

2018年12月7日10:00～
札幌第1合同庁舎2階講堂

総務省「政策評価に関する統一研修」

評価とは何か

同志社大学政策学部・大学院総合政策科学研究科

山谷(やまや)清志

もくじ

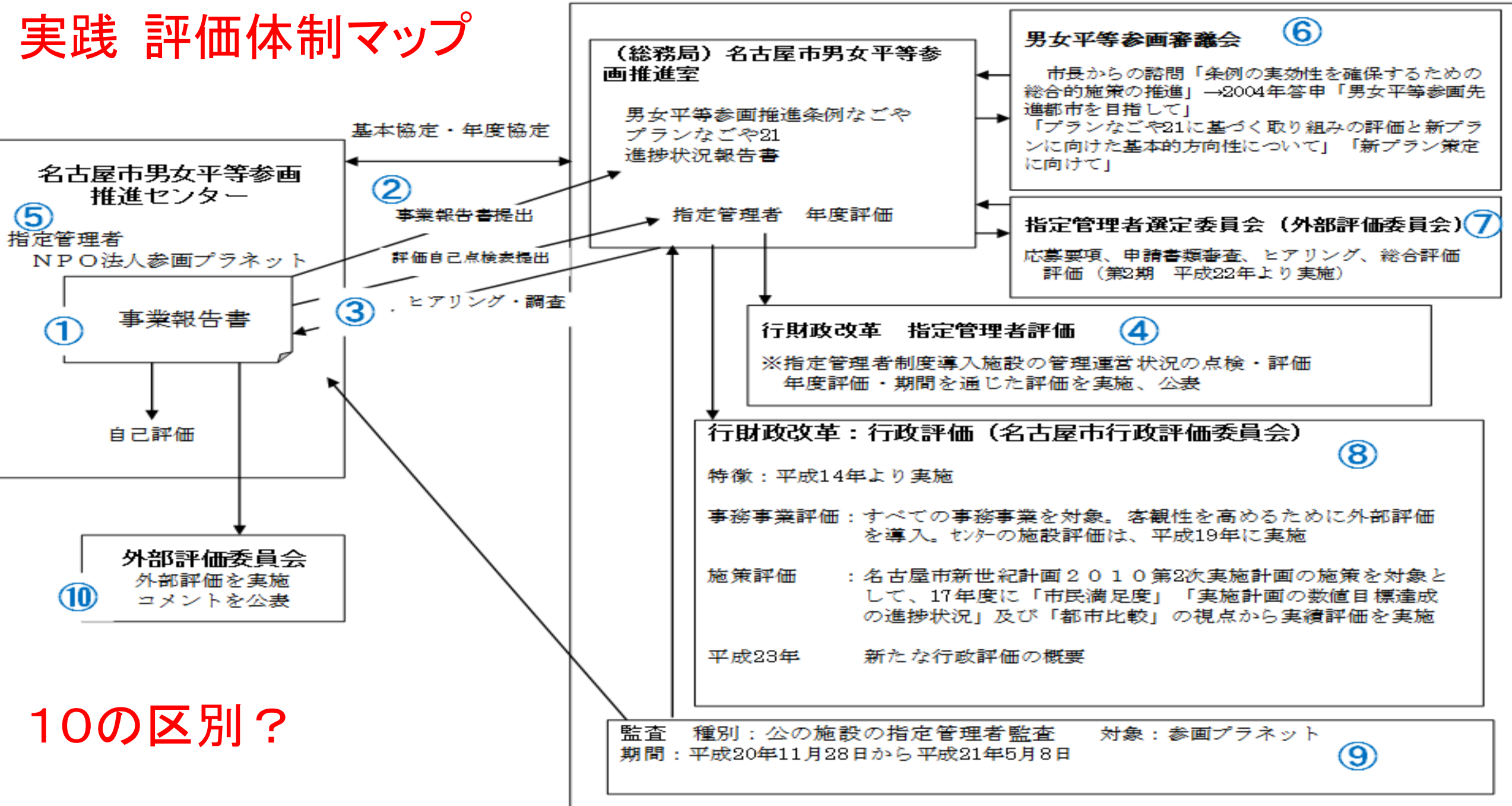
- I あらためて「評価」とは
- II プログラム評価について
- III 業績測定について

I あらためて「評価」とは

...

