平成 13 年度経済産業省委託調査

研究開発プロジェクト等の評価手法に関する調査

報 告 書

平成１４年３月

財団法人 政策科学研究所

はしがき

本報告書は、経済産業省が行った平成１３年度調査研究『研究開発プロジェクト等の評価手法に関する調査』を、（財）政策科学研究所が受託実施した成果をとりまとめたものである。

我が国では欧米に遅れて「評価の時代」が始まっている。しかし、評価対象が広がり、評価が複雑化するにつれ、評価を巡る混乱も生じている。この背景には評価に関する枠組みと基礎的な概念の整理と実態的な展開の解説を行っている基盤的な情報が流通・共有されてこなかったこともある。評価関連の研究者・実務家や関心を持つ当事者等が、個々の局面で健全で信頼できる評価に貢献し成果を蓄積するとともに、そのあり方を巡って建設的なコミュニケーションが可能になるような知的共通基盤を作るために、本格的な評価のテキストを作ることが要請されている。

本調査は、評価のマネジメントを行う人材を中心にした本格的な入門テキストの内容と構成を持つように、近年の研究開発プロジェクト等の評価のあり方を概括することを意図している。もちろん、評価の対象や局面により必要な知識は異なり、一報告書でカバー出来るわけではないが、先ず、中小規模のプロジェクト評価の制度を運営するレベルに焦点を置いたテキストを提案したものである。

本調査の成果が、今後の我が国の評価関係者の理解の深化と評価システムの学習的進化に結びつけば幸いである。

末尾ながら、調査に協力いただいた専門委員会や有識者の方々に深く感謝の意を表したい。今後とも当研究所は我が国の評価システムの健全な進化と成熟を支えるような調査研究活動を展開する所存である。

平成１４年３月

財団法人　政策科学研究所

目　次

＜参考資料＞

本報告書の概要

本報告書は『研究開発プロジェクト等の評価手法に関する調査』の表題の下で行われた作業を取りまとめたものであるが、実質的には、行政レベルにおける研究開発評価の「本格的入門テキスト」の形式で編纂されている。「本格的入門テキスト」の意味するところは、現下の我が国の評価状況に焦点を絞り、その実態の改善に役立つ解説資料として編集しようとしたところにある。

報告書の性格と構成

評価には、その目的や対象、フェーズ等により、難易度に大きな違いがある。我が国では比較的取り組みやすい局面、すなわち、（プログラムに従属した）プロジェクトの事前評価や機関評価などから実践に入ってきた。その他の評価はこれらに比べると格段に困難で、本格的な取り組みのためには専門的な人材による支援も必要になる。このような状況をふまえると、先ず、プロジェクト評価や機関評価などに携わる一般の評価人材のための本格的なテキストブックを整えることから始める必要がある。本報告書は、この点に実務上の焦点をあてつつ研究開発評価の全体像を描くことを課題とした。科学技術ピアレビューを担う研究者への評価法のガイダンスにもなるものである。

研究開発評価の全体像の理解を支援するために、以下の構成でテキストをまとめた。

すなわち、

①研究開発システム全体を特徴づける評価の理念

②評価対象の諸局面を含む評価の推進体制のあり方

③評価の方法論

④実務的な取り組みを深めるための具体的な評価作業のあり方

的確な評価能力を身につけるには、さらに、研究開発マネジメントの知識や方法論が必要である。また、評価の方法論を、全体像を理解したうえで使いこなすためには、事例を学ぶ必要がある。こうした次の段階への知識やアプローチの概略については、参考資料にまとめ関係者の便に供した。

行政における技術経営と評価（第１章）

今後の施策は「行政経営」という新しい枠組みに適応した立場から展開することが不可欠であり、「評価制度」が行政全般の多様な局面で導入されつつある。先んじて評価が導入された研究開発分野では、固有の困難性のために、これまでの試行においては、必ずしも的確な評価がなされてきたとはいえない。この背景には、評価に関する枠組みや基礎的な概念、その実態的な展開のために必要な基盤的な知が、関係者に流通・共有されてこなかったことがある。評価の質的改善への建設的コミュニケーションと協働を可能とするようなテキストを作ることが要請されている。

