ウエーブレット解析では局所的な時間－スケール解析を示し，両者により間欠カオスの時系列データの局所的な特徴を図示した。図示した解析結果は，いずれも Devaney や Schuster 等が定義するカオスの性質を持っていることが判明した。また，信号再生の観点から複素フーリエ，ウエーブレット，非線形多変量の３つの解析を比較した。さらに，間欠カオスの跳躍点の検出を B スプラインウエーブレット解析で試みた。信号再生に関しては複素フーリエが最も良好であり，また跳躍点の検出も B スプラインウエブレットで可能であることがわかった。」（2005年２月）

7）．マルチフラクタル解析

「DLA シミュレーションとフラクタル解析[13]」

ここでは「非整数微分・積分を応用して，非整数ブラ

ウン運動のDLA挙動をシミュレーションした。クラスター解析にはマルチフラクタル次元を導入した。また，DLA形成に要するエネルギーとフラクタル次元の関係を４次近似曲線で示した。これにより，DLAの形成はマルチフラクタル的であることが判明した。ただし，解析上の問題点としてクラスターを覆う最適セル配置のアルゴリズムが未確立であるために，マルチフラクタル解析の精度が粗雑となった。この精度向上のためには，問題点についてのさらなる研究が必要である。」（2005年12月）

以上，演繹的立場からのデータサイエンス取り分け，その中のコンピュータサイエンスの実施であるが，理論からアルゴリズムを見出し，それをプログラミングする素養が必修であることを断言する。

C．データ解析の基礎となる科目と実践（物理学実験，海洋観測に関わって）

◎物理学実験（基礎教育）に携わって

物理学実験（測定とデータ処理）を1984年～1994年の10年間担当，その間に物理学実験教育で行った内容を記す。

科学は「同じ条件で同じ実験を行えば，誰でも同じ結果が得られる」ということが必修である。このためには，測定したデータの精度が重要となる。物理学実験で言えることは，実験計画の段階であらかじめ測定精度を設定して計器を決め，それで測定段階に入るのが常であるが，これを踏まえないでデータを測定しても，予測れた物理法則には至らないということである。著者は，インターネットに流通しているデータを鵜呑みにすることには危惧している。信頼性の保証が成されていないものが多く，「悪貨は良貨を駆逐する」という観が強い。

インターネットが普及して，目的に叶ったデータと思えるものを扱う場合，提供者はそのデータの精度・信頼度を明示すべきである。たくさんのデータ，各方面からの多種多様なビッグデータがインターネットから配信されたとしても，その信頼性が保証されていない限り，それは「デマである」と揶揄されても反論できないだろう。それをデータセット化してデータ分析し，そこから洞察された結果を意思決定に反映させることは，妄想に近いとしか言いようがない。その典型として，今日の「新型コロナウィルス」の風評がある。これまでに人類が経験したことのない得体の知れないデータ（進化する変異株）を扱っているのであるが，インターネット上のワクチン接種に関して「良いのか悪いのか」「デマ」が飛び交っている感じが否めない。ここにデータサイエンスの科学的倫理が問われていると思う。

著者は，すべての学生に「如何に正しいデータ（エビデンス）を得るか」という演習・実験の授業を施し，正しいデータ処理法を身に着けさせることの重要性を強く推奨する。その一例として著者が10年間に渡って関わってきた物理学実験を次に掲げる。

ここで，データ分析，データ解析，データ処理の術語を定義しておく。「データ分析」とは，データがどんな要素で構成されているのか，主要なカテゴリーを見出すことで，データを見直すことである。「データ解析」とは，データの構成要素を理論的・統計的に解析して洞察することである。「データ処理」とは，データの収集・分析・加工の一連のプロセスのことである。

1)．「本学の物理学実験教育について（教育方法）[14]」

ここでは「学生数が１クラス100名程度の多人数教育について，より効果的な物理学実験教育を展開するにはどのような方法がベータなのか，その教育方法を提案した。この教育方法は実験教育を物理学（座学教育）の補助的手段と捕らえるのではなく，実験を通して学生の工学的な適正（精度を踏まえた測定，正しい測定技術とそのデータ処理の手法など）の向上を目指すことを目的とし，授業初めに学生に教員から独自の教育手法として試問を行い，測定方法にきめ細やかな指導を行うという教育方法を採用した。その教育成果は次回報告する。」（1985年３月）