Evaluation revisited

実践 評価体制マップ



10の区別？

1. 評価とは

「評価は物事のメリット、値打ち、意義を体系的に明らかにすることである」 (Scriven, 1991:139)

“Evaluation is the systematic determination of merit, worth and significance of things”

評価 = 事実特定 + 価値判断
(Evaluation) (Factual identification) (Value determination)

客観的な情報を正しく提供

出所: 佐々木亮(2015)評価士養成講座テキスト

2. 評価の目的(1)

① 形成的評価 (Formative evaluation)

「コックが調理中にスープを味見すること」

- ・事業や活動の開発・**形成**
- ・継続的**改善**、効率、節約、生産性
- ・**立案**、**実施**段階で行う評価

② 総括的評価 (Summative evaluation)

「客が出来上がったスープを味わうこと」

- ・事業や活動の**成果**の把握、影響監視
- ・プログラム継続・縮小・拡大の判断
- ・**一定期間経た段階**で行う評価

相互に補完し合うことで、より効果的なプログラム形成につながる



2. 評価の目的(2)

- 評価情報を**何のために**使うのか (Scriven, 2004)
 - 事業や活動の改善を目的としたもの(形成的評価)
 - 介入のアカウンタビリティの確保を目的としたもの (総括的評価)
 - 科学的な知識・知見を生み出すことを目的としたもの

その他にもあるが、重要なことは、

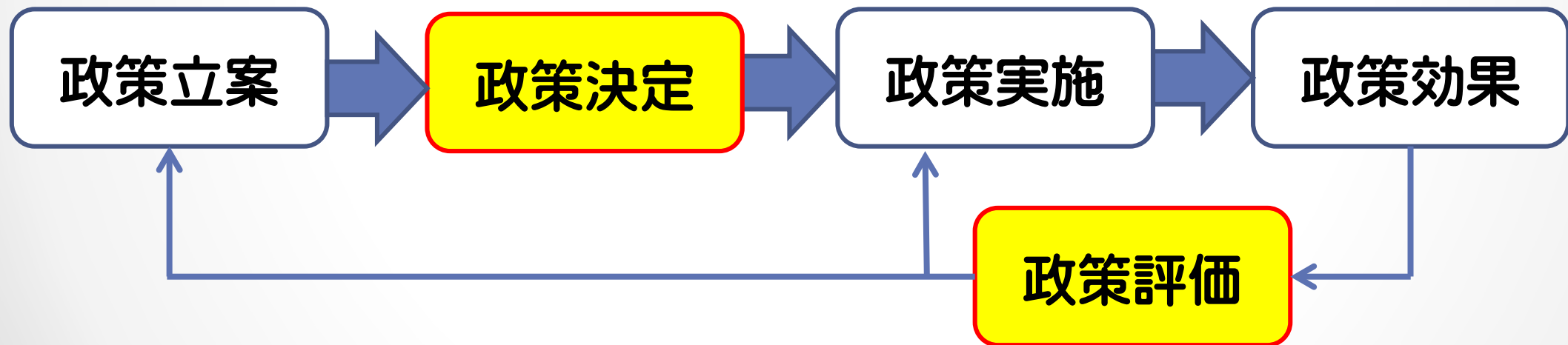
- **評価の「目的」**をまず明らかにすること。
- それに合わせて**評価方法**(誰が、いつ、どのような手法で行うのか)を選択する必要がある。

3. 政策評価の目的

●「政策評価」とは何か？

○政策評価活動は、意思決定者に「情報提供活動」の一種です。

→「政策評価」は、「政策決定」とは異なります。



- 政策実施の改善へのフィードバック ➡ 形成的評価
- 政策立案へのフィードバック ➡ 総括的評価

4. 評価の流れ ①～⑩

① 評価の目的決定

② 評価の対象選定

③ 評価のコスト

④ 評価の基本的方針 外部による統制・内部の質保証、
定量・定性、透明性

⑤ 評価の実施体制 内部評価 外部評価、第三者の有
識者の選定

評価の流れ 続き

- ⑥ 評価規準：有効性、効率、節約、公平性、持続可能性、適正な関連 (relevance)
- ⑦ 評価の実施方法 評価プロセス、評価マニュアル、ガイドライン
- ⑧ 評価の基本スケジュール
- ⑨ 評価結果 公表方法と情報公開
- ⑩ 評価の活用とフォローアップ

Ⅱ プログラム評価について

...