研究開発評価の理念（第２章）

研究開発評価システムでは、評価システム全体を統合する理念が重要である。評価の理念に多様性があることは、異なった制度・土壌のもとにある各国のシステムを比較すると見えてくる。

フランス、イギリス、ドイツ、アメリカ、そして韓国の評価システムを大局的に比較し特徴づけた。

評価システムの理念は当該国の状況と組織風土に適応するように定め、評価制度の設計と運用制度全体を貫く必要がある。我が国では、そのマネジメント上の蓄積を踏まえると、支援型の評価制度、すなわち評価者と被評価者が一体となり評価に当たるinclusive（内在的）で支援的な評価制度であるべきと考えられる。一方で、この方式が内包する欠陥を補うために、明示性の確保や第三者によるチェックとフィードバックメカニズムの導入が必要である。

評価の仕組みと推進体制（第３章）

（１）評価推進体制の全体像

研究開発評価の推進体制の全体像を理解するためには、評価担当組織が置かれた外部体制と評価組織の内部体制の双方のレベルでその概要を把握する必要がある。とくに我が国では前者の理解を支える教育機会が殆ど無いことに留意すべきである。

評価制度を設計・運用するために必要な知は多岐にわたるが、外的な認識枠組みとして、①評価一般に関わる考え方や仕組み、②評価対象である科学技術のマネジメント、③研究開発を担う組織現象の特徴、④評価法の運用に必要な基盤的手法、に関する知がある。また、評価の内部構造として、①評価システムを貫く評価理念、②評価対象の捉え方、③評価を実際に行うためのデータの測定・収集・分析と評価結果に結

びつける評価の方法論、④評価人材のあり方や分担、などに関する知がある。本報告書では、評価の内部構造に関わる問題を本文で扱い、その他は「参考資料」において解説を加えた。

（2）評価の諸局面

評価の局面として、プロジェクト／プログラム・制度／施策・政策、及び研究者／研究グループ（業務担当組織）／機関という評価対象の階層性や、事前評価、中間ないし途上評価、事後評価（さらに直後と追跡に分ける）という評価時点のフェーズ区分があるが、これらの組み合せにより評価として取り組む内容が大きく異なることを示した。

（3）評価の内容

評価内容の基本要素を４種の観点から整理し、それぞれの特徴や論点を説明した。

①目的系：対象が目指すものの妥当性について評価する内容である。

②コスト系：費用の量やその構成比等の質に関わる内容である。

③成果系：成果についての内容であり、科学技術の内在的価値（科学技術としての価値）と外在的価値（さらに経済的価値と社会的価値に区分できる）を含む。目的として定めた主題に関する意図的成果と非意図的な副次的成果、直接的成果と間接的成果の区分が必要であり、さらにまた、アウトプット、アウトカム、インパクト、アディショナリティ、スピルオーバーなどの概念による区分もある。

④マネジメント系：実績やパフォーマンスと採用されたマネジメントとの関連において適否を論ずる内容である。

（4）評価人材

評価を担う人材には、主に、①評価のマネジメントを担当する人材、②社会経済性の評価パネリストと分析エキスパート、③科学技術ピアレビューアの３者が必要である。それぞれの評価人材に必要な資質や、その適切な養成・確保の方法についてまとめた。

（5）評価推進体制のまとめと課題

評価対象別に、評価の体制や目的、内容などについての特徴と、我が国で今後評価の推進上とくに取り組むべき課題をまとめた。評価推進体制のまとめと課題に用いた区分は、①プログラムのもとで展開されるプロジェクト（従属プロジェクト）②プログラムに基づかないプロジェクト（独立プロジェクト）③プログラム、④施策と狭義の政策（制度の複合体）、⑤全体に関わる総合政策、である。

評価の方法（第４章）

（１）評価論

評価一般の論理が、評価の基礎力として必要である。

①評価の枠組み

評価結果は意思決定そのものではなく、評価と意思決定とを区別することが必要である。外在的評価と内在的評価、明示的評価と暗黙的評価などの基礎的枠組みを解説した。

②評価の原理

評価とは対象の価値を明確にすることである。そのため、評価対象の状況を評価値の集合として明示的に表すことが必要である。これに価値観に相当する評価要素毎の重要度を乗じたものが評価結果としての価値である。これらを行列表現して定式化し、システム論の表現形式で解説した。