2)．「本学の物理学実験教育について（教育成果）[15]」

ここでは「前報で提案した教育方法の成果を，試問と実験状況，レポートの作成状況，これらの総合評価の３点で示した。また，実験を積み重ねる毎のデータ処理能力の向上を評価した。試問の導入は実験意欲を高める点で効果が見られたが，それだけ実験時間を奪うためデータ処理能力の向上には効果が見られなかった。データ処理中心の指導を実験回数が増えるにつれて徐々に増やしていき，データ処理から得られえる洞察に重点を置いた授業展開を提案した。」（1986年３月）

物理学実験では，あらかじめ見込まれた精度の測定器を用意して，その範囲でできるだけ正確な手法で測定されたデータこそが，その後の洞察に役立つことの重要性を説いた。誤差の３つの公理が示すように，データは誤差を伴う。「出来りだけ正確にデータを測定すること」が重要であり，これは，データサイエンスにおいても同じである。

次に，座学での「データ解析」の授業を行う中で，次のような知見を得た。

3)．「本学のデータ解析の教授法（加重型回帰分析法とその結果）[16]」

ここでは「授業科目「データ解析」に関してその問題点を反省して望ましい授業の一例を提示した。そこでは，実験計画の立て方，母平均の区間推定，母平均の差の検定（非等分散性よりウェルチの検定），単回帰分析，単回帰線形判別分析等の統計解析の手法の学習の流れで授業展開を提示した。次に，データ解析のアイディアとして加重型回帰分析法を試みて，その有意性を誤差の少ない模擬的なデータを使って示した」（1997年12月）

◎帰納的な研究（フィールドワーク）に携わって

1990年～1996年の７年間，九州大学応用力学研究所を中心とした沖縄西方・東方の黒潮の海洋観測に参加し，フィールドワークを行った。データ分析の過程で現象のモデリングを行い，事実として起こっていることは何なのか，その理解を深めた。

4)．「沖縄東方黒潮の流動解析と力学解析（データ解析）[17]」

ここでは「1990年７月に沖縄東方海域のH－LINEで曳航観測したADCPデータをもとに，ADCPデータの解析方法と結果を報告した。また，この海域に流れる黒潮続流の流動解析，力学解析を行った。さらに，これらの解析結果をもとに，この海域の流れ場の簡単なモデリングを行った。このことから，この海域に黒潮続流が潜入してくる際に，独特のメカニズムが潜んでいる可能性を示唆した。沖縄東方海域の流動解析の報告は希少である。」（1995年７月）

5)．「地衡流平衡の度合いについて（データ解析）[18]」

ここでは「黒潮がどの程度地衡流平衡にあるのかを，ADCPデータとCTDデータの解析結果の比較で示した。比較の方法は両者から得られたそれぞれの鉛直流速分布状況を比較するもので，本論文のオリジナルである。結果として，潮流れの陸棚斜面近傍を除いて，主流部において地衡流の平衡性が確認された。また，逆に黒潮が地衡流平衡にあることが立証されればADCPデータの水平流速分布から鉛直速分布が計算できるという手法を提案した。」（1996年８月）

6)．「黒潮流れのモデリングについて（データ解析の手法）[19]」

ここでは「流れを数値シミュレーションする場合のモデリングの重要性，モデリングによる流動現象解明の必要条件，十分条件の検討，流動現象解明のため簡単なモデリングから複雑なモデリングへ移行する山登り法を提案した。流動現象の一例として黒潮流れを取り上げて，その力学計算モデル，地衡流調節モデル，観測データによるモデル，１と1/2層モデルへと山登り的にモデリングを移行していく数値シュミレーション手法を論じた。」（1996年８月）

海洋観測に限らずフィールドワークによるデータ解析では，現象を捉えるためにモデリングを行い，現象の理解を深めていくことが通常である。それはまるでブラックボックスの中に手を突っ込んで中にあるものを探っていく感じだが，この帰納的な解析こそ，ビッグデータの解析には不可欠ではないかと思う。