Program Evaluation

5. プログラム (Program) とは

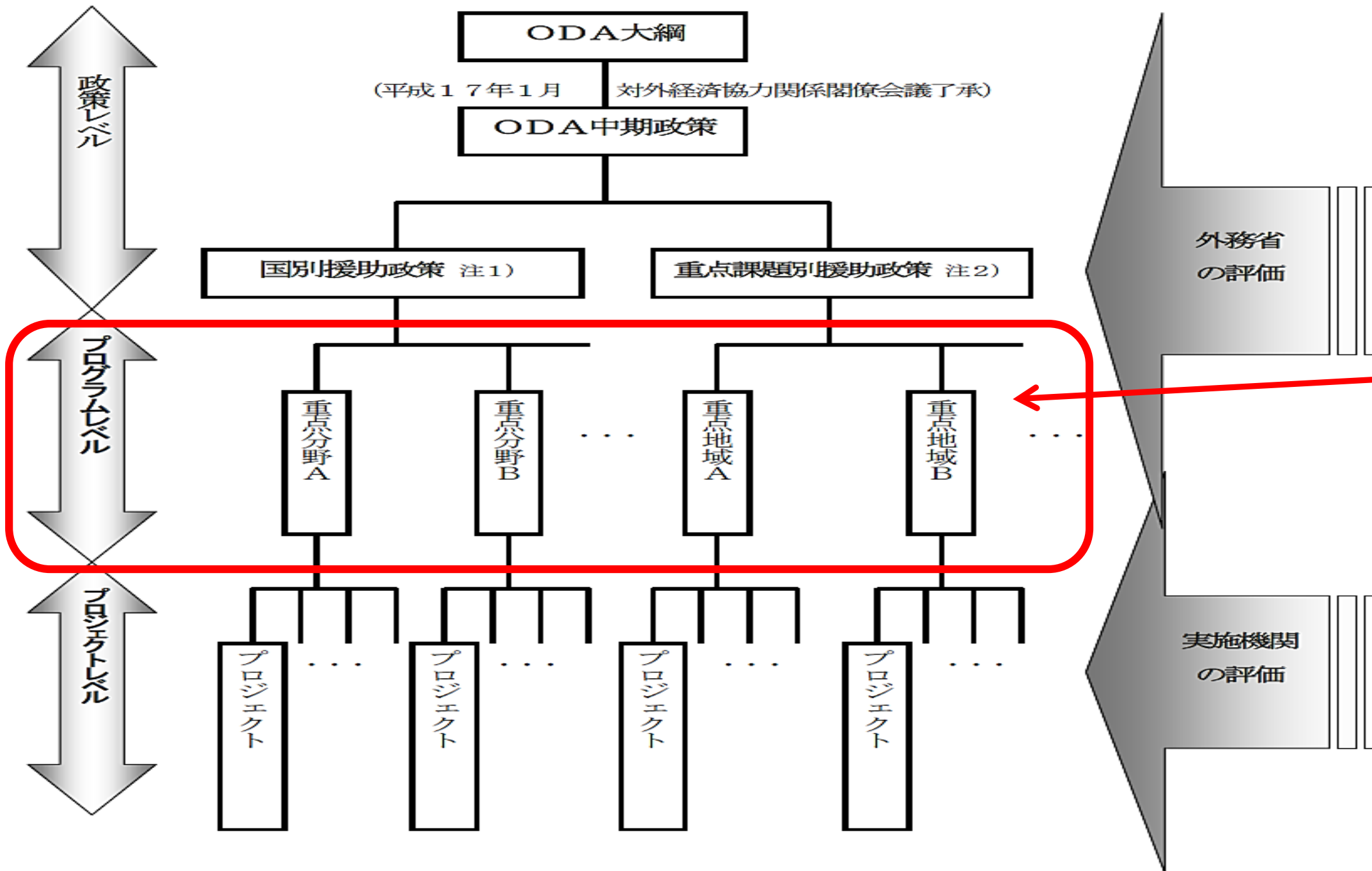
- プログラムとは、**ある社会目的を達成するための一連の活動群**をさし、それらの活動を**実施に導くルール**、制度、組織体制や人材を含む取組み全体をさすものである。
- 特定の組織が主体となって実施する場合もあれば、複数の組織・関係者が協働で実施する取組みもある

プログラム 要するに

- 何をしたい、問題は何か **what** ニーズ
- どうやってやるのか **how** 政策手段5つ
- コストは誰が負担するのか **how much**
- どこでやるのか **where**
- いつまでやるのか **when** スケジュール
- なぜやるのか **why** **accountability**
- 見込みはあるのか **feasibility** **sustainability**

現場の専門 プログラム

- 「塀の中」 高齢者の社会復帰を支えるプログラム 福祉知識 会話 体操
- JAXA 宇宙ビジネス重視へ 宇宙基本計画工程表の見直し
- 与野党 医療と介護 消費増税にむけ大まかな改革項目と実施時期を列挙した 社会保障改革プログラムを閣議決定 2012年8月21日
- 沖縄科学技術大学院大学
- 岩手県 International Linear Collider: 国際リニアコライダー 「宇宙誕生と物質構造の謎を探る、素粒子物理学の国際共同プロジェクト」
- 青森県 国際熱核融合実験炉 (ITER)