③評価手順の基本設計

評価の流れに沿って、評価ドキュメント（文書資料）の基本構成、困難であるが具体的に定めなければならない評価項目／評価基準／評定区分とその構造、についての内容と考え方を示した。

④評価結果の利用

評価のための評価や評価の形骸化に陥らないよう、評価結果の有効な利用を図り、意思決定結果が開示される必要がある。

（２）評価手法の体系

原理的かつ実践的な見知から、評価に関する調査・分析・評価手法の体系化を行った。調査法、分析法、評価法は互いに連関しているが、方法論としては峻別する必要がある。調査法は、対象の質に関わる情報収集と数量に関わるデータ収集に分けることができる。評価手法は、定性的評価と定量的評価に区分できるが、半定量的評価や定性・定量評価法を組み合わせる総合的評価を加えて分類することが適切であり、各々の特徴や代表的手法を解説した。一方、単純評価と、エキスパートの支援が必要な複合評価という区分も用いた。

これらの区分に基づき、特徴や代表的手法を解説しつつ、評価手法の体系をまとめた。

（３）評価結果の表現

評価値の相対的な位置や意味を表す評価結果の表現方法は重要で、この表現方法自体が高次の評価法である。

研究開発評価の方法論（第５章）

上記を踏まえて、現在の我が国の研究開発評価の向上を図るうえで重要な方法論に焦点を絞り、具体的に説明した。

はじめに、評価対象（従属プロジェクト、プログラム・政策、独立プロジェクト、機関）と評価フェーズ（事前、途上・直後、追跡）、及び評価内容（科学技術性、社会経済性）と評価フェーズ（事前、その他）のそれぞれ２軸をとり、これら２平面上の各評価局面での方法論の有効性についての概観を行った。

扱ったものは大要次のものである。

基礎的調査方法としては、インタビューとヒアリング、サーベイ（アンケート）、事例調査・事例研究を、測定法としては、評点づけと順位づけ、実証データと統計的データを、分析法としては、ビブリオメトリックス（論文分析、特許指標）、テクノメトリックス、エコノメトリックス（寄与率、波及効果、外部性、必然性）などの重要概念を扱い、代表的なミクロアプローチである BETA 法を紹介した。

次いで定性的評価法であるピアレビュー法とパネル法を、その留意点を含め説明した。

また、半定量的方法である評点法では、評定段階の設定、評点づけと総合化の段階別について、比率評価法では、ランキング、ベンチマーク、ポートフォリオ、コントロール・グループ・アプローチ、前後比較、非事実仮説等の方法を、また指標法では、指標間の比の算出と指標の総合化に分けて説明した。

総合的評価法であるシステム評価法としては主に Decision Management 法を、ロジック評価法としてはケース評価法とレトロスコピック評価法を、レビュー法としてはフォーサイト、フォーカス・グループ、デルファイ法を紹介した。

評価推進体制の策定（第６章）

最後に、評価推進体制を具体的に設計する際の論点をまとめた。ここで必要なことは、評価の枠組み、すなわち評価システムと評価体制に関する枠組み、つまり広い意味での評価方法とそれを実施運営する際の人の問題を明確にすることである。

評価推進体制の策定上での基本的なポイントを、評価対象によらず共通のものと、本報告書で焦点を置いてきた従属プロジェクトの事前評価と中間・直後評価、機関の途上評価について整理した。加えて、プログラムの追跡評価、政策の評価、独立プロジェクトの事前評価については、評価の困難さのポイントとそれに対する基本的なアプローチを説明した。それぞれ海外の代表的な評価事例をあげて参考に供した。

調査目的・内容・方法・体制

1．調査目的

効率的・効果的に研究開発プロジェクト等を実施するためには、厳正かつ適正な評価が必要であり、そのためには、適宜評価のあり方を見直し、その都度適切な評価手法等の開発を推進していくことが必要である。このため、最近のプログラム等に対する内外の評価の動向をふまえ、研究開発評価の概念と枠組みを整理する。あわせて、研究開発プロジェクトやプログラム等に関する評価システムの実態を把握し、現状における評価システムのあり方を検討する。

2．調査内容

（1）研究開発評価の枠組みのまとめ

研究開発プロジェクト等の評価システムのあり方を評価関係者が相互に検討し、適正な評価手法等の運用を可能とする前提として、研究開発評価の枠組みを概念的に整理する。

①評価についての概念の整理

研究開発を対象とする評価を検討する際の前提として、評価一般についての概念等を整理する.