D．データサイエンスに関する一連の研究について

データサイエンスに関する一連の研究について，著者が1982年～2002年，2008年～2021年に取り組んできたテーマと対比する。その中間については前述の「B.（当時，データサイエンスを意識していない）著者の研究」で述べた。前半の21年間は数値解析，後半の15年間はインテリア・住居学に関してであるが，これらはもちろん，データサイエンスを意識して行ったものではないことを付言する。

数値解析では，MAC法，近似解法，線形関数解析，非線形関数解析，Lie微分などを駆使し，NS方程式の解についてチャレンジした。それを次のようにまとめた。

「数値解析に関するいくつかの提言―関数解析的見地から―[20]」

内容：「NS方程式は初期値―境界値問題に限定すれば，それを数学的証明により厳密解として示すことよりもむしろ，関数解析的に解のクラスを定義して数式のアプリオリ評価を行い，解を数値解析的に特定することの方が，解の存在性や一意性を具体的に示したことに成るのではないか，ガロア理論を援用して体の拡張により特定された解，あるいは正しく解析接続された解であると保証されているのであれば，数学的証明抜きでもコンピュータ内部で正しくプログラミングされた演算で得られた数値解の正当性を，著者は認め得る者である。むしろコンピュータ内部での機械的な演算の正確さからして，正しくプログラミングされた数値解はそのまま正解であり得ると主張する。」（2020年12月）

インテリア・住居学では，線形計画・遺伝アルトリズム，グラフ理論，構造モデリング，マトロイド，ヘドニック・アプローチ，ISM法，AHP，非線形計画などを駆使し，最適な間取りについてチャレンジした。それを次の様にまとめた。

「インテリアと住居学に関わって―数学を手段とした取り組み―[21]」

内容：「著者は14年来，様々な数学的手法を，インテリアあるいは住居の内部動線，室内動線，ゾーニング等に当てはめた研究を進めてきた。その動機は，「プロの間取り図には何か数学的に抽出でき得る根拠が見出せるのではないだろうか」ということであり，研究を推進する毎にその答えは「イエス」であることを主張する者である。著者は数学的手法としてニューラルネットワークを用いたAIについては研究を行ってはいないが，スマホを見ながら自分が趣向するインテリアあるいは住居を選定することが当たり前の時代がすでに来ていることを予見する。」（2020年12月）

いずれも数学的な解析による一連の研究であるが，「基礎研究」として一つのターゲットにどのくらいの解析を試みていくのか，これは学者ならではの妙味である。それには「基礎科学力」が当然伴う。「基礎研究」には実利を無視して黙々と研究を積み重ねていくものが多々ある。「いったいこれが何の役に立つのか」と自問したくなるものは，データサイエンスが実利を目的とするならばナンセンスであると言えるだろう。このあたりが，データサ

イエンスと「基礎研究」との分岐点ではないだろうか。実利を主とするのであれば，データエンジニアリングと言う術語の方が妥当かも知れない。

E. データサイエンスについて，別の角度からのアプローチ

本報告を記すに当たって最も著者が主張したい根幹は，**「自己点検・第三者評価の文化」**である。この見解は著者が2007年～2021年の14年間，大学の業務として（一般財団法人）大学・短期大学基準協会に登録されたALO（認証評価連絡調整責任者）や評価員を経験したことによる。

今日，大学間では共通の観点によりエビデンスに基図いた自己点検・第三者評価が行われている。この文化は国境のないインターネット上でも必須ではないだろうか。組織はデータに関して責任を持つべきである。と言うのは，インターネット上で得られたビッグデータは世界中の利用者から得られた共通概念であり，これを単独のある組織のみが独占し，自己都合のいいデータを操ってある方向に向かわしめるのは，世界の公平性に偏りが生じる要因があると思うからである。その対策として，公官庁に社会倫理・企業倫理・情報倫理の立場から点検項目（観点・区分・テーマ・基準等）を具体的に示していただいて，今日，大学では当たり前のように行われている**自己点検・第三者評価報告書**[22]の作成とその第三者評価員による相互評価を行う文化は，グローバル時代の情報社会にとって必須の課題であると著者は主張すると共に，ビッグデータの公共利用性を担保する上でも重要な課題であると強調したい。