ODAのプログラム

プログラム：政策手段と政策デザイン

政策目的

① 分配（公共財の提供）

再分配（福祉）

② 規制および規制緩和

③ 広報、PR、教育、ナッジ

④ 税制（租税政策・減税）

⑤ 制度創造、改変（地方分権、民営化、介護保健制度など）

6. プログラム評価の定義

- プログラム評価とは社会調査の方法を活用し、社会プログラムによる介入の有効性を体系的に調査し、評価を行うこと。その評価は、プログラムを取り巻く政策的・組織的な文脈を考慮した方法で行われるもので、社会状況を改善するための活動の情報源となるものである(Rossi, Lipsey & Freeman, 2004: 16)
- 評価とは、プログラムや政策の改善に寄与するための手段として、明示的または暗示的な規準と比較しながら、プログラムや政策の実施やアウトカムを体系的にアセスすること(Weiss, 1998:4)

ことば

総括する (summarize)

評定する (rating)

優劣をつける (ranking)

比べる (compare)

見積もる・価格づける (estimate)

格付・等級づけ (rating)

定量化する (quantify)

事実関係を書き記す (describe)

全体を総覧する (review)

分類する (classify)

はかる (測る・量る measure)

総合調査する (survey)

調査研究する (research)

試す (test / experiment)

精査する (investigate)

検査する (audit)

振り返る (reconsider)

「評価する」・褒める (appreciate)

値踏みする (appraise)

事前予測 (forecast)

事前評価 (assess)

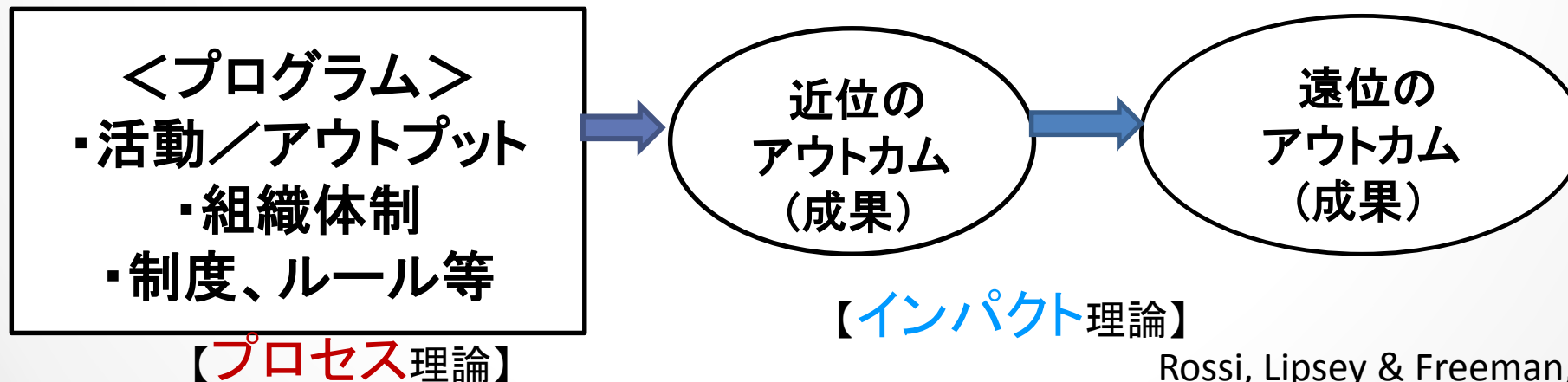
監理 (monitor)

監査 (audit)

監察 (inspection)

7. プログラム理論 (Program Theory)という考え方

- ・「プログラムの実施と成果との間に存在する媒介メカニズム」(Weiss)
- ・プログラムの作戦や活動内容を決定するにあたり活用できる考え方であり、「めざす成果を得るためには、おそらくこういう活動が必要であろう」という「仮説」である
- ・その「仮説」が有効か、有効にするにはどうしたらよいか、を問う
- ・「インパクト理論」と「プロセス理論」のふたつの理論が包含される



Rossi, Lipsey & Freeman, 2004: 140

8. ロジック・モデル

●ロジック・モデルとは何か？

→プログラム理論を踏まえてプログラムを表現する汎用的枠組み(世界標準)である。

●なぜ、ロジック・モデルが必要なのか？

→プログラムの運営方法や実施プロセスを明らかにすることで**ブラック・ボックス評価**を回避できる。(インパクト理論のみならずプロセス理論の明確化)

