

平成 26 年度
データサイエンティスト育成研究会
< 報告書 >

平成 27 年 3 月 17 日

一般社団法人コンピュータソフトウェア協会
データサイエンティスト育成研究会

目次

はじめに	2
I. 概要		
1. 目的	3
2. 方法	4
3. 目標	6
4. プログラムについて	7
II. プログラム詳細		
1. 本プログラムが目指すデータサイエンティスト(DS)	8
2. プログラムの構成	9
3. データ活用プランニング	10
4. DSが必要とする知識とスキル	16
5. DS 育成手順	24
III. 支援ツール	25
IV. 補足		
1. プログラムの活用	31
2. 育成プログラムの課題	31
V. 考察	32
おわりに	33

はじめに

平成 26 年度データサイエンティスト育成研究会は、CSAJ の研究会として前年度 6 カ月間の活動成果を引継ぎ、平成 26 年 4 月 15 日（キックオフ）にスタートいたしました。その後、情報サービス業界視点で「データサイエンティスト」を育成するための教育プログラム開発を目標として 2 ヶ月に 1～2 度のペースで研究会を開催いたしました。研究会では、作業の効率化を図るため、グループワークを取り入れ検討・議論を重ねました。その成果物としての「データサイエンティスト育成プログラムの詳細な枠組」を中心に平成 26 年度の活動についてご報告させていただきます。

研究会の開催・推進にあたりましては、ご参加いただいた会員各位に多大なご支援ご協力を賜りました。この場をお借りして深謝いたします。

データサイエンティスト育成研究会 主査 昆 凡子

I. 概要

1. 目的

当研究会は、平成 25 年度研究会設置申請において以下の目的を掲げ、設置された。平成 26 年度は、平成 25 年度の成果をベースに「データサイエンティスト育成プログラム」の詳細について討議検討を行い、その成果を本報告書にまとめた。

大量データの有効活用が大きなテーマとなっている昨今、これを操る「データサイエンティスト」の育成が急務になっている。「データサイエンティスト」は、企業や政府機関、自治体の持っている大量データを分析し、企業の業績を飛躍的に伸ばす役割を担いつつある。そこで、「データサイエンティスト」の定義、資格要素など、国内においてまだまだあいまいな部分を調査研究し、その結果を踏まえ、CSA Jとして、国に提言していけるような内容作りを目指す。

平成 25 年度初期段階で検討し、以下を目的とすることで研究会として合意した。



情報システム（情報サービス）業が、ビジネスにおいてイニシアティブを取るためのデータサイエンティスト像を明確にし、その育成について概要を定める。

特に、情報サービス業がイニシアティブをとるためと強調した背景には、情報サービス業が従来どちらかという受身でサービスを提供してきたことを打破したいという研究会メンバーの積極的な姿勢がある。ビッグデータの活用など、経営や事業に能動的に参画する方向を提示する良い機会という認識である。

I. 概要

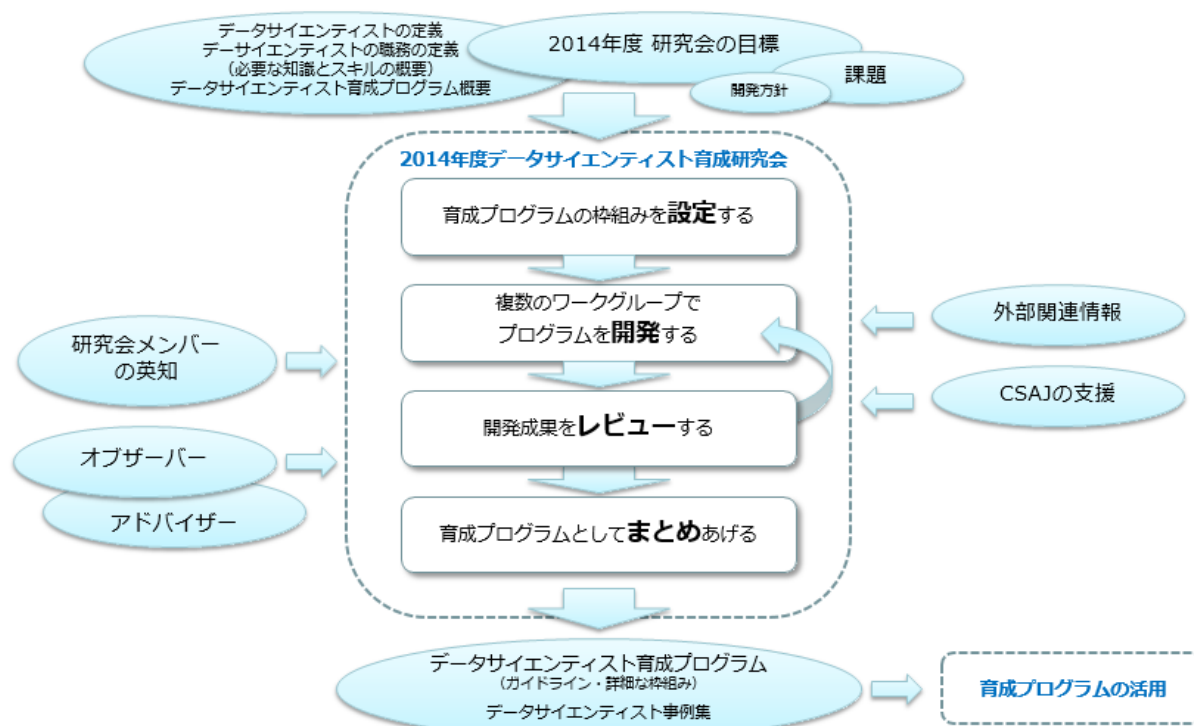
2. 方法

平成 26 年 4 月以降第 2 火曜日を中心に CSAJ 会議室を利用して参加企業による研究会を実施した。必要に応じてテーマに合った知識を有するオブザーバ、講師を招聘した。

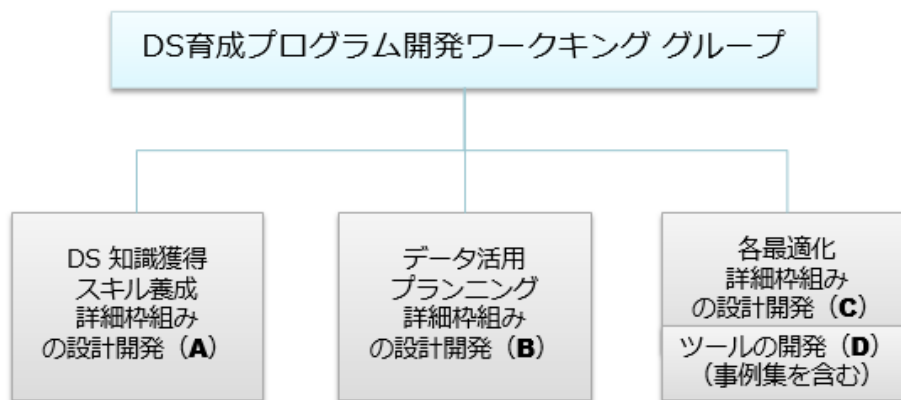
第 4 回～第 6 回は、ワーキンググループ 3 チームに分かれ、テーマ毎に分担して作業を行い研究の効率を高めた。各回の作業結果はコアメンバーがとりまとめ、その資料を研究会にて討議し、合意形成してゆく方法を採用した。

タイトル	開催日	概要	参加者数
第 1 回 研究会	平成 26 年 4 月 15 日(火)	キックオフ (メンバー紹介、前年度活動の確認と活動計画書の検討承認)	17 名
第 2 回 研究会	平成 26 年 5 月 20 日(火)	データサイエンティスト (以降 DS) 育成プログラムの枠組みを検討する	14 名
第 3 回 研究会	平成 26 年 6 月 12 日(木)	DS 育成プログラムの枠組みを決定し、詳細開発の手順を検討し設定する	16 名
第 4 回 研究会	平成 26 年 7 月 15 日(火)	DS 育成プログラムの詳細開発を複数のワークグループで分担し行う	21 名
第 5 回 研究会	平成 26 年 9 月 24 日(水)	DS 育成プログラムの詳細開発内容のレビューを行いつつ開発を継続する	17 名
第 6 回 研究会	平成 26 年 10 月 21 日(火)	各ワーキンググループの成果を研究会全体で確認・調整する	12 名
第 7 回 研究会	平成 26 年 12 月 17 日(水)	DS 育成プログラムの詳細開発内容のとりまとめを行う	12 名
第 8 回 研究会	平成 27 年 2 月 17 日(火)	とりまとめた DS 育成プログラムのレビューを行う	15 名
第 9 回 研究会	平成 27 年 3 月 17 日(火)	研究会報告と講師による講演	XX 名