ここに，ビッグデータを扱う組織はその公共性を担保するためにもエビデンスを明示し，定期的に自己点検・第三者評価用の報告書を作成して第三者評価員あるいは同業者間で相互評価を受審し，その認証は社会に公表すべきであることを提案する。

ところで，データサイエンスに期待する者には次の「4つの神話・勘違い[4]がある」（p50）とされている。それを著者なりに記述する。

(1) AI結果を鵜呑みに受け止めてしまうこと

データサイエンティストの洞察が浅ければ，それなりの結果しか得られないばかりか，意思決定に誤謬をもたらす。AI結果よりもその組織の意思決定にとって最適なデータ解析は有り得るし，現に各企業はそれを企業秘密として実施している。

(2) 「ビッグデータを参照すればそれが正しい」と言う誤解

小さな組織が扱うデータに有益性が有りながら，それをインターネットに計上しないからといって無視する暴挙，逆にその組織にとってビッグデータは自己の組織に最適な参照となり得るのかという検証の必要性。

(3) AIソフトを絶対的なものと信じ込んで導入すること。

データサイエンティストが作成したAIソフトを，自己組織の意思決定にそのまま起用したことによるミスマッチ，AIでの解析結果がその組織にとって最適である保証はない。解析結果を洞察して意思決定を行うのはあくまでもその組織の人間である以上，自己組織の規模にふさわしい独自の解析手法が取られている場合が多いことの現実。

(4) 人間とAIソフトの対決結果の如何により，AI結果を優先する暴挙。

個々の人間には，社会倫理，自己ポリシー，自己同一性がある。その個性をAIとの対戦で安易に修正する必要などない。AIはあくまで，あるデータサイエンティスト達がある組織の有益性で作成したアルゴリズム・ソフトに過ぎない。

上述の（1）～（4）を踏まえ，ここに著者は次のことを提案したい。

組織はその収益のために，自己組織のガバメントの意思決定を中心にPDCAを回しつつ発展していると思わいたのだが。その過程の中でデータサイエンティスト達が行うのは「ビジネス理解・データの理解・データの準備・モデリング・評価・展開」のサイクルである「CRISP-DMのライフサイクル[4]」（p.76，p.83）であり，これはPDCAの主としてDCの繰り返しに対応している。これにより自己組織にふさわしい最適アルゴリズムが開発・更新されデータのクリーニング等を経て最適なデータ分析が可能となるのである。組織はその成果を査定して改善計画を立て次の進展に向けて取り組む，これがPDCAである。

組織はこれらの取り組みを点検項目に照らして自己点検報告書とまとめ，それを同業者が相互評価を行う。これを公官庁が認証をするという第三者評価の文化を形成する，このような認証（アクレジテーション）はビッグデータの公共利用性を担保する上で有効であり，それなくして逆にインターネット上でいい加減なデータが流出しそれを分析することは意味のあることではない。先にも述べたが「悪貨は良貨を駆逐する」という観が強い。

さらに，大学教育において学生に自己点検・第三者評価のプロセスを教授することは，今後，不可欠であると提案する。著者は，教養・リテラシーでのデータサイエンス教育が単なるデータの収集・分析・加工，すなわちデータ処理のスキルとして位置づけられることには抵触する。データサイエンスは個々の組織のガバメントに不可欠なPDCAを回すことの，主としてDC部分のデータを科学的に分析・解析する役割を担っている。そこから得られた洞察・意思決定は第三者評価取り分け査定（アセスメント）を導入させることで，社会倫理・企業倫

理・情報倫理にも耐えうるものでなくてはならない。単にAI任せの無責任な取り扱いに終始したのでは，社会の発展を誤った方向に向かわしめる可能性があるのではと思う。

このことを踏まえ，教育・リテラシーでは，学生には組織の**PDCA**を学ばせる中でその目的達成のためにデータサイエンスを如何に実現するのか，相互評価とは如何にあるべきかといった視点での教育が不可欠であることを提案する。