→介入のメカニズムを明らかにすることで体系的な**評価(5階層)**が可能になる。

→目的の設定、指標の検討を行うときの枠組みとして活用できる。

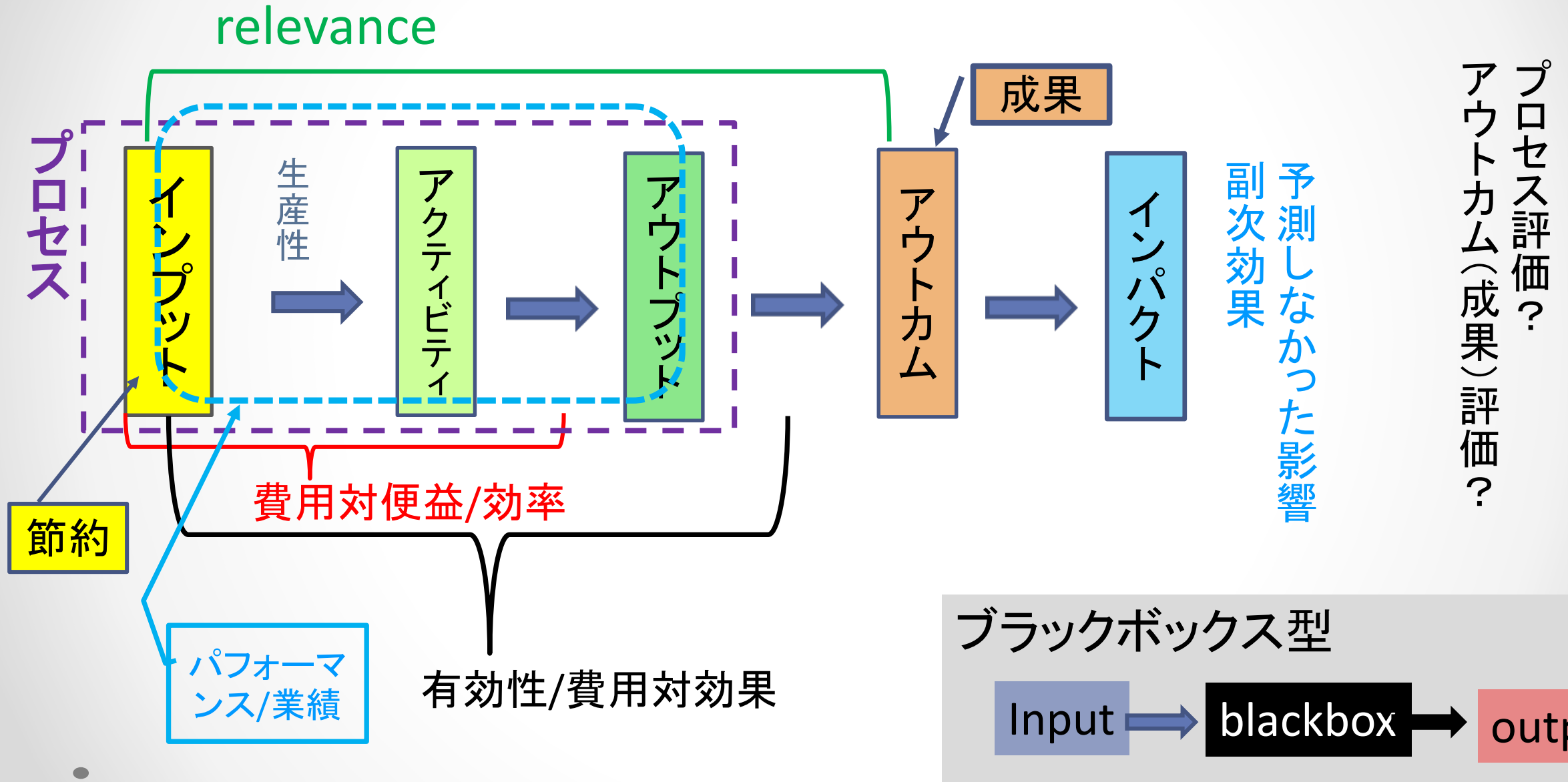
→ステークホルダー間で、プログラムのコンセプト、作戦を検討し、共有するためのツールとして活用できる