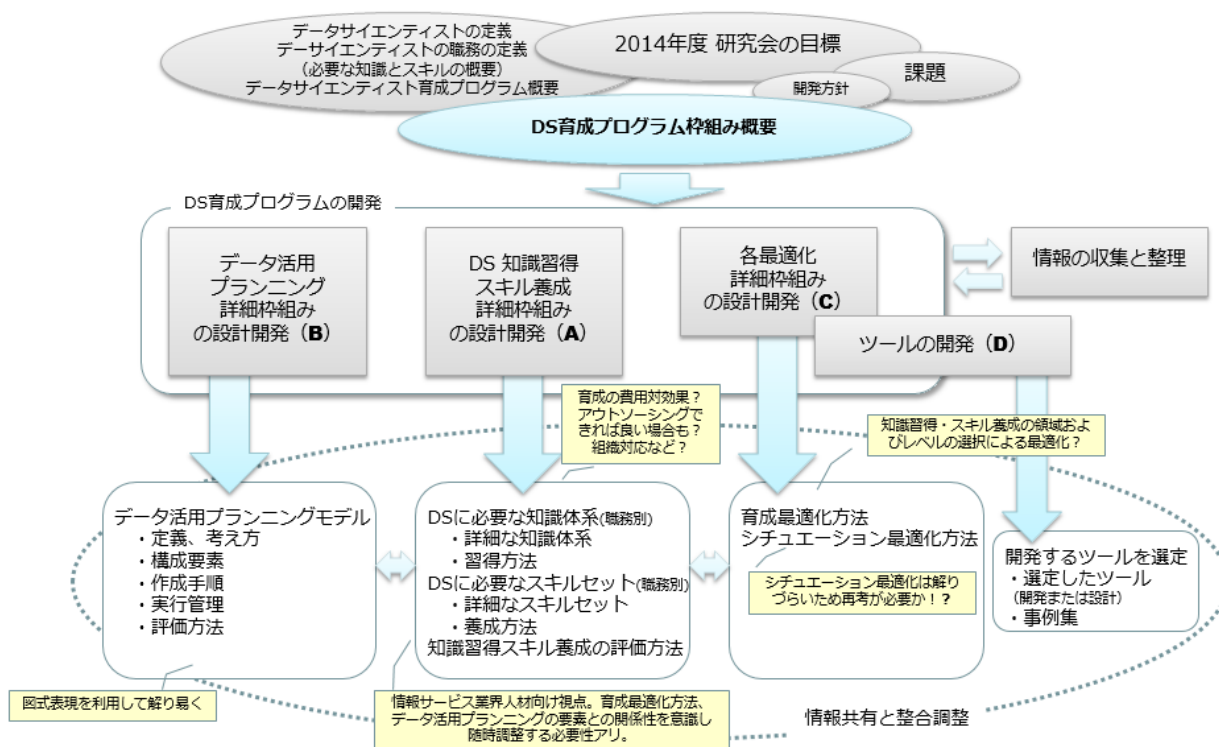
研究会の目的を達成するための手順を下図のように設定し、順次研究活動を進めた。



ワーキンググループは、テーマ毎に以下3チームを編成した。当初C・Dを別チームとしが、最終的にはひとつのチームとして活動した。各チームは数名で構成し、研究会の開催日以外にもそれぞれ独自に活動を行った。



ワーキンググループは以下の指針で活動を行った。



I. 概要

3. 目標

平成 26 年度の当研究会の目標は、4 ページの手順図のアウトプットとしてもあるように、「データサイエンティスト育成プログラム」の開発である。研究会メンバー目標達成のためにリソース、時間および期間を考慮し、プログラム全体をカバーすることを条件に表現粒度について調整しながら開発を行った。

情報システム（情報サービス）業が、ビジネスにおいてイニシアティブを取るためのデータサイエンティスト像を明確にし、その育成について概要を定める。



2013年度において、当該研究会はデータサイエンティスト育成のための枠組みを検討整理し、概要としてまとめる活動を行ってきた。2014年度は、2013年度の成果を基に情報サービス業界視点で捉えたデータサイエンティスト育成のための具体的な教育プログラムの開発を目指す。育成プログラムは、情報関連部門の人材がデータサイエンティストとして、企業経営や事業に対し、積極的にアプローチし利益貢献できる人材を育成するためのものとして開発する。併せてデータサイエンティスト事例集をとりまとめる。



【目標】

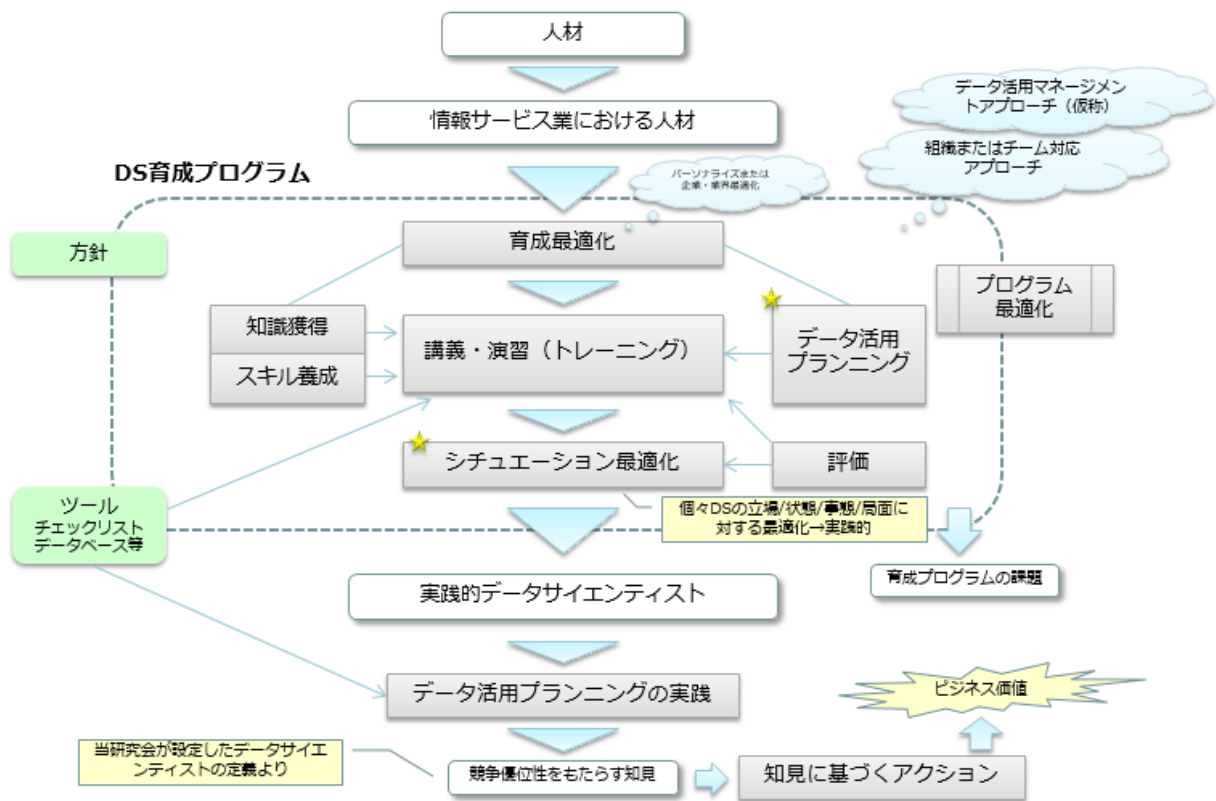
当研究会が定義する「データサイエンティスト」を育成するための教育プログラムを開発する。

（「データサイエンティスト育成プログラム」の開発）

I. 概要

4. プログラムについて

データサイエンティスト育成プログラムは、情報サービス業における人材を実践的データサイエンティストに育成するためのプログラムである。育成プログラムは、研究会で定義したデータサイエンティストを創出するための要素を手順も含め提示するものである。育成プログラムの領域と位置づけを概要図として以下に示す。