3. お わ り に

これまでの考察と提言をまとめる。

1)． 1980年代ごろから各大学に情報関連の学科が急激に新設されたが，2004年ごろから陰りが見えはじめ，著者が勤務していた経営情報学科は募集停止に追い込まれた。2005年ごろから各大学に福祉関連の学科が新設されるにつけ，情報関連の授業科目はその姿を次第に消していったように思われる。著者が担当した情報関連の授業科目は，大学教育に於いては情報技術革新のその時代に流行したものに過ぎず，1974年に示された「データサイエンスの原理」に基づくカリキュラムの将来の方向性を俯瞰できるものではない。これでデーサイエンティストを育成する教育システムであるとはとても言えない。

著者にとって，これがデータサイエンスに近い研究であると見なせるのは，学生の卒業研究で呉市の地域活性化研究で取り組んだ「呉市の安心・安全な生活環境の確保」で，「呉市内の売物件データの分析から分かること[6)]」（2014年３月）程度である。

売物件データはインターネット上から取り込んだ。インターネットで呉地域から取り込んだソースデータは，インスタンス（行）が480件，属性（列）が33項目で公表されたデータセットであるが，その中から有効な属性を選定するまでには，相当な時間を費やした。ここでは機械学習は行っていない。

呉市という自治体を対象に「空き家問題」を１年間掛けてインターネットからデータを取り込んで分析するという地域活性化研究に取り組んだ。許された時間内で結果を出すというのは，教育面で指導していくにはやや負担が大きすぎるのではという感が否めない。

2)． データサイエンスは，科学である以上はこの２面性を備えている。１つは演繹的方法（論理的推論に基づく体系的整合性），他の一つは帰納的方法（実験や観測に基づく経験的実証性）である。データサイエンティストはその２面性の素養を培ってきた者ということになる。

演繹的立場からのデータサイエンス取り分け，その中のコンピュータサイエンスの実施であるが，理論からアルゴリズムを見出し，それをプログラミングする素養が必修であることを断言する。

科学は「同じ条件で同じ実験を行えば，誰でも同じ結果が得られる」ということが必修である。このためには，測定したデータの精度が重要となる。インターネットが普及して，目的に叶ったデータと思えるものを扱う場合，提供者はそのデータの精度・信頼度を明示すべきである。

著者は，すべての学生に「如何に正しいデータ（エビデンス）を得るか」という演習・実験の授業を施し，正しいデータ処理法を身に着けさせることの重要性を強く推奨する。

フィールドワークによるデータ解析では，現象を捉えるためにモデリングを行い，現象の理解を深めていくことが通常である。この帰納的な解析こそ，ビッグデータの解析には不可欠ではないだろうか。

3)． 数学的な解析による一連の「基礎研究」として「数値解析に関するいくつかの提言」，「インテリアと住居学に関わって」の２例を示した。一つのターゲットにどのくらいの解析を試みていくのか，それには「基礎科学力」が当然伴うが，「基礎研究」には実利を無視して黙々と研究を積み重ねていくものが多々ある。データサイエンスが実利を目的とするならば，このような取り組みは敬遠されがちになる。このあたりが，データサイエンスと「基礎研究」との分岐点ではないだろうか。実利を主とするのであれば，データエンジニアリングと言う術語の方が妥当かも知れない。

4)． 本報告を記すに当たって最も著者が主張したかった提言は，「自己点検・第三者評価の文化」である。

今日，大学間では共通の観点によりエビデンスに基図いた自己点検・第三者評価が行われている。この文化は，国境のないインターネット上でも必須ではないだろうか。参画した組織は，データに関して責任を持つべきである。

データサイエンスは個々の組織のガバメントに不可欠なPDCAを回すことの，主としてDC部分のデータを科学的に分析・解析する役割を担っている。そこから得られた洞察・意思決定は第三者評価取り分け査定（アセスメント）を導入させることで，社会倫理・企業倫理・情報倫理にも耐えうるものでなくてはならない。単にAI任せの無責任な取り扱いに終始したのでは，社会の発展を誤った方向に向かわしめる可能性があることに危惧する。

ビッグデータを扱う組織は，その公共性を担保するためにもエビデンスを明示し，定期的に自己点検・第三者評価用の報告書を作成して，第三者評価員あるいは同業者間で相互評価を受審し，その認証は社会に公表すべきべきであることを提案する。