②研究開発システムについての概念の整理

評価対象である研究開発システム・制度について、概念と特徴、要件を整理する。なお、研究開発システム・制度の目標設定系にかかわる概念は、特に評価システムのあり方に関連するので、展開・整理する。

③評価システムについての概念と枠組みの整理

最近の評価システムに関する動向を踏まえ、研究開発システムを対象にした評価システムの設計／運営系にかかわる枠組みとあり方を示し、内容と留意点を整理する。

（2）研究開発プロジェクト等の評価実態の整理

研究開発のプロジェクトやプログラム等の評価の実態調査に基づき、評価手法の概要をまとめると同時に、設計・運用されている評価システムの概念体系に沿ってまとめる。

①評価システムの概念体系に沿ったまとめ

国際的に評価が設計・実施・改革されている多様な状況を、理念・目的や環境・コンテクストの差異に留意しつつ、評価システムの概念的体系に従って、整理する。この中で評価システムの実践的な機能の共通理解を促すとともに、今日的な動向、課題と展望も明らかにする。

②評価実態の整理

研究開発プロジェクトやプログラム等の事前評価、中間・事後評価及び追跡評価等において利用されている評価手法等の実態調査を実施し、その概要と特徴を取りまとめる。

評価は文化とも関係が深いので、比較に留意しつつアメリカ、イギリス、フランス、ドイツなどの海外の注目すべき考え方や事例を紹介する。

3．調査研究体制

調査研究の専門性などに留意し、当研究所内に下記の構成の専門委員会を設置して指導と助言を得た。

＜専門委員会＞

委員長	平澤　冷	政策研究大学院大学 教授
委員	丹羽　冨士雄	政策研究大学院大学 教授
	村上　路一	（株）宇宙情報技術研究所 副社長
	桑原　裕	（株）ホエールズウェブ・ドット・コム 取締役
	西村　吉雄	東京大学大学院工学系研究科 教授 （平成１４年２月１５日まで日経ＢＰ社編集委員）
外部協力	林　隆之	大学評価・学位授与機構 助手

本調査研究を実施した当研究所内の体制は以下のとおりである。

＜実施体制＞

総括責任者	大熊　和彦	（財）政策科学研究所主席研究員・研究部長
	伊東　慶四郎	（財）政策科学研究所主席研究員
	猪瀬　秀博	（財）政策科学研究所主席研究員
	元川　浩司	（財）政策科学研究所主任研究員
	川島　啓	（財）政策科学研究所研究員
	斉藤　文子	（財）政策科学研究所研究員
	大谷　卓史	（財）政策科学研究所客員研究員

序

現在我が国においては、「評価」が行政の多様な局面で導入されつつある。「研究開発評価」はこの動きに数年先んじて本格的に取り組み始められたが、課題の困難性のために依然として戸惑いや混乱の渦中にある。

本報告書は『研究開発プロジェクト等の評価手法に関する調査』の表題の下で行われた作業を取りまとめたものであるが、実質的には行政レベルにおける研究開発評価の「本格的入門テキスト」の形式で編纂されている。「本格的入門テキスト」の意味するところは、現下の我が国の評価状況に焦点を絞りその実態の改善に役立つ真に適合した解説資料たらんとするところにある。

具体的には、①研究開発の評価作業やその運営に携わる行政官や資源配分機関の職員のための、「研究開発評価システムの設計と運営」に資する一般的知識、②評価作業の中でも最も多くの関係者が関わる、「プロジェクトの事前評価のあり方」と、③「機関評価のあり方」との質的改善に資する詳細な知識についてである。