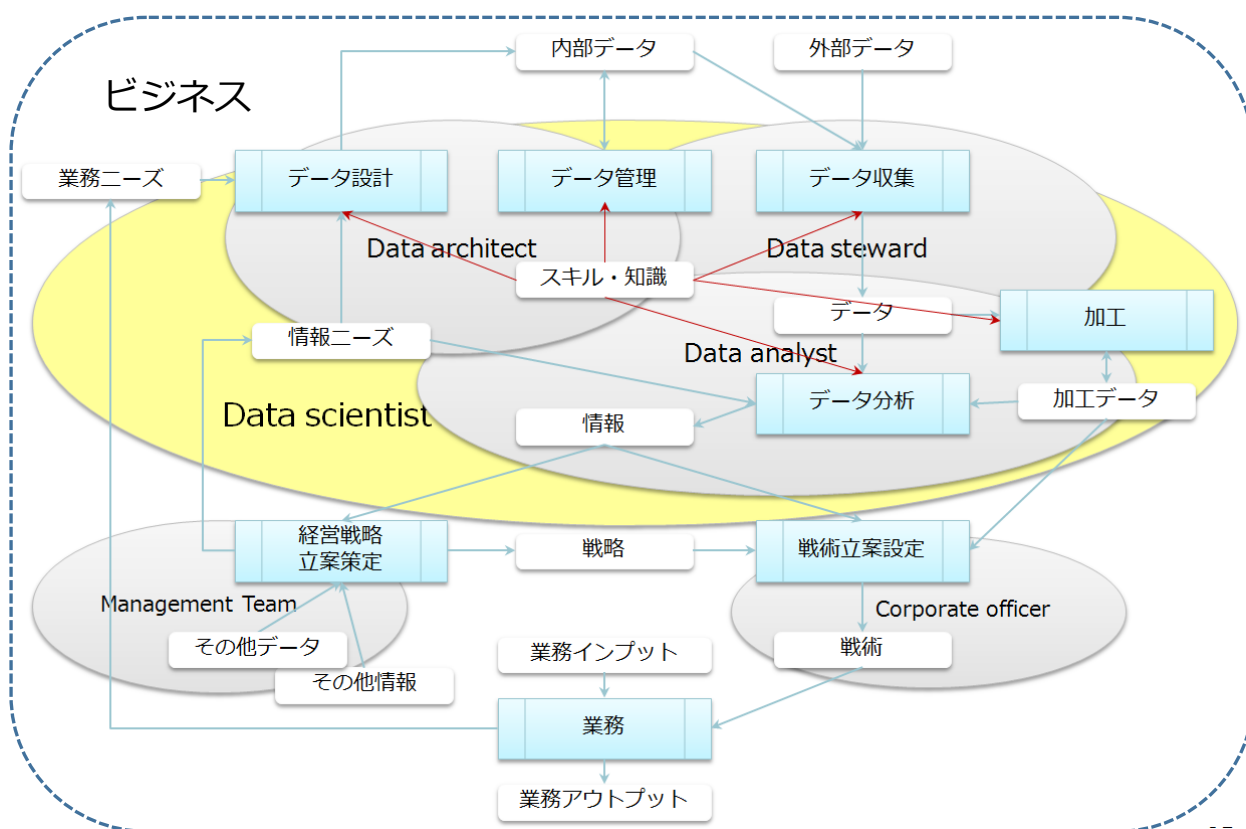
Ⅱ. プログラム詳細

1. 本プログラムが目指すデータサイエンティスト

データサイエンティストとは、

「ビジネスにおいて、競争優位性をもたらすために、データの収集・加工・分析に優れた専門性を発揮し、知見を引き出す人材」 である。

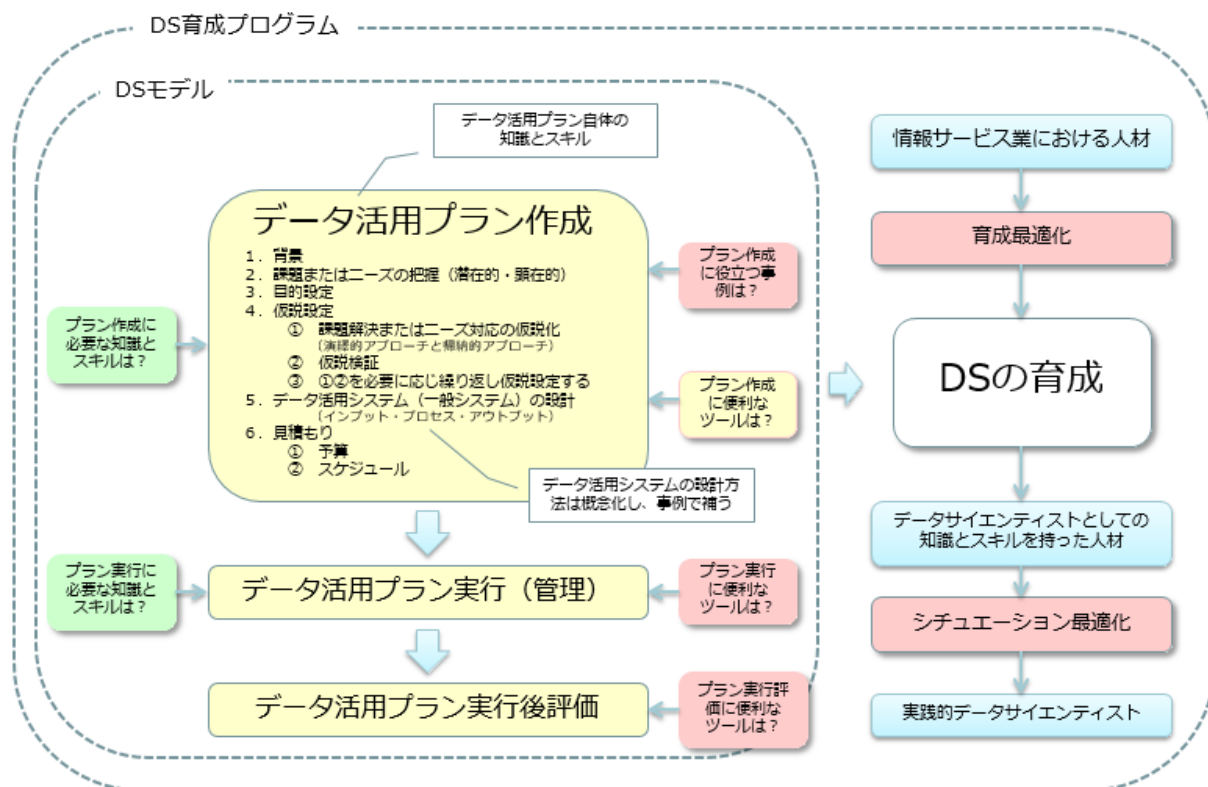
本研究会ではデータサイエンティスト（DS）を昨年度の活動で上記の通り定義した。本年度研究では、昨年度に定義した DS を目指すための育成プログラムを提示する。DS は、各方面で定義されている。表現はちがうもののそれぞれの視点での定義であり類似している。正しい定義というものがあるわけではない。参考までに、本研究会が DS を定義する際、検討のために作成した DS 領域についての概要図を以下に示す。



II. プログラム詳細

2. プログラムの構成

データサイエンティスト育成プログラムは、データサイエンティストのモデル（DSモデル）とその育成手順を提示することで構成している。DSモデルは、当研究会が平成25年度に定義した「職務」を行う人材の汎用的活動をモデル化したもので、データ活用プランを作成できる人材としている。DSはデータ活用プランを作成し、提案する。DSは、提案が決裁された場合、提案内容の実行管理を行い、実行後の評価までを職務とし遂行する。各職務に必要なスキルと知識については整理し、II.4.で提示している。DSモデルの人材育成には、対象となる人材毎に育成の効率化を配慮し、スキルの養成と知識の習得に関する範囲と内容を最適化する。さらに、実践的なデータサイエンティストになるためには、DSそれぞれにシチュエーション毎の最適化が必要である。



Ⅱ. プログラム詳細

3. データ活用プランニング

本節は、**B チーム（ワーキンググループ）** が担当

当該研究において、「データ活用プランニング」は、データサイエンティストが真っ先に行う職務（職務1）として定義している。言い換えればデータサイエンティストは「データ活用プランニング」を行える人材としている。そのことは、DSの定義における冒頭の「ビジネスにおいて・・・」で役割がより明確になる。企業や事業においてビジネスは、概ね計画に従い活動を行う。活動の中で競争優位性をもたらす知見を引き出すためにデータを活用する。そのために計画を立てる。それが「データ活用プランニング」である。

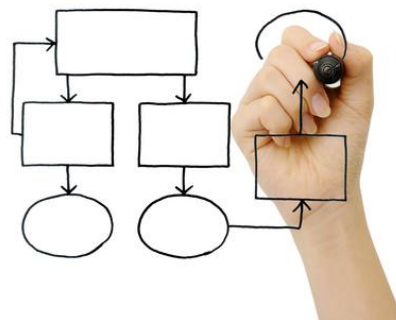


1) プラン作成の目的

プラン作成の目的は以下の通りである。

- ① 活動の目的に合致した知見を得るため、データ活用に関する方法、手順、リスク、必要資源などを明確にし、経営等に対してデータ活用活動投資の判断材料を提供する
- ② データ活用プロジェクトまたは業務の遂行に必要な情報を包括的に提示する
- ③ データ活用の問題や障害を確認し、データ活用活動の信頼性と生産性を高める

「データ活用プランニング」の手順は、当研究会メンバーが慣れ親しんでいるシステム設計概念を取り入れた。但し、ここでのシステムは「一般システム理論」が定義している概念的なシステム「システムは互いに作用している要素からなるもの」、「システムは目的に向かって動いている」の視点で使用している。一般的に、システム=コンピュータシステムとして捉えられることがほとんどのため、以降「活用プラン」でシステムを設計するとしていることがコンピュータシステムをプランニングし、それを提案することではないということに十分注意する必要がある。



2) データ活用プランニングの手順

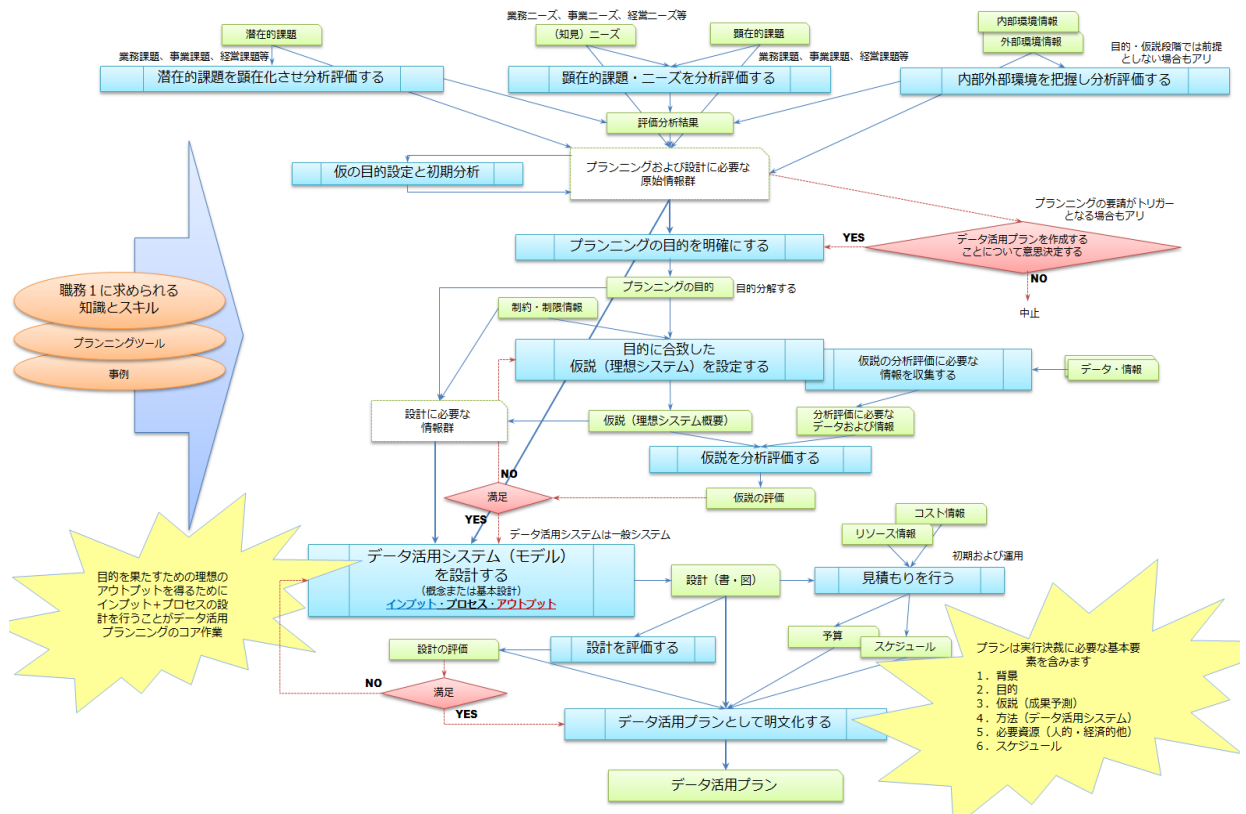
プランニングの手順は以下の通りである。

- ① プランニングの目的を明確にする
- ② ①の目的に合致した理想システムを設定する（理想システム≡仮説）
- ③ データ活用システムを設計する
- ④ ①～③を提案としてまとめる