さらに，大学教育において学生に自己点検・第三者評価のプロセスを教授することは，今後，不可欠であると提案する。教育・リテラシーでは，学生には組織のPDCAを学ばせる中で，その目的達成のためにデータサイエンスを如何に実現するのか，相互評価とは如何にあるべきかといった視点での教育が不可欠であることを提案する。

要　　旨

著者が担当した情報関連の授業科目は，大学教育に於いては情報技術革新のその時代に流行したものに過ぎず，1974年に示された「データサイエンスの原理」に基づくカリキュラムの将来の方向性を俯瞰できるものではない。

著者にとって，これがデータサイエンスに近い研究であると見なせるのは，2014年に呉市という自治体を対象に，「空き家問題」を１年間掛けてインターネットからデータを取り込んで分析するという地域活性化研究に取り組んだことである。インターネットから取り込んだソースデータから有効な属性を選定するまでには，相当な時間を費やした。また，ここでは機械学習は行っていない。

データサイエンスは，科学である以上，演繹的方法（論理的推論に基づく体系的整合性）と帰納的方法（実験や観測に基づく経験的実証性）の２面を持つ。データサイエンティストはその２面性の素養を培ってきた者ということになる。

演繹的立場からのデータサイエンス取り分け，その中のコンピュータサイエンスの実施であるが，理論からアルゴリズムを見出し，それをプログラミングする素養が必修であることを断言する。

インターネットが普及して，目的に叶ったデータと思えるものを扱う場合，提供者はそのデータの精度・信頼度を明示すべきである。フィールドワークによるデータ解析では，現象を捉えるためにモデリングを行い，現象の理解を深めていくことが通常である。この帰納的な解析こそ，ビッグデータの解析には不可欠ではないだろうか。

「基礎研究」には実利を無視して黙々と研究を積み重ねていくものが多々ある。データサイエンスが実利を目的とするならば，このような取り組みは敬遠されがちになる。このあたりが，データサイエンスと「基礎研究」との分岐点ではないだろうか。実利を主とするのであれば，データエンジニアリングと言う術語の方が妥当かも知れない。

データサイエンスは個々の組織のガバメントに不可欠な PDCA を回すことの，主として DC 部分のデータを科学的に分析・解析する役割を担っている。ビッグデータを扱う組織は，その公共性を担保するためにもエビデンスを明示し，定期的に自己点検・第三者評価用の報告書を作成して，第三者評価員あるいは同業者間で相互評価を受審し，その認証は社会に公表すべきべきであることを提案する。

教育・リテラシーでは，学生には組織の PDCA を学ばせるべきで，組織の目的達成のために PDCA を回す中で如何にデータサイエンスが実現できるのか，また相互評価とは如何にあるべきかといった視点での教育の提供が不可欠であると提案する。