●ロジック・モデルの構成要素は何か？

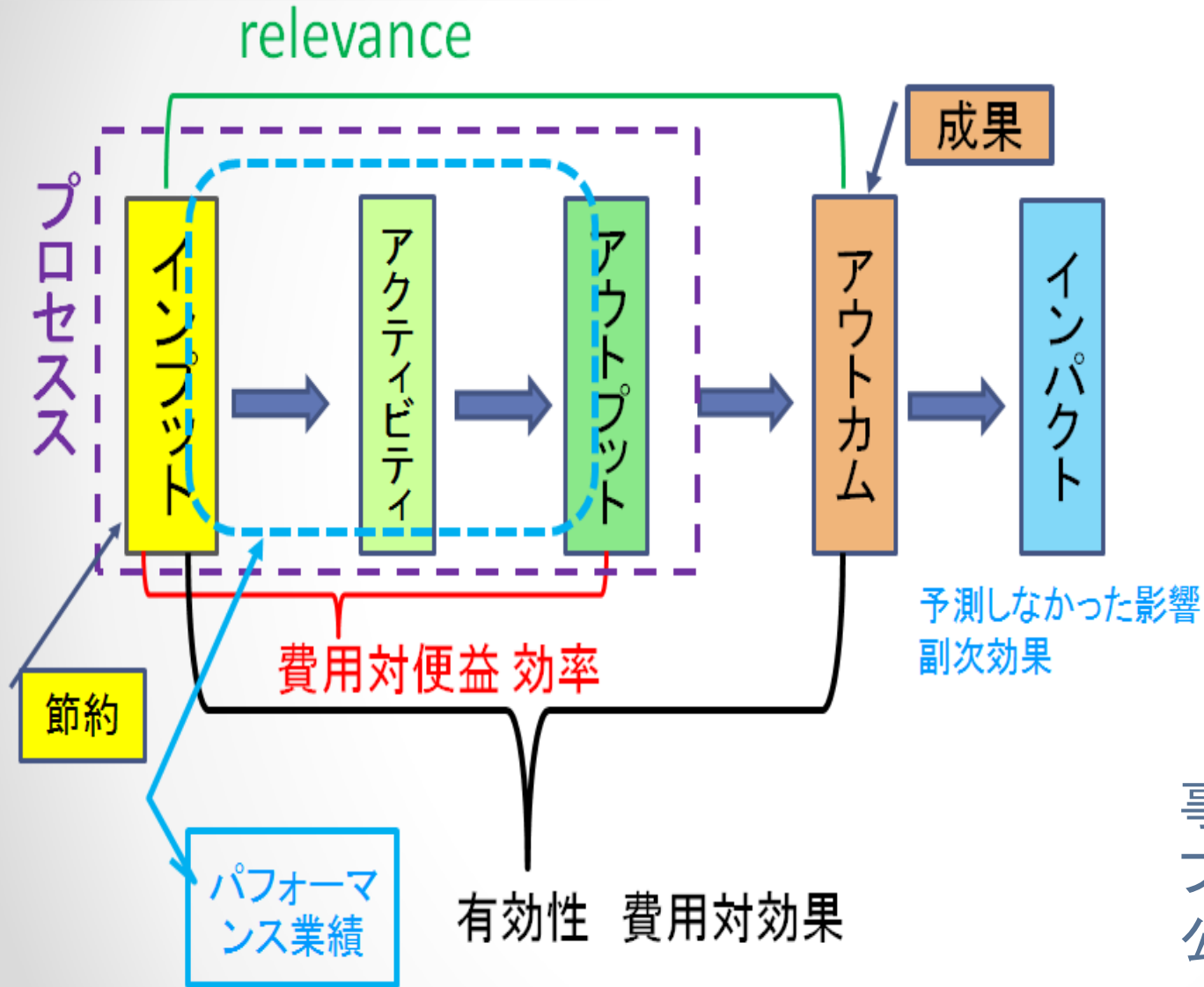
→4つの基本要素でプログラムを表現するものである。

「インプット」「アクティビティ」「アウトプット」「アウトカム」

ロジックモデルの考え方



成果指標 測定指標 実績指標

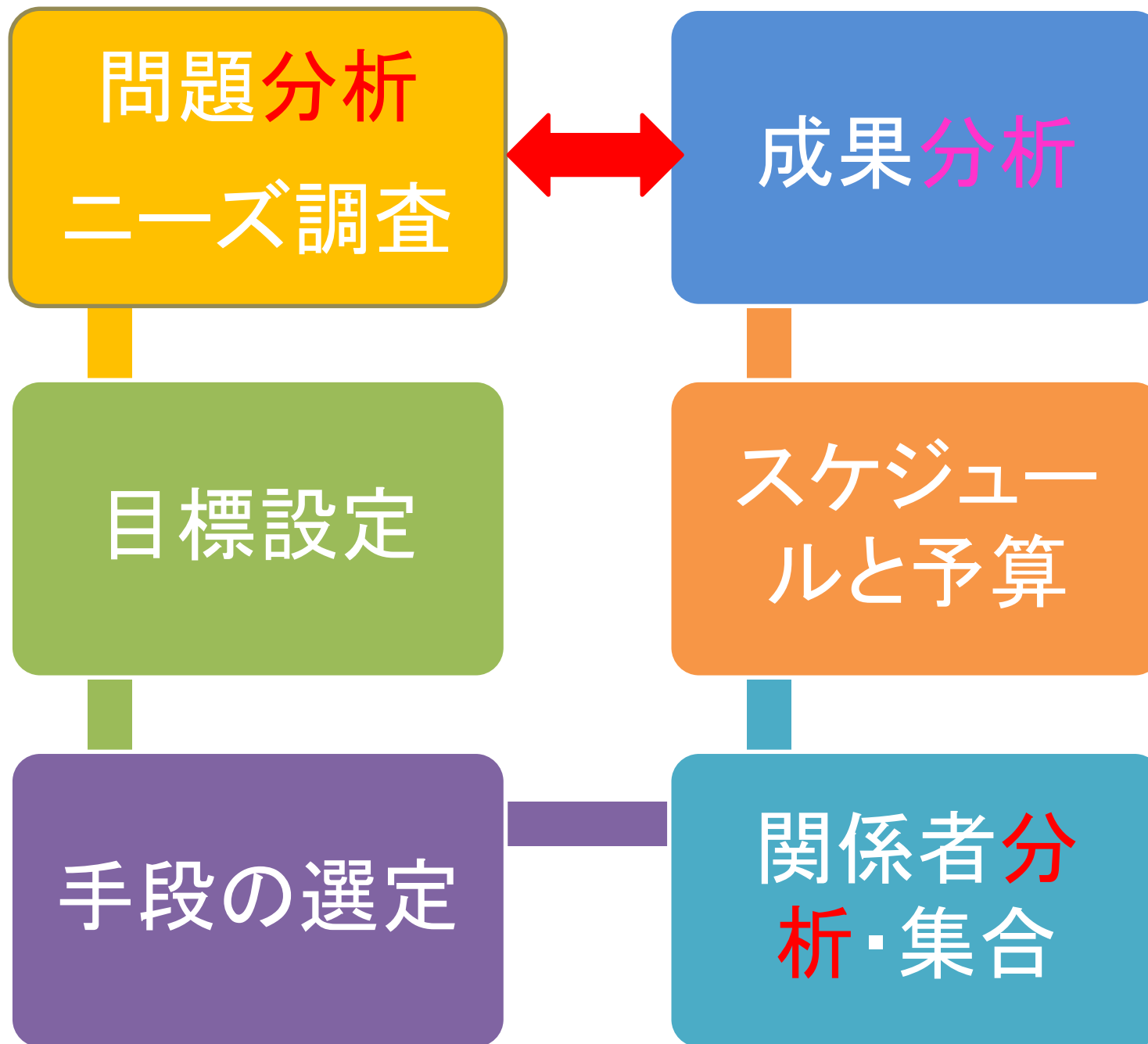


2016年度救急搬送 560件
2017年度救急搬送 320件

大学進学率の向上
観光客数 有効求人倍率
幸せだと思ふ人の比率
観光客が市民病院不払い

事務事業評価はパフォーマンス測定
プログラム評価はアウトカム調査
公共事業評価は事業事前評価

ロジックが通っているか



プログラム評価の5階層～体系的に介入を評価する～

(Rossi, Lipsey and Freeman, 2004: 80)

プログラムのコストと効率性の評価

Assessment of Program Cost and Efficiency

プログラムのアウトカム／インパクト評価

Assessment of Program Outcome/Impact

プログラムのプロセスと実施の評価（プロセス評価）

Assessment of Program Process and Implementation

プログラムのデザインとセオリーの評価（プログラム理論評価
／セオリー評価）

Assessment of Program Design and Theory

プログラムのニーズ評価

Assessment of Need for the Program

Ⅲ 業績測定について

Performance Measurement

...

9. 業績測定 Performance Measurementとは

- 実績測定、業績評価、パフォーマンス・メジャーメントとも

- ＜業績測定のエッセンスというべき定義＞

- サービスやプログラムのアウトカムと効率を定期的に測定すること(Hatry, 1999)

- ＜業績測定という手法の定義＞