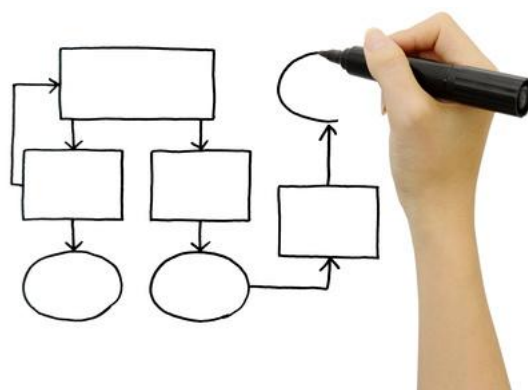
（ブレイクした手順は次ページに「データ活用プランニング手順概要」を参照）

【データ活用プランニング手順概要】

以下に示す手順概要は、インプット→プロセス（ファンクション）→アウトプットの形式でデータ活用プランニングの手順を示したものである。当研究会が検討した結果としての標準的な手順を示しており、実際に適用する際は、ケースによってモディファイしての活用となる。



当該「データ活用プランニング」では、「データ活用システム（前ページの図左下部）」を設計する手順を提示している。一般的なコンピュータシステム同様にアウトプットから設計する。目的に合致したアウトプットを設計する。その際、事例や経験をベースに理想的なアウトプットを想定し、それらに対し具体的で入手可能な内外データを摺り合わせる。アウトプットは、データや情報であったり知見であったり、その表現レベルはプランニングの目的により最適化する。アウトプット設計に当たっては、常に実施可能性を確認しながら行うため、インプットとプロセスの概要は想定しながら行う。この設計作業には、当該研究で定義した DS 職務の「知識とスキル」を持っていることが望まれる。また、「支援ツール」は、設計の生産性と信頼性を向上させる。



3) データ活用システムの設計手順

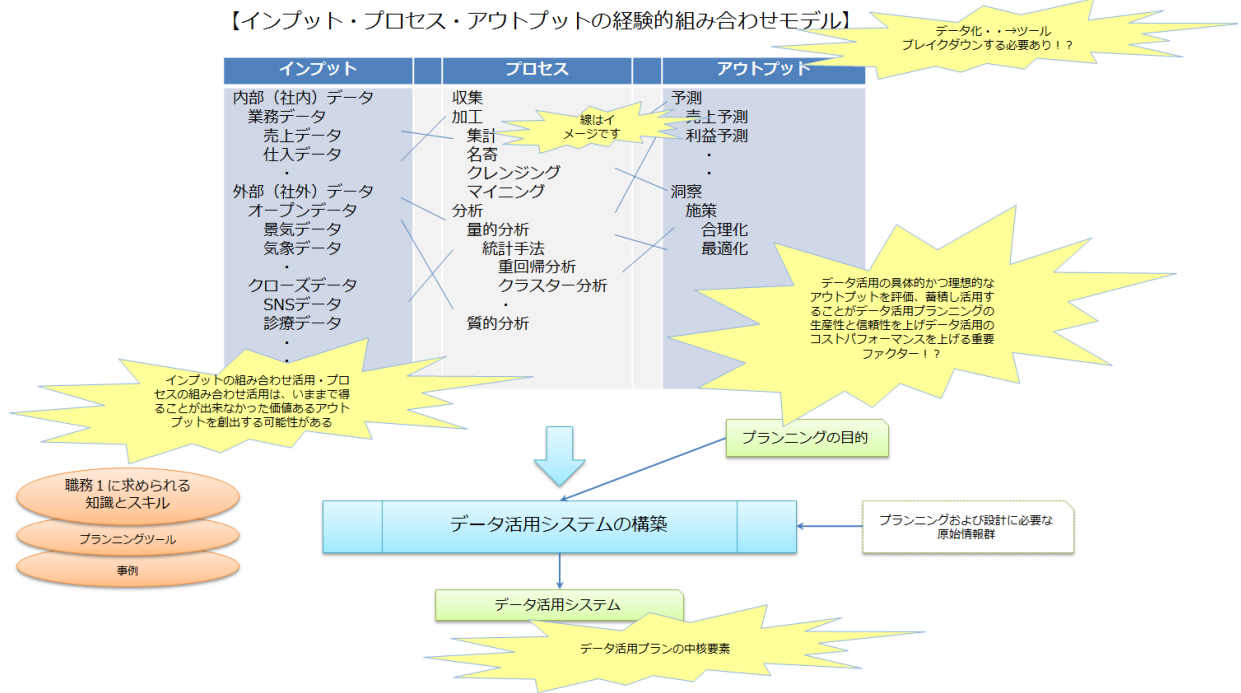
データ活用システムの設計手順は以下の通りである。

- ① プランニングの目的に合致したアウトプットを設計する
- ② 事例、経験、入手可能データによりインプットおよびプロセスを設計する
次ページ、「データ活用システムの設計イメージ」を参照

(設計手順のイメージは、次ページ「データ活用システムイメージ」を参照)

【データ活用システムの設計イメージ】

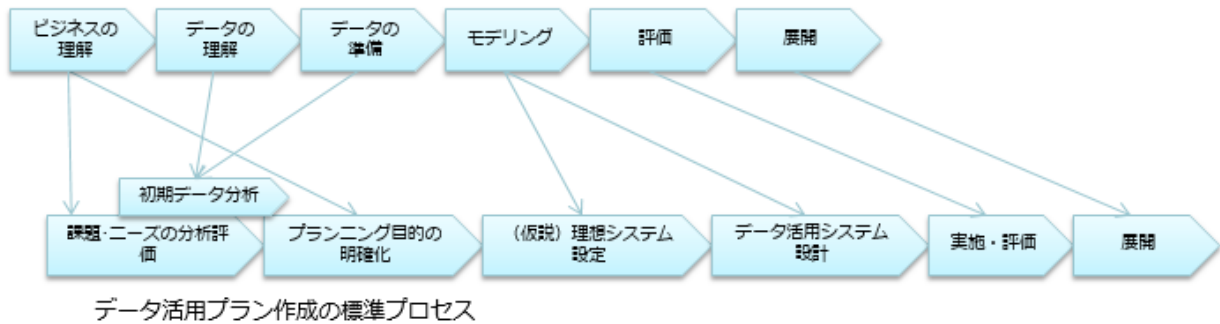
データ活用システムの設計は、活用事例などを参考にしつつ設計者の創造力を必要とする。そのためにDSは広い視野とゼネラリストとしてのスペシャリティが求められる。下図は、インプット、プロセス、アウトプットそれぞれの要素についてスキルと知識を持ち創造力で組み合わせを設定してシステム設計していくイメージを示している。



4) CRISP-DM との関連について

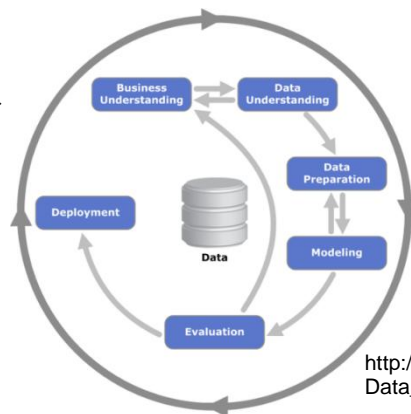
「データ活用プラン」の作成、実施・評価、展開のプロセスは、CRISP-DM（**Cross Industry Standard Process for Data Mining**）のプロセスに類似している。CRISP-DMはデータマイニングの標準プロセスである。データ活用プランニングは、DSの標準作業（活動）として定義している。プロセスの対応は下図の通りである。

CRISP-DMのデータマイニングの標準プロセス



【参考】

CRISP-DM のプロセス



http://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining#mediaviewer/File:CRISP-DM_Process_Diagram.png

Ⅱ. プログラム詳細

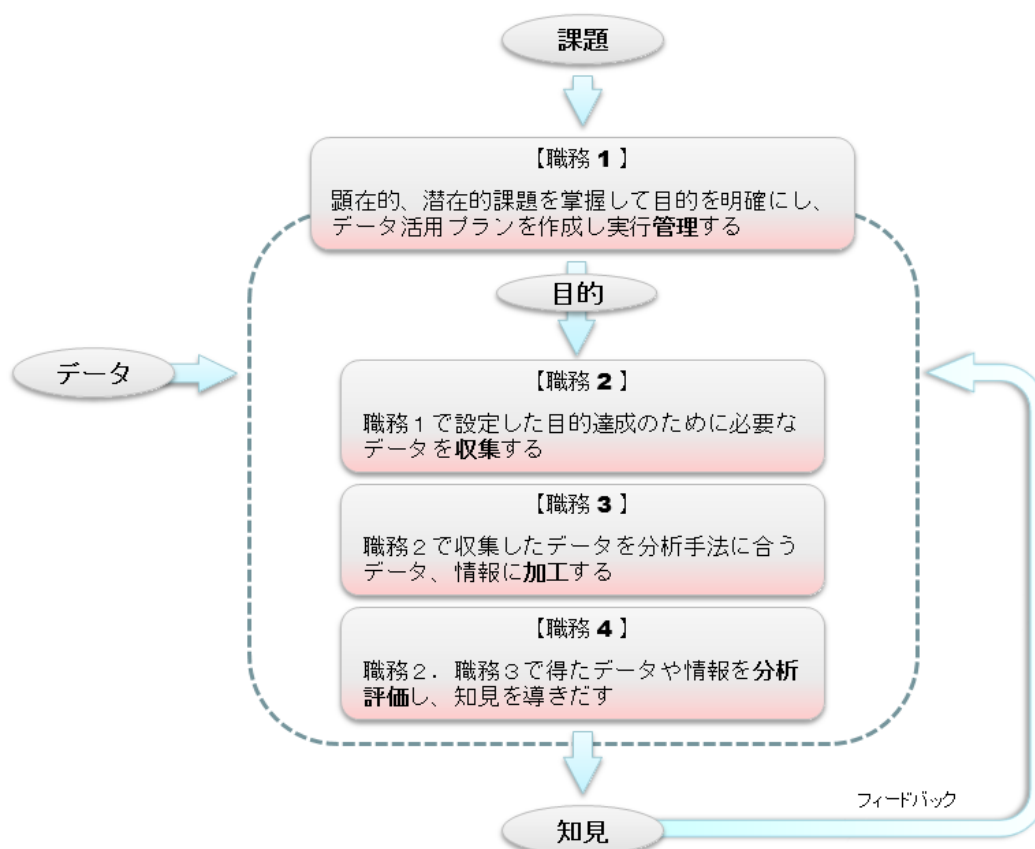
4. DSが必要とする知識とスキル

本節は、**A チーム（ワーキンググループ）**が担当

本ワーキンググループは、平成 25 年度データサイエンティスト育成研究会で定義した、データサイエンティスト（以下 DS）の「職務とタスク」「職務に求められる知識とスキル」をベースとして、「DS 知識習得スキル養成詳細枠組みの設計開発」を行った。