しかしながら、研究開発評価の中でも、その取り組みのためにより高度な知識と深い経験を要する評価課題については、本報告書の埒外とした。これらの高度な評価課題は、特別に訓練された「評価エキスパート」の支援を必要としていて、本報告書の範囲を超えているからである。ちなみにそれらの評価課題とは、④プログラムに基づかないで「独立に展開される（多くの場合大型で長期にわたる）プロジェクトの事前評価」、⑤政策や施策の見直し作業に欠かせない「プログラムの途上ないし追跡評価」、⑥そしてさらには、政策形成プロセスの設計や上位の戦略のあり方と切り離せない「政策の事前評価」等である。これらの評価課題は②や③の課題よりむしろ重要であるが、一般の評価関係者が安易に立ち入るべきではないと考える。この立場から、ここでは①の中で一般的知識として、これらの高度な課題についての評価のあり方に触れる程度にとどめた。なお、これら困難な評価課題の適切な運営のためには、桁の違う資源配分と継続的な努力が必要であることを特記しておきたい。

研究開発の評価活動が、我が国において形骸化することなく、真に有効な手段として定着することを念願し、本報告書がその一助となることを期待している。

第１章　行政における技術経営と評価

１．１　行政管理から行政経営へ

現在我が国が置かれている状況は、かつてキャッチアップ体制の下で高度成長を遂げた状況とは全く異なる枠組みの中にある。経済成長率でみると、73 年の第一次オイルショックの直後から低成長の時代に移り、また 80 年代の中頃には GDP/人口が世界のベストテン入りし、名実ともに経済的なフロントグループの一員にもなった。それに加えて先進国全体の成熟社会への移行、社会全体の傾向としての知識社会化、そしてまたベルリンの壁の崩壊後 90 年代に入ってからは本格的なグローバル経済体制へとさまざまな社会経済的枠組みが変わってきた。

このような枠組みの変化は、民間企業はもちろんのこと行政においても、かつての「行政管理」の視点からではなく、「行政経営」という新しい枠組みに適応した立場からの施策展開が必要になったことを意味する。例えば、低成長の時代に入ったならば、限られた余剰資源を有効に活用するために資源配分において「選択と集中」が必要となり、その選択と集中のための論理を生み出す能力が重要となる。キャッチアップ体制の下では、先進国に見習ってターゲットを定めることができたが、フロントグループの一員としては、固有の問題を解決するために独自の戦略をつくる必要があり、その戦略形成機能が新たに必要となる。また、成熟社会では、供給側ではなく受容側を重視する政策が重要となり、いわゆるニーズ型の、社会ディマンドを重視した政策の枠組みを強化する必要がある。キャッチアップ体制の下では供給側を強化する政策で十分であったが、成熟社会ではもはや受容者側を無視しては政策の効果が上がらない。従って受容者側の社会ニーズやディマンドを分析し、あるいは社会経済のインパクトを分析する、このような新たな能力が必要になる。キャッチアップ体制の下では、国内政策を重視するだけで十分であったが、グローバル経済の枠組みになって、国際政策が特に重要性を帯びてくる。国際化に対応した政策には、グローバル・ウォッチングがまずその大前提として必要であり、また、国際的なベンチマーク等の比較調査も重要になる。このような、世界のデータを収集しそれを独自に分析する能力がやはり新たに必要となる。知識社会では、専門分化した知識がさらに専門分化していくことの繰り返しの中で、本質的に発散的な思考傾向を生み出し、知識が個別に分化し発

散的に展開される。従って、こうした自然の傾向に対して知識の統合を促す新たな政策が必要になる。統合化政策にはシステム思考やシステム論等を踏まえた新しい思考のメカニズムが要求される。また一方で、知識社会そのものはソフト社会でもあり、思考世界を扱える一連の能力が要求される。

これらの能力をまとめると、社会や世界の状況を的確に把握するための分析的能力と、新しい状況に適合した社会システムや制度を構想し実体化する構成的能力の２種類に要約される。

１．２　新行政経営における評価の位置付け

我が国では、1997 年から新たな行政改革に取り組み始め、2000 年までの間にその準備段階を終えて、2001 年から新組織形態に移行した。このような組織改革に続いて、新たな組織形態に適合した新しい行政経営の方式が模索されている。その第一段階として、「評価制度」が行政全般の中に導入され、例えば 2001 年度から発足した独立行政法人に対し、機関評価に当たる評価を担当する評価委員会が設けられた。また、政策評価を各省の内部において行うことが義務づけられ、それを定めた法律が 2002 年の４月から施行された。