- インプット・作業(活動)・アウトプット・アウトカム・効率(生産性)などの一連のデータを定期的に収集し報告すること(U.S.GAO 1992)
 - プログラムの実績、特に事前に設定した目標や基準の達成度合いについての体系的で継続的なモニタリング(U.S.GAO 2012)

- ＜業績測定による評価システム・制度の定義＞

- 業績(performance)の定期的な測定・評価・対策提言・公表」(上野・上野2007)

代表的使用例

SUSTAINABLE DEVELOPMENT GOALS

世界を変えるための17の目標

1 貧困をなくそう



2 飢餓をゼロに



3 すべての人に健康と福祉を



4 質の高い教育をみんなに



5 ジェンダー平等を実現しよう



6 安全な水とトイレを世界中に



7 エネルギーをみんなにそしてクリーンに



8 働きがいも経済成長も



9 産業と技術革新の基盤をつくろう



10 人や国の不平等をなくそう



11 住み続けられるまちづくりを



12 つくる責任 つかう責任



13 気候変動に具体的な対策を



14 海の豊かさを守ろう



15 陸の豊かさも守ろう



16 平和と公正をすべての人に



17 パートナーシップで目標を達成しよう



SUSTAINABLE DEVELOPMENT GOALS

2030年に向けて
世界が合意した
「持続可能な開発目標」です

10. 業績測定 of 指標に求められるもの

- **妥当性**のある指標を設定すること
 - 業績測定において妥当性とはシンプルに「測定すべきものを測定していること」
 - ロジック・モデルに基づいて指標を設定することが推奨されている。
- 目標値は**明確に**設定すること
 - 根拠となるデータに基づいて、意味が明らかで曖昧さのない数値を
 - 目標達成の難しさが異なれば、達成度合いの評価も異なる。
- 実績値と目標値は**的確に比べる**こと
 - 最終的な水準とそこまでの変化のどちらに着目するか、などの的確に選択
- 指標間・プログラム間の**比較、集計・集約は丁寧に行う**こと
 - 比べられないものは比べないこと