アウトプットのイメージとしては、情報サービス業界人材向け視点で、育成最適化方法、データ活用プランニングの要素との関係性も意識し、DS に必要な知識体系（職務別）の詳細な知識体系と学習目標を作成することにある。

以下、平成 25 年度に報告した資料（平成 25 年度データサイエンティスト育成研究会報告書より）



職務に対応した知識とスキル

職務 **1** 顕在的、潜在的課題を掌握して目的を明確にし、データ活用プランを立て、実行管理する

- 職務1のタスク
- | |
|-----------------------------|
| 1.1 目的を具体化し、明確にする |
| 1.2 目的達成のための作業設計を行い作業計画を立てる |
| 1.3 作業の実行を管理する |

職務 **2** 職務1で設定した目的達成のために必要なデータを収集する

- 職務2のタスク
- | |
|------------------------------|
| 2.1 必要データを策定する |
| 2.2 必要データの収集のための設計を行い、計画を立てる |
| 2.3 必要データを収集する |

職務 **3** 職務2で収集したデータを分析手法に合うデータ、情報に加工する

- 職務3のタスク
- | |
|-----------------------------|
| 3.1 収集データを評価し加工の必要性や方法を吟味する |
| 3.2 収集データの加工計画を立てる |
| 3.3 収集データを加工する |

職務 **4** 職務2、職務3で得たデータと情報を分析評価し知見を導き出す

- 職務4のタスク
- | |
|---------------------------------|
| 4.1 収集、一部加工されたデータと情報を合目的に分析評価する |
| 4.2 分析結果を知見として取りまとめる |
| 4.3 知見を合目的にまとめ提示する |

職務に求められる知識とスキル

職務 1 顕在的、潜在的課題を掌握して目的を明確にし、データ活用プランを立て、実行管理する

職務 1 のタスク	1.1 目的を具体化し、明確にする
	1.2 目的達成のための作業設計を行い作業計画を立てる
	1.3 作業の実行を管理する

■ 知識

- 1K1 経営の知識
- 1K2 業界の知識
- 1K3 業務の知識
- 1K4 マーケティングの知識
- 1K5 ITの知識
- 1K6 分析のための知識
- 1K7 プロジェクト管理の知識

■ スキル

- 1S1 戦略立案スキル
- 1S2 IT活用スキル
- 1S3 プロジェクト管理スキル
- 1S4 コミュニケーションスキル

職務 2 職務 1 で設定した目的達成のために必要なデータを収集する

職務 2 のタスク	2.1 必要データを策定する
	2.2 必要データの収集のための設計を行い、計画を立てる
	2.3 必要データを収集する

■ 知識

- 2K1 業務の知識
- 2K2 ITの知識
- 2K3 データの知識

■ スキル

- 2S1 データの収集/整理スキル
- 2S2 大規模データ、特殊データの処理基盤スキル
- 2S3 ITスキル

職務 3 職務 2 で収集したデータを分析手法に合うデータ、情報に加工する

職務 3 のタスク	3.1 収集データを評価し加工の必要性や方法を吟味する
	3.2 収集データの加工計画を立てる
	3.2 収集データを加工する

■ 知識

- 3K1 業務の知識
- 3K2 ITの知識
- 3K3 データの知識

■ スキル

- 3S1 データの可視化スキル
- 3S2 大規模データ、特殊データの処理基盤スキル
- 3S3 ITスキル

職務 4 職務 2. 職務 3 で得たデータと情報を分析評価し知見を導き出す

職務 4 のタスク

- | |
|---------------------------------|
| 4.1 収集、一部加工されたデータと情報を合目的に分析評価する |
| 4.2 分析結果を知見として取りまとめる |
| 4.3 知見を合目的にまとめ提示する |

■ 知識

- 4K1 経営の知識
- 4K2 業界の知識
- 4K3 業務の知識
- 4K4 マーケティングの知識
- 4K5 ITの知識
- 4K6 分析のための知識

■ スキル

- 4S1 戦略立案スキル
- 4S2 プレゼンテーションスキル
- 4S3 IT活用スキル
- 4S4 分析スキル

これらの職務ごとのタスクを行うために必要となる“知識”“スキル”をベースに本ワーキンググループでは、職務で求められる知識とスキルの“知識”“スキル”の項目に対して発散思考で具体的な“知識項目”“スキル項目”を洗い出し、洗い出した項目の収束（とりまとめ）を行い、それぞれに必要な“知識項目”“スキル項目”（スキルセット）を明確化した。

次に、できあがった“知識項目”“スキル項目”（スキルセット）を体得することを目的として、各スキルに対して“スキル習得の学習目標”を設定した。

“スキル習得の学習目標”の設定に当たっては、ID（インストラクショナルデザイン）手法を参考にし、“条件”“対象”“基準”“タスク”を意識して作成した。

また、取りまとめにあたっては、i コンピテンシ・ディクショナリ（IPA）も参考に作成した。

【結果】

「職務 1 実施にあたってのスキル習得の学習目標」を以下の通り設定した。

職務1 顕在的、潜在的課題を掌握して目的を明確にし、データ活用プランを立て、実行管理する

職務1のタスク

- 1.1 目的を具体化し、明確にする
- 1.2 目的達成のための作業設計を行い作業計画を立てる
- 1.3 作業の実行を管理する

知識	知識項目		
1K1 経営の知識	経営分析手法		
1K2 業界の知識	対象業界の動向		
1K3 業務の知識	対象業務の知識		
1K4 マーケティングの知識	マーケットリサーチ手法	マーケット・顧客分析	
1K5 ITの知識	データベース知識	データ分析ツール	
1K6 分析のための知識	各種統計手法	データマイニング	
1K7 プロジェクト管理の知識	プロジェクト企画・計画	プロジェクト実行管理	

スキル	スキル項目			
1S1 戦略立案スキル	マーケットリサーチ	顧客要件分析	プロモーション立案	戦略立案手法
1S2 IT活用スキル	データマイニング処理	集計・統計の基礎知識		
1S3 プロジェクト管理スキル	PMBOK			
1S4 コミュニケーションスキル	コミュニケーション	ファシリテーション	プレゼンテーション	ネゴシエーション
	課題発見力			

スキル習得の学習目標

1S1 戦略立案スキル	データ分析ツール・手法を十分理解した上で活用し、経営レベルの要求に応じた戦略を立案できる
1S2 IT活用スキル	立案した戦略に対して、実践できるレベルでIT活用プランを立てプロジェクト化することができる
1S3 プロジェクト管理スキル	プロジェクト管理手法に基づき対象プロジェクトに対して求められるQCDを達成するためのマネジメントができる
1S4 コミュニケーションスキル	戦略目標達成に向けステークホルダー、プロジェクトメンバーとプロジェクト運営に必要なコミュニケーションが図れ、計画通りにプロジェクトを進められる

「職務2 実施にあたってのスキル習得の学習目標」を以下の通り設定した。

職務2 職務1で設定した目的達成のために必要なデータを収集する

職務2のタスク

- 2.1 必要データを策定する
- 2.2 必要データの収集のための設計を行い、計画を立てる
- 2.3 必要データを収集する

知識	知識項目	
2K1 業務の知識	データの選定	必要データの入手
2K2 ITの知識	データベース知識	各種データソース関連知識
2K3 データの知識	データの構成要素	

スキル	スキル項目			
2S1 データの収集/整理スキル	情報収集力			
2S2 大規模データ/特殊データの処理基盤スキル	地図データ	画像データ	動画データ	音声データ
	ソーシャルデータ	クラウド技術		
2S3 ITスキル	データマイニングツール	統計分析ツール	情報集計ツール	

スキル習得の学習目標

2S1 データの収集/整理スキル	分析に有効となるデータを明確にし、対象のデータの入手方法を確保し、入手したデータを加工・分析しやすいように整理することができる
2S2 大規模データ/特殊データの処理基盤スキル	分析の目的達成のために必要となる大規模データ/特殊データを蓄積・加工するためのITインフラを構築もしくは用意できる
2S3 ITスキル	各種データ分析ツール・手法を用いて、大規模データ/特殊データを加工・分析、収集し管理することができる

「職務3実施にあたってのスキル習得の学習目標」を以下の通り設定した。

職務3 職務2で収集したデータを分析手法に合うデータ、情報に加工する

職務3のタスク

- 3.1 収集データを評価し加工の必要性や方法を吟味する
- 3.2 収集データの加工計画を立てる
- 3.2 収集データを加工する

知識	知識項目		
3K1 業務の知識	必要データの加工		
3K2 ITの知識	データ編集・加工ツール	データ集計ツール	データマイニングツール
3K3 データの知識	データの構成要素		

スキル	スキル項目			
3S1 データの可視化スキル	データの表・グラフ化	データ可視化ツール		
3S2 大規模データ、特殊データの処理基盤スキル	地図データ	画像データ	動画データ	音声データ
	ソーシャルデータ	クラウド技術		
3S3 ITスキル	データマイニングツール	統計分析ツール	情報集計ツール	