研究開発に関しては、1995 年の科学技術基本法が成立したことに端を発し、翌年策定された第一期の基本計画に合わせて、研究開発評価を十分に行うべきことが、基本計画の中で定められ、そして研究開発評価に関わる「大綱的指針」が 1997 年に制定された。研究開発に関しては、他の行政課題に一歩先んじて、マネジメントを整備する段階に入ったと言える。

この第一次の大綱的指針は、さらに 2001 年度から施行された第二期の科学技術基本計画に合わせて改訂され、2002 年度までの間に、第二次の大綱的指針の内容を踏まえた、各省レベルの第二次の「評価指針」が策定されたところである。第一次の大綱的指針は「機関評価」と「プロジェクト評価」を中心課題として取り上げたのに対して、第二次の大綱的指針ではそれらの質的向上に加え、「制度や施策」の評価、これは実質的にはプログラムの評価に相当するが、それと「研究者」の評価という新たな項目とが加えられた。

機関評価は 1997 年から毎年国立研究機関を中心にして実施されてきたが、多くの国立研究機関が独立行政法人に移行した後、初めての独立行政機関の評価が 2002 年の７月を目途にして行われようとしている（表２－２参照）。

このような流れの中で、評価に関しかなりの混乱がみられる。まず、研究者の側を中心として、評価に対する深く沈潜したアレルギーがある。

研究開発評価に対する研究開発者の懐疑的な想い

・評価の意義が分からない
—何の役に立つのか
—時間の無駄ではないか
—研究に没頭する時間をさかれる
—余分な負担がかかる

・評価は自らが既に行っている
—さらに評価を重ねる必要があるのか
—研究の意図や動機は研究者の個人的なものであり、他と比較して評価することはできないのでは
—自分の研究分野では自分が第一線で研究しているので、そのような深い知識を持っていない他者は妥当に研究内容を評価できないのでは

・最近の評価の必要論への懐疑
—事前評価や中間評価はともかくとして、研究が既に終了してしまった段階で行う事後評価は形式的な意味以外にどんな意義があるのか。評価結果の活用の仕方がないではないか
—社会経済的に役に立つ研究へ傾斜していく中で、評価を強化することによって自由に展開できる基礎研究ができなくなってしまうのでは
—研究結果さえまともに出しておけば評価を行う必要はないのでは

また、本格的に始まったプロジェクト評価を担うべきパネルメンバーも評価に不慣れであり、さらにはプロジェクト評価の評価システムを設計する立場にある行政官も、未知の領域へ向けてスタートするという状況にあって、評価に対する経験や知識の乏しいまま評価システムを設計し、そしてまたそれを運用せざるを得ない状況に置かれている。もう一方で、特殊法人改革の一環として「経済性評価」を各特殊法人一律に義務づけたように、評価制度を推進する側にも「研究開発評価」の特殊性に対する理解が十分でないために惹起された混乱も見られた。

このような状況の結果として、定型的な評価手法が評価対象の適性を考慮しないまま導入されたり、あるいは外国で用いられている評価制度を、評価システム全体の中の整合性を考慮しないで導入する等のちぐはぐな行動が見られ、その混乱に拍車をか

けてきた。また一方で、評価それ自身が対象によらずいずれも同様に簡単に処理できるものであるかのような単純な認識や、評価制度を導入することによって直ちに運営の効率化がもたらされるとする楽天的な期待も散見される。

評価対象によって、またその評価の目的やフェーズによって、評価の難易度には非常に大きな開きがある。日本において研究開発評価の課題としてまず導入された「プロジェクトの事前評価」は、もしプログラムの下でそれを展開するのであるならば、すなわち「従属的なプロジェクト」であるならば、評価はそれほど困難ではないし、また、機関評価も実績評価を中心とした途上評価であるわけで、機関のミッションないし目的、あるいは機関の中長期計画が明確に定められているならば、その評価もそれほど難しいものではないと言える。

このような、比較的取り組みやすい評価局面から、我が国では評価の実践に入って来たが、この後に控えている評価課題は、これらに比べて格段に難しい課題であり、それは、「政策評価」や「プログラム評価」であり、そしてまた「独立して展開される大型のプロジェクトの事前評価」等である。政策評価を行うためには、その前段階として、その政策が内包するプログラムや制度の分析が行われなくてはならないし、そのプログラムや制度が内包するプロジェクトや実施状況等がさらに分析されなくてはならない。このような階層的な広がりを持った評価対象をとりまとめて評価し、その結果を政策の見直しに反映させることが必要になる。このような本格的な分析は、いわゆる分析のエキスパートによって初めて可能となるが、我が国にはそのような人材はほとんど存在しない。評価マネジメントを適切に遂行できる人材とともに、早急にそのような人材の養成に取り組むことが必要である。

このような状況の中で、「プロジェクト評価」や「機関評価」を担うマネジメント人材のための実務的なテキストブックがまず必要であることが認識された。そしてまた同時に、今後取り組まなくてはならないより困難な評価課題、例えば「プログラム」の社会経済的側面のインパクト分析等の専門的な課題については、少なくともその概要について整理しておく必要がある。しかしながら、この種の課題に本格的に取り組むためにはそれなりの専門家の養成が必要であり、そのエキスパートを養成するためのテキストブックは、プログラムの下で展開される「プロジェクト評価」や通常の「機関評価」のための内容に比しあまりにも内容のレベルが異なるので、本報告書ではこの難易度の高い部分のテキストブックについては今後の課題とし、本格的な記述を展開していない。

一方、科学技術のピアレビューを担う科学技術専門分野の研究者に対しては、評価

システムの運営を担当するマネジメント人材が、このテキストブックの中で説明される研究評価システムの全体像を踏まえて、評価ピアパネルを開催する際に、導入的なガイダンスを行うことが適切であろう。

ピアレビューア自身を評価者に仕立て直すためのテキストブックという目的は、従って本プロジェクトでは意図していない。研究者が評価パネルメンバーとなる際に期待されることは、その研究者が専門とする研究領域についての知見を動員し、ピアレビューアとして、提案されている内容の科学技術的側面を評価することである。ピアレビューアの役割を、その限りの評価機能に限定すべきである。しかし、そのようなピアレビューアの見解はまさに彼らにしかできない貴重なものであり、評価のマネジメント担当者は、彼らの見解に他の必要な判断要素も含めて評価全体の意思決定を行うべきである。

現在我が国において先ず必要とされる研究開発評価の質的改善のあり方は以上のようなものである。

第２章　研究開発評価の理念

この報告書は、研究開発評価の全体像を理解することを目的として構成されている。すなわち、①研究開発システム全体を特徴づける評価理念、②評価対象の諸局面を含む評価の推進体制のあり方、③さらには評価の方法論の概説と、④重要な研究開発評価法についての具体的な説明、⑤そして実務的な取り組みを深めるために「プロジェクト評価」と「機関評価」を中心にした具体的な評価作業のあり方についてである。

プロジェクト評価と機関評価は現在我が国で中心的に展開されている研究開発評価のカテゴリーである。さらに進んではプログラムの評価、及び制度や政策の評価等の課題が残されているが、それらについては本報告書では評価システムの全体像を説明する中でまず触れることとし、また具体的な取り組みの方向性については「評価推進体制の策定」の章でその概要を示すにとどめた。その理由はプログラム評価や政策評価は、それらを十分に行うためには先ず体制を整備する必要があり、具体的な方法論を議論するのは体制の整備に合わせ次の段階の課題であると考えるからである。

研究開発評価に取り組む前提として、さらに必要なことは、研究開発マネジメントの知識、あるいは方法論を具体的に使いこなすための、方法論の基盤に相当する知識やスキルである。これらについては本報告書の巻末にその概略を解説する（参考資料１及び２）。また、研究開発の全体像を理解した上で、それらを使いこなすためには、多くの事例を学ぶ必要があり、具体的な研究開発評価の事例に関して、やはり参考資料の中にそれらをまとめた（参考資料３）。これらの知識を援用しながら、あるいはこれらの解説を参照しながら、以後の記述について理解を深めて頂くことを希望する。

研究開発評価システムは言うまでもなく、システム全体として無矛盾の仕組みである必要がある。そのような無矛盾の仕組みにするためには、いわばシステム全体の統合的な理念が必要になる。そのような理念の下に矛盾の無い形で評価システム全体がつくられていることが望ましい。

2．1　主要国の事例

主要国の特徴的な理念に相当する仕組みを見ると、実にバラエティに富