11. 業績測定とプログラム評価の補完関係

- ①手法としての業績測定の基本は、プログラム評価のプロセス評価・アウトカム評価の階層にほぼ相当する。
- ②評価システム・制度としての業績測定の目的は、個々のプログラムにとどまらず、プログラムの集合や機関全体の継続的なパフォーマンス向上や効率化である。
- ③両者は、目的の異なる評価であり、役割分担によって互いに補完すべき(下表)。
- ④業績測定の結果を踏まえてプログラム評価の対象を選ぶ、プログラム評価の結果を踏まえて業績測定の指標を設定するなど、両者の連携も。Nielsen&Hunter[2013], Hatry[2013]

項目	業績測定	プログラム評価
タイミング	定期的・継続的	定期的でない
評価対象(数)	多数(しばしば全数)を対象に、一律	少数を対象に、1本ごと
情報の深さ	シンプルな業績データ	因果関係を追究
評価体制	内部評価が基本	専門的リソースが必要

12. 評価

科学風のウソが生産される背景

(1) 構造的背景 最初から「結果ありき」の政策評価が選好される

- 「勇気ある撤退」を全く評価しない**組織風土**
- 予算・人員を巡る担当部署での「限定合理性」や「縦割主義」による**視野狭窄**
- 関係者の「**思込み**」や「欧米では～」型の**権威主義**
- 行政庁側の**中立的でない**評価要請や過度の分析・評価者への介入・圧力
- 分析・評価側の**客観性の欠落**や金銭的・競争的・学術的動機に基づく利害相反

(2) 技術的背景 –

都合のよい政策評価の「結果」が捏造される

- 統計試料の出典・年度の恣意的な取捨選択や混合・補間・加重
- 代表性の欠落や異常値の残留など偏った試料からの外挿・拡大推計
- 分析期間・対象における外的要因をせず、むしろ悪用する
- 分析手法における前提条件の乱用・逸脱や基礎的理論の無視
- 潜在的な偏差・誤差の軽視と強引な解釈

終 より良い評価で世の中を変える15のポイント

1. 評価を理解し、使いこなす **人材育成と適正配置**
2. 評価に使う **基礎データ(evidence)**の確保と蓄積
3. 研究者と実務家の **評価リテラシー**の交換、普及
4. 評価事例の **アーカイブ化**
5. 評価プロセスの **透明性**確保
6. 政策実務家と評価研究者が **情報共有**、協働
7. 政策専門家に対する「シビリアン・コントロール型評価」強化

8. 評価**費用**の確保
9. 研究者と実務家の**人事交流**
- 10.**監査**と評価プロの役割分担
11. 大学と大学院における評価関係**講義の正規科目化**
- 12.**国会や議会**における評価部門の設置
13. マネジメント評価とアカウンタビリティ評価の分業
- 14.**シチズンシップ教育**、18才選挙権教育に評価を組み込み
込み思考力増強、虚偽言説を回避
15. 評価のセクショナリズム回避（課毎に評価しない）

アカウントビリティの意味を再確認

1. Political accountability	政治責任。
2. Constitutional accountability	制度の適切さ。
3. Legal accountability	合法性、合規性、準拠性、デュー・プロセス。
4. Administrative accountability	手続き(process)の妥当性、適切性、行政監察。
5. Management accountability	業績指標の達成。パフォーマンス。
6. Professional accountability	専門職、研究職の能力。
7. Policy / program accountability	政策結果 (outcome) 責任、政策手段選択の責任。

1950s/60s

1970s

1980s

1990s

2000~

Policy sciences
システム工学
社会工学 PPBS
行動科学 管理科学

Systems analysis

program

From analysis to evaluation

Project analysis
cost-benefit analysis
micro-economics

政策工学

Policy Analysis

政策アセスメント

Policy evaluation
system, process, program,
accountability

政策レビュー

Program evaluation
プログラム要素, エビデンス
Program theory

Monitor & Evaluation

政治学・行政学

input → blackbox → Output (policy)



accountability / Plan ~ Do ~ See

GAO

Evaluation research
実践課題としてのプログラム改善
ex. 福祉 医療 教育

Management Review

UK 1970s

Thatcherism
NPM

Reinventing government
1992

Performance measurement
Output 測定
GPRA 1992 → outcome 測定
performance Management
performance budgeting 業績予算

MbO / TQC

PDCA

2003

目標管理型実績測定 政策チェックアップ 行政評価

事業評価

総合評価

実績評価

評価の歴史