スキル習得の学習目標

3S1 データの可視化スキル	各種理論、ツールを用い収集・加工したデータを判断、意思決定しやすく可視化することができる
3S2 大規模データ、特殊データの処理基盤スキル	目的達成のために大規模データ/特殊データをデータの構成要素を見極め分類、整理ができる
3S3 ITスキル	各種データ編集/加工ツール、データ集計ツール、データマイニングツール等を利用して、可視化したデータをそれぞれの特徴を理解したうえで加工できる

「職務4 実施にあたってのスキル習得の学習目標」を以下の通り設定した。

職務4 職務2. 職務3で得たデータと情報を分析評価し知見を導き出す

職務4のタスク

- 4.1 収集、一部加工されたデータと情報を合目的に分析評価する
- 4.2 分析結果を知見として取りまとめる
- 4.3 知見を合目的にまとめ提示する

知識	知識項目			
4K1 経営の知識	経営戦略手法			
4K2 業界の知識	対象業界の動向			
4K3 業務の知識	対象業務の知識			
4K4 マーケティングの知識	マーケットリサーチ手法	マーケット・顧客分析	マーケット戦略	
4K5 ITの知識	プレゼンツールの知識	統計分析ツール	情報集計ツール	データマイニングツール
4K6 分析のための知識	各種統計手法	戦略を意識した分析手法		

スキル	スキル項目			
4S1 戦略立案スキル	マーケットリサーチ	顧客要件分析	プロモーション立案	戦略立案手法
4S2 プレゼンテーションスキル	プレゼンソフト	ドキュメント技法	プレゼン技法	
4S3 IT活用スキル	データマイニングツール	統計分析ツール	情報集計ツール	
4S4 分析スキル	課題発見力	各種統計手法	データ分析ツール	データマイニング

スキル習得の学習目標

4S1 戦略立案スキル	課題と目的を把握し経営戦略手法を用い分析結果を戦略に活かせる形に反映することができる
4S2 プレゼンテーションスキル	課題と目的を把握し分析結果を企業の戦略に合った知見にまとめ、論理的に、分かり易く説明することができる
4S3 IT活用スキル	分析に必要な各種情報技術、ツールを用いそれぞれの特徴を理解したうえで分析結果・検定結果を分かり易く表現できる
4S4 分析スキル	各種統計手法や、戦略を意識した分析手法の知識をベースに、ステークホルダー等の意思決定に直接役立つ知見を提供できる

これらの結果は、データサイエンティストのキャリアレベル判断の基準として使用したり、データサイエンティスト育成にあたっての育成方針設定、研修開発の指針として活用していただけることを目的としている。

今後の課題としては、知識項目、スキル項目の粒度にはばらつきがあり、今後精査の余地が残っている。

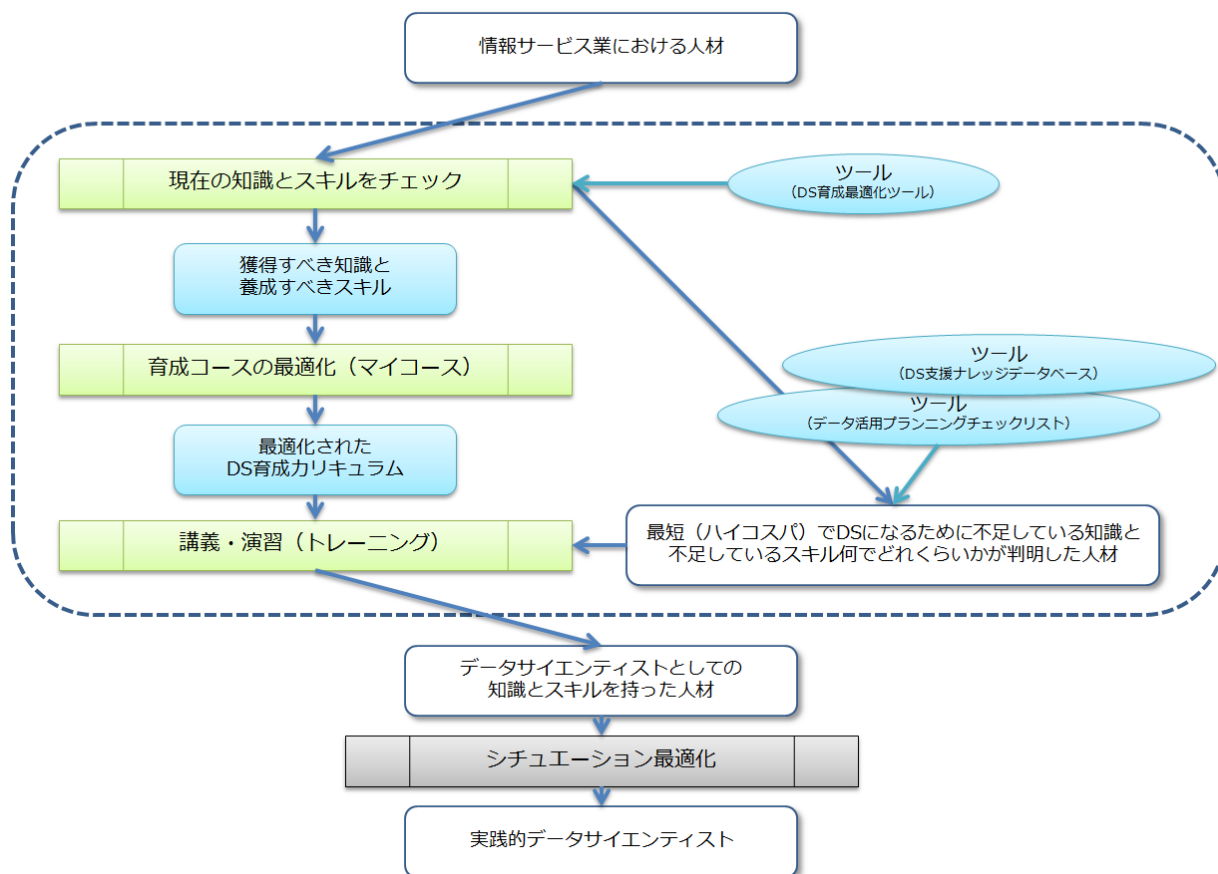
また、達成レベル計測を意識した学習目標とはなっていないものも有る。

今後は学習目標がデータサイエンティスト育成の条件の多くを満たせるように精度をあげていくことが望まれる。

II. プログラム詳細

5. DS 育成手順

情報サービス関連に携わる人材に対し、データサイエンティスト（当該研究会の定義）として、企業経営や事業に対し積極的にアプローチして利益貢献できるように必要な基礎知識を提供し実践できるようにトレーニングを実施するものである。その手順を以下に示す。



Ⅲ. 支援ツール

本章は、C・Dチーム（ワーキンググループ）が担当

本研究会では、情報サービス業の人材をDSとして育成する為の研究を行った。そして、先の「Ⅱプログラム詳細」で説明した通り、本研究会が考えるDSとは何か。また、DSとして必要な知識、スキルは何かを明確にしてきた。「Ⅲ 支援ツール」では、DSを育成する為にどのような「支援ツール」が必要であるか検討した。各ツールについては、そのツールの必要性に関して説明するものであり、ツール提供に向けては、今後精査しなければならない点が多々ある。また、いくつか具体的な名称をあげて説明した部分は、例として記したものである。

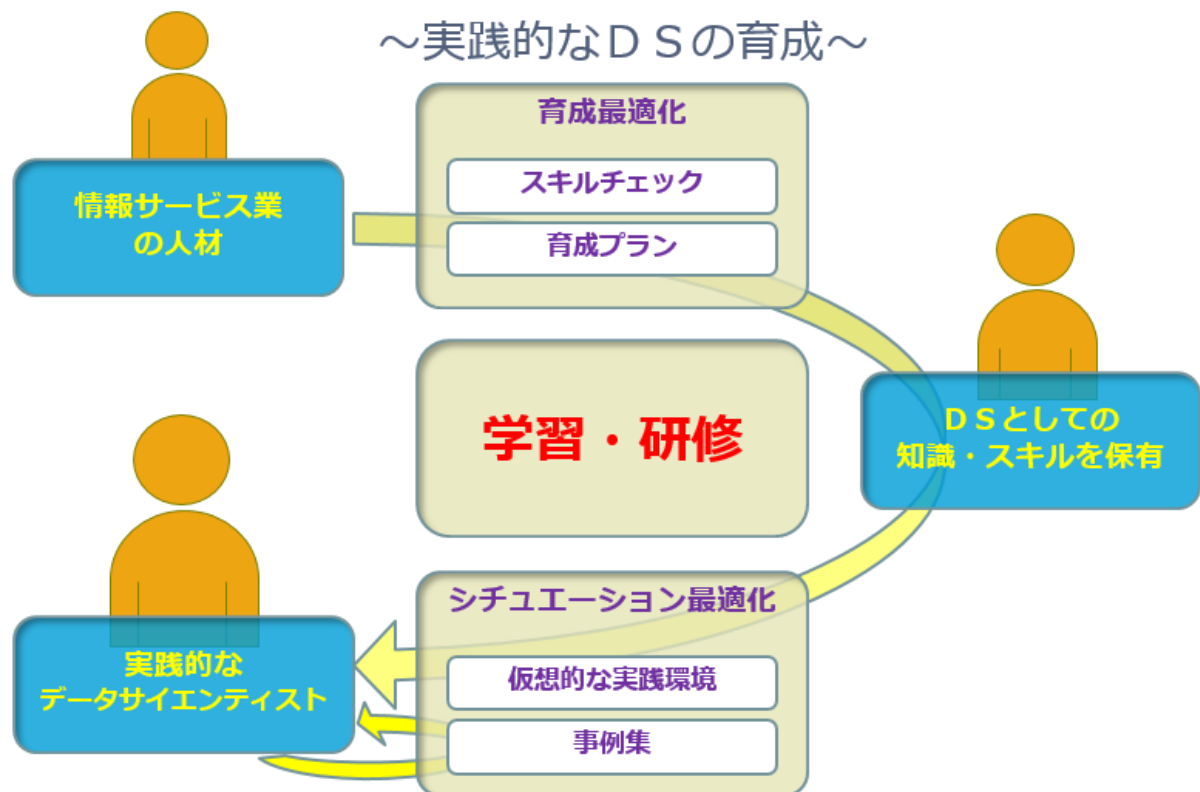
育成に必要な支援ツール

即戦力として活躍するDSを本研究会では「実践的なデータサイエンティスト」と定義している。支援ツール検討チーム（C・Dチーム）では、情報サービス業の人材を「実践的なデータサイエンティスト」に育成する為にどのようなツールが必要か検討した。

DS育成に必要な支援ツールの整理にあたり、DSとしての知識・スキルに乏しい人材をDSとしての知識・スキルを保有した人材にステップアップさせるまでの部分を最適に行う為の方策を「育成最適化」と表現している。その部分で必要なものは「スキルチェック」と「育成プラン」という2つのキーワードである。

さらに、実践的なDSに育成させる、または、DSのレベルアップを最適に行う事を「シチュエーション最適化」と表現した。そこで必要となるのが、「仮想的な実践環境」と「事例集」の2つのキーワードである。

以降、計4つのキーワードについて検討した内容を順に報告する。



1. スキルチェック (育成最適化)

DSとして必要な知識・スキルを習得するために、まず、各技術者のスキルレベルを認識する必要がある。さらに、各技術者に即した育成プランを組み立てて学習する事が必要である。また、一定レベルのDSであったとしても、更にレベルアップするために現時点のスキルを確認する事は重要である。そのようなことから最適な育成のためのツールとして「スキルチェック」が必要であると考えている。

スキルチェック用のツールの方向性としては、本研究会においてDSの職務を1～4に定義し、各職務に対応した知識・スキルを整理した。その流れからスキルチェックも職務毎のチェックシートとして用意する事を基本に考えている。ツールは、それらのチェック要素に関して、自己判断形式でチェックするツールを想定している。その特徴は以下の通りである。

- ① スキルチェックはポータルサイトのサービスとして公開する。
- ② 評価したデータはポータルサイト内に管理され随時更新できる。

参考にチェックシートのイメージをご紹介します。評価基準などは今後十分な検討が必要だと認識している。

職務 1：顕在的、潜在的課題を掌握して目的を明確にし、データ活用プランを立て、実行管理する

分類	基準	評価要素	配点	評価
1S1 戦略立案スキル	データ分析ツール・手法を十分理解した上で活用し、経営レベルの要求に応じた戦略を立案できる。	必要な知識・スキルを保有し実務経験が充分にある	5	
		必要な知識・スキルを保有し若干の実務経験がある	4	
		必要な知識・スキルは保有するが実務経験がない	3	
		知識・スキルを習得中である	2	
		知識・スキルを習得していない	1	
1S2 IT活用スキル	立案した戦略に対して、実践できるレベルでIT活用プランを立てプロジェクト化することができる。	必要な知識・スキルを保有し実務経験が充分にある	5	
		必要な知識・スキルを保有し若干の実務経験がある	4	
		必要な知識・スキルは保有するが実務経験がない	3	
		知識・スキルを習得中である	2	
		知識・スキルを習得していない	1	
1S3 プロジェクト管理スキル	プロジェクト管理手法に基づき対象プロジェクトに対して求められるQCDを達成するためのマネジメントができる。	必要な知識・スキルを保有し実務経験が充分にある	5	
		必要な知識・スキルを保有し若干の実務経験がある	4	
		必要な知識・スキルは保有するが実務経験がない	3	
		知識・スキルを習得中である	2	
		知識・スキルを習得していない	1	
1S4 コミュニケーションスキル	戦略目標達成に向けステークホルダー、プロジェクトメンバーとプロジェクト運営に必要なコミュニケーションが図れ、計画通りにプロジェクトを進められる。	必要な能力があり実務経験が充分にある	5	
		必要な能力があり若干の実務経験がある	4	
		必要な能力はあるが実務経験がある	3	
		必要な能力を習得中である	2	
		必要な能力を習得していない	1	

2. 育成プラン (育成最適化)

スキルチェックにより、技術者が習得すべき知識・スキルの整理が可能となる。次に必要な事は育成プランの組み立て、つまり、どのようにすればその知識・スキルが習得できるかといった部分となる。

技術者が育成プランを組み立てる際、どのようにすれば知識・スキルの習得が可能なのかをアドバイスする支援が必要であると考えている。

ツールとしての役割は、知識・スキル習得の為の情報を幅広く提供することある。その中から、各技術者が適した方法を選択できるような環境作りをツールとして想定し、発信する情報は、以下のものを想定している。

- 1) 教材の紹介
- 2) セミナーの紹介
- 3) 関連資格・認定試験の紹介

3. 仮想的な実践環境 (シチュエーション最適化)

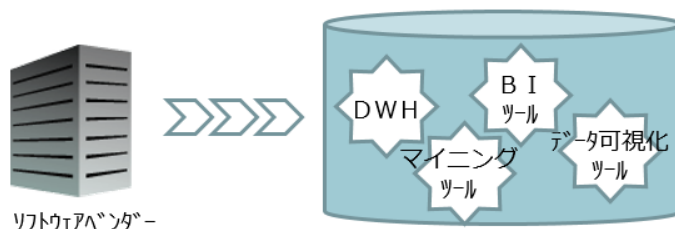
DS として必要な知識が身に付いた人材が、さらにスキルアップする為に必要な事は実践的な経験であると考えている。また、一定レベルのDSであったとしても、新たな業務を行う際、業種・業務・分野が異なることで、実務で役立つ環境・ソフトウェア・ツールなどは異なる事が多いと考えられる。様々な状況のなかでDSとして活躍する為には、実務に近い環境で実際に試すことを支援する仕組み・環境が必要であると考えている。

ツールとして考える支援環境の方向性は以下の通りある。

- 1) 各種ソフトウェア・ツールが体験的に利用できる
- 2) 疑似的なデータを使い、実験的に利用することができる

(データマイニング系)

- ・ R言語統合開発ツール
- ・ RapidMiner
- ・ カスタマーリングス
- ・ BIRT Analytics
- ・ Visual Mining Studio
- ・ など



(DWH・BI・データ活用系)

- ・ AGRA
- ・ Dr.Sum EA
- ・ QlikView
- ・ InfoFrame DWH Appliance
- ・ Pentaho
- ・ Apache Hadoop
- ・ AnalyticMart
- ・ など



(注意)

上記のご紹介は例として記載するもので、記載するものが全てではなく、また、見直しが必要である点に注意下さい。

4. 事例集 (シチュエーション最適化)

DS が行うデータ活用プランニングや、プラン実行において、最新の技術や最適なツールを用いる事は非常に重要である。また、プランを検討するうえで、成功、失敗を問わず先行事例は大いに役立つ情報と言える。各種事例を共有し、効率化に役立てる事が重要である。

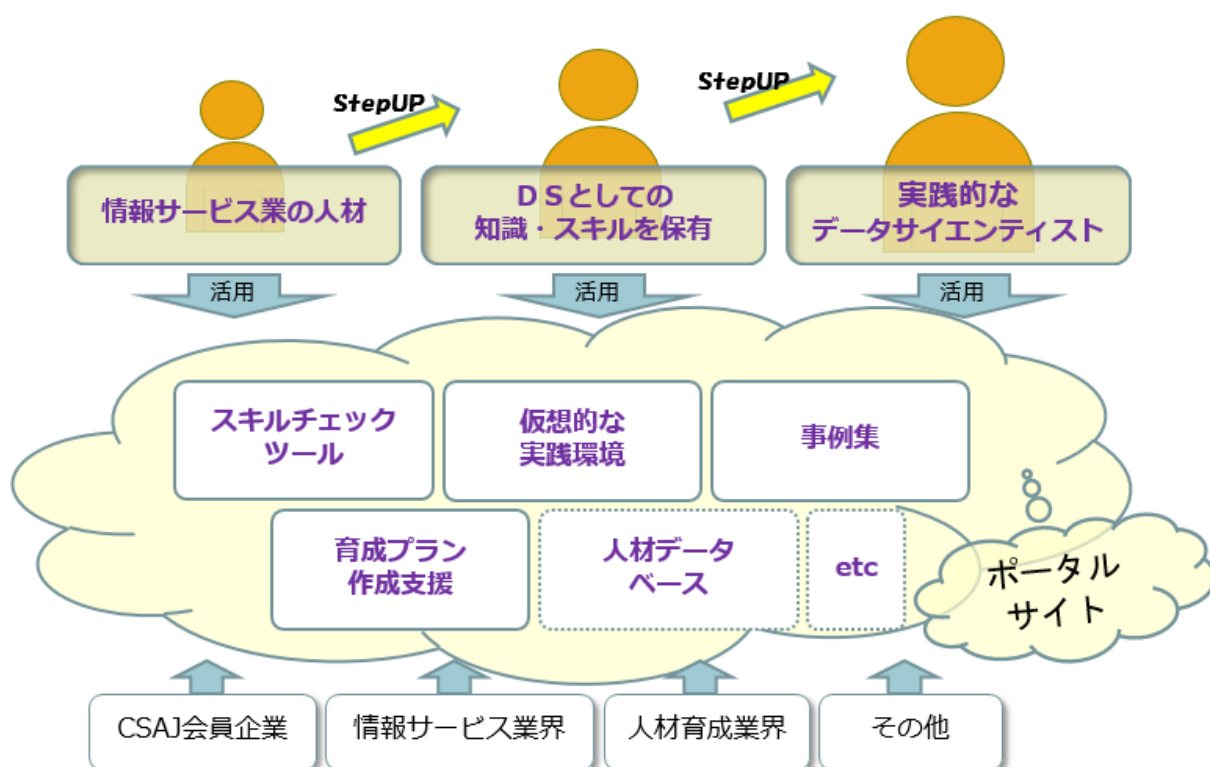
事例を参考にする側が容易に目的の事例を見つけ出せる仕掛けが必要である。

- 1) 各事例をカテゴリ分けしスムーズな閲覧を可能とする
- 2) 事例の登録に手間が掛からない仕掛けとする
- 3) 目的の事例が簡単に検索できる

ポータルサイトの位置づけ

これまで説明した4つのツールの提供方法は、利便性の面から、ポータルサイトを構築してサービスする事が望ましい。例外的に Web からの操作ではサービスできない部分もあると思うが、原則的にはサイトのサービスとして提供する。当面のサービス内容は、本研究会の主旨に沿って「DS 育成の最適化に必要なサービス」が中心と考えているが、将来的には DS のための総合ポータルサイトとして位置づけられるように発展させることも重要である。

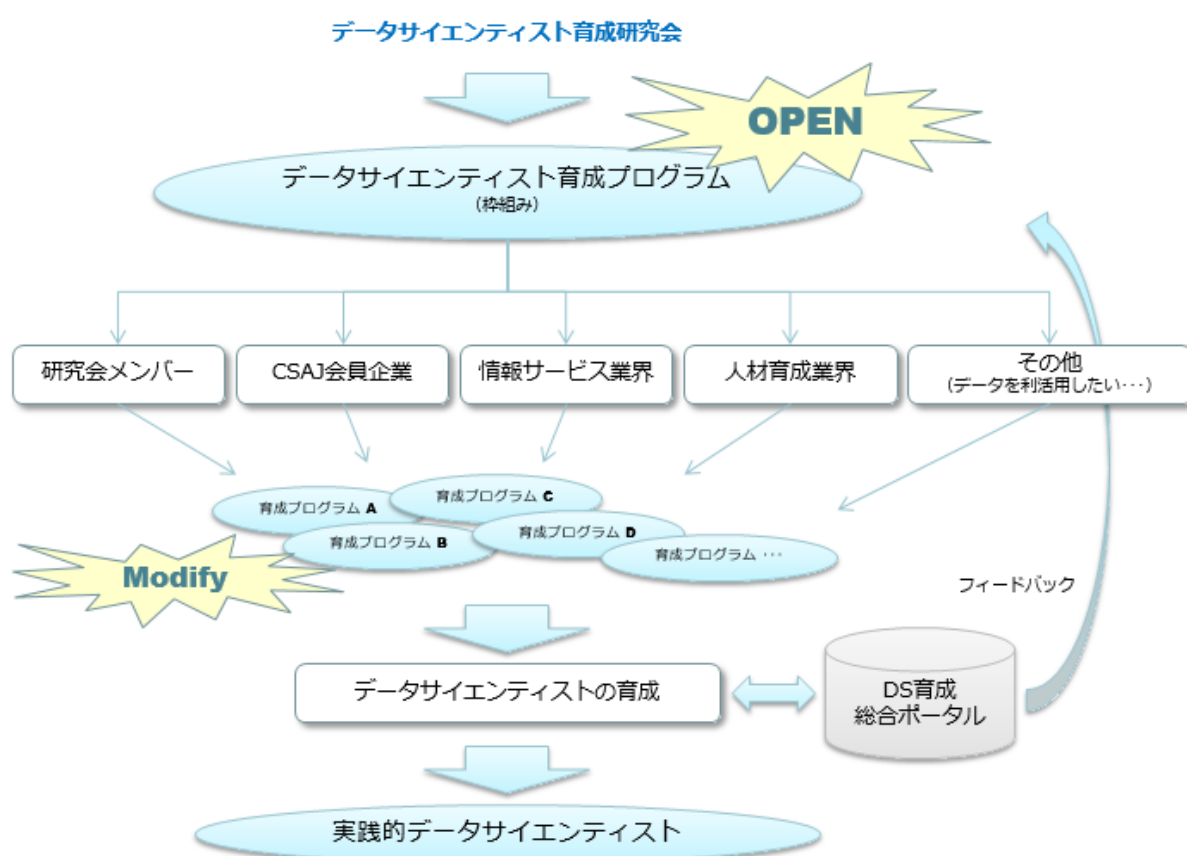
DS、関連ソフトウェア製品の提供ベンダー、DS を必要とするエンドユーザ、各関係団体など各ステークホルダーがお互いに利用価値のあるサイト構築を目指すべきである。



IV. 補足

1. プログラムの活用

当該「データサイエンティスト育成プログラム（枠組み）」は、情報サービス業視点で提示させていただいたもので、研究会の外に対してもオープンなものとしている。プログラムの活用については自由にモディファイして活用していただきたい。



IV. 補足

2. 育成プログラムの課題

研究会活動の中で、解決していないまたは検討が不十分な内容については課題とした。以下に課題を列挙する。これら課題は、本研究活動の成果物の一部として今後の活動に引き継ぐことを望む。

1. データサイエンティストの活用対象データはビッグデータだけではない

データサイエンティストにスポットライトが当てられた背景にビッグデータがあるためビッグデータを分析し活用する人材として捉えられることが多い。当該研究会において、ビッグデータは重要な活用対象であるが、その他のデータも活用対象としてきた。今後、活用対象データについては明示して種々対応することが必要かもしれない。十分議論されていないので今後の課題としたい。

2. データサイエンティストに求められる知識およびスキルは広範である

データサイエンティストという人材を定義したが、その知識およびスキルは広範となり、それらすべてを一人で網羅し、獲得することは可能かもしれないが現実的ではない。今後の検討の中で組織対応するなどを検討し、その管理と一部の役割をデータサイエンティストに求めるなどその辺り明確にすることを今後の課題としたい。

3. 知見を引き出した後のフィードバックを含めた対応についてはどう扱うか

知見を引き出した後工程について十分議論されていないので今後の課題としたい。

4. 継続的なデータ資産活用のため、メタデータを集中管理する必要がある

データ活用のための、現在にマッチしたデータ辞書の整備が重要。

5. データ活用プランニングは、情報システム設計開発的アプローチをしているため新規システム開発提案と見なされやすい

データ活用プランニングをデータ活用システムの設計としたため、一般的なコンピュータシステムの設計開発と扱われる可能性が高い。データ活用システムは、一般システム概念であり「データを活用するしくみ」のことである。

6. データサイエンティストに求められる知識とスキルは環境変化に対応してアップデートが必要である

知識とスキルは、当研究会の DS 定義と職務に対応したもので環境変化、適用対象によって随時アップデートが望まれる。

V. 考察

平成 26 年度「データサイエンティスト育成研究会」は、25 年度の成果を引き継ぎ「データサイエンティスト育成プログラム」の詳細な枠組みを提示することを目標に進めた。研究成果は、世の中で定義が定まっていない「データサイエンティスト」について情報サービス業（業界）視点で定義し、その人材をモデル設定して育成するという枠組みで表現した。

データサイエンティスト（以降 DS）について、本研究会が定義した内容からデータ活用プランニングができる人材としてモデル化した。プランニングは、企業活動・事業活動等ビジネスにおいては必須の作業であり、DS の重要な職務である。DS の職務は、プランが決裁され実行される際、少なくともその管理と成果の提示までをその領域とする。プランの大小やアプローチ内容等により、組織的な対応が必要なことは DS のカバーすべき領域の広さから自明である。このような DS に必要なスキルと知識について、本研究では職務に対応して検討し設定した。これらの設定は、DS の育成に指針を与える。支援ツールは DS 育成を文字通り支援する。

以上の観点でまとめた「DS 育成プログラム」は、DS の定義からの演繹的アプローチを中心に行ったため成果は総じて概念的かつ概要レベルである。今回のアプローチ方法については、リソースおよび時間的な制約から適当であったと評価している。概要レベルであることについても同様との認識である。成果物の不完全さについては、今後このプログラム活用される方々が補正していくことを望む。

これまで情報システム業から顧客に対するアプローチは、業務プロセスの改善が大半であり、その評価は業務の生産性や信頼性の向上によるコスト削減であった。本プログラムにより育成する DS はこれまでとは顧客に対するアプローチがちがう。本研究成果の活用により育成された情報サービス業の新たな人材が、「データ活用プランニング」というアプローチを積極的に行い、顧客に価値ある知見を提供できることを願う。願いが叶ってはじめて本研究の価値を見いだすことができる。

おわりに

本研究会は、メンバーの総意をベースとし、すべてメンバーの経験と英知により結果を導き出しました。その結果、前年度同様に既成概念に囚われることなく自由な発想で議論を行うことができました。今回の研究会成果が CSAJ 会員、研究会メンバー、情報サービスに関わる方々に少しでもお役に立てば幸いです。

最後に、研究会の開催、推進にあたりましては CSAJ 様関係各位ならびにご多忙の中ご参加いただいた会員各位に多大なご支援ご協力を賜りました。この場をお借りして御礼申し上げます。ありがとうございました。

研究会メンバー（敬称略）

（主査）

昆 凡子 株式会社シムコス

（メンバー：五十音順）

青木 明彦	株式会社算法
岡木 勇	株式会社フォーラムエイト
小倉 健一	日本システム開発株式会社
加藤 秀治	リコージャパン株式会社
木村 雅晴	株式会社シムコス
国井 勝則	オデッセイヒューマンシステム株式会社
小南 利仁	株式会社システム・ロジックス
昆 充生	株式会社シムコス
武田 大輔	リコージャパン株式会社
武智 秀昭	株式会社パスカリア
谷川 智久	日本システム開発株式会社
中原 崇	株式会社パスカリア
西村 紀彦	株式会社ミック
日角 健二	株式会社シムコス
本田 義久	リコージャパン株式会社
丸山 啓之	株式会社富士通マーケティング
水野 保	ジャパンラーニング株式会社
山田 晃嗣	株式会社シムコス

（オブザーバ）

佐々木 宏	立教大学経営学部経営学科（教授）
丹下 博詞	アグラ株式会社（代表取締役）
獵古 敏明	日本システムランド株式会社

（事務局）

小林 哲 一般社団法人コンピュータソフトウェア協会

平成 26 年度データサイエンティスト育成研究会 報告書
平成 27 年 3 月 17 日 第 1 版

一般社団法人コンピュータソフトウェア協会 (CSAJ)
〒107-0052
東京都港区赤坂 1-3-6 赤坂グレースビル 4F
TEL : 03-3560-8440 FAX : 03-3560-8441 <http://www.csaj.jp/>