引　用　文　献

1) 総合イノベーション戦略推進会議：人間中心の AI 社会原則，2019年３月
2) 数理・データサイエンス・AI 教育プログラム認知制度検討会議：「数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）」の創設について，2020年３月
3) 数理・データサイエンス・AI 教育プログラム認知制度検討会議：「数理・データサイエンス・AI 教育プログラム認定制度（応用基礎レベル）」の創設について，2021年３月
4) ジョン・D・ケレハー＆ブレンダン・ティアニー（今野紀雄　監訳，久島聡子　訳）：「データサイエンス」，NEWTON PRESS，2020年１月
5) 渡辺　治：コンピュータサイエンス（計算を通して世界を観る），丸善出版，2019年４月（第２刷）
6) 石田真由美，古川博仁：「呉市内の売物件データの分析から分かること」，呉 OCN 地域活性化研究として採択，2014年３月
7) 古川博仁：直交射影行列を用いた多変量解析（データ解析の数学的理論），呉大学短期大学部紀要　第２号　pp.15−34，1998年11月
8) 古川博仁：帰納的関数を用いた数値計算の基礎的理論（代数系の構造と計算可能性）―アルゴリズム―，呉大学短期大学部紀要　第５号　pp.1−22，2001年７月
9) 古川博仁：確率微分方程式による流体粒子の挙動（測度論的確率論と数値解析）―数値解析―，呉大学短期大学部紀要　第６号　pp.1−21，2002年７月
10) 古川博仁：地域分析と数値シミュレーション（地域メッシュによる解析）―数値解析―，呉大学短期大学部紀要　第７号　pp.21−25，2003年７月
11) 古川博仁：制御に関する一考察（変分原理と最適化問題），呉大学短期大学部紀要　第７号　pp.27−41，2003年７月
12) 古川博仁：間欠カオスに関するデータ解析（複素フーリエ，ウエーレット，非線形多変量の３つの解析），呉大学短期大学部紀要　第８号　pp.29−47，2005年２月
13) 古川博仁：DLA シミュレーションとフラクタル解析，呉大学短期大学部紀要　第９号　pp.19−44，2005年12月
14) 古川博仁：本学の物理学実験教育について（教育方法），広島工業大学研究紀要　Vol.19　No.23　pp.97−99，1985年３月
15) 古川博仁：本学の物理学実験教育について（教育成果），広島工業大学研究紀要　Vol.20　No.24　pp.127−135，1986年３月
16) 古川博仁：本学のデータ解析の教授法（加重型回帰分析法とその結果），呉大学短期大学部紀要　第１号　pp.55−66，1997年12月
17) 古川博仁：沖縄東方黒潮の流動解析と力学解析（データ解析），呉女子短期大学紀要　第９号　pp.71−79，1995年７月
18) 古川博仁：地衡流平衡の度合いについて（データ解析），呉女子短期大学紀要　第10号　pp.77−84，1996年８月
19) 古川博仁：黒潮流れのモデリングについて（データ解析の手法），呉女子短期大学紀要　第10号　pp.85−92，1996年８月
20) 古川博仁：数値解析に関するいくつかの提言―関数解析的見地から―，広島文化学園短期大学紀要　第53号　pp.1−6，

2020年12月
21） 古川博仁：インテリアと住居学に関わって―数学を手段とした取り組み―，広島文化学園短期大学紀要　第53号　pp.7－10，2020年12月
22） 平成30年度広島文化学園短期大学自己点検・評価報告書，本学ホームページ，http://www.hbg.ac.jp/info/jouhoukoukai/index_top.html，（2019年 4 月）

Summary

The Information-related subjects in charge by the author were just subjects that was popular at that time in the area of information technology innovation in university education，and I couldn't get a bird's eye view of the future direction of the curriculum based on the "Principles of Data Science" presented in 1974.

For the author, it is able to be regarded a regional revitalization study targeting a local government called Kure City to take in data from the Internet and analyze it over a year over the "vacant house problem" in 2014 as the research close to data science. It spent a considerable amount of time to select valid attributes from source data fetched from the Internet. Also, machine learning was not performed here.

Data science is as long as science, it has two sides: a deductive method, that is systematic consistency based on logical reasoning and an inductive method, that is empirical demonstrability based on experiments and observations. Data scientists were able to be said that they have cultivated cultivate the two-sided background.

From a deductive standpoint, in the field of data science, especially in the field of computer science, data scientists need the ability to find algorithms from theory and program them.

When the Internet becomes widespread and handles data that serves its purpose, the provider should clearly indicate the accuracy or reliability of the data. In data analysis by fieldwork, it is normal to perform modeling to capture the phenomenon and deepen the understanding of the phenomenon. In an inductive method, such a method is essential for big data analysis.

There are many "basic researches" that ignore actual interests and silently accumulate research. If data science is for profit, such efforts tend to be shunned. I think this is the turning point between data science and "basic research." If the main focus is on actual profits, I think it is preferable to call data engineering rather than the term data science.

Data science is mainly responsible for scientifically analyzing the data of the DC part of turning PDCA, which is indispensable for the administration of individual organizations.

Data science has the role of scientifically analyzing the data of the DC part mainly in order to turn PDCA, which is indispensable for the administration of individual organizations. Organizations that handle big data provide evidence to ensure its public nature, regularly prepare reports for self-inspection and are necessary to undergo mutual evaluation by a third-party evaluator or between peers. I suggest that the certifications qualified there should be published to society.

In education and literacy, I propose to students. Students should learn the PDCA cycle of the organization, and how to realize data science as they turn the PDCA cycle to achieve the purpose of the organization.

It is essential for them to be educated from the perspective of what mutual evaluation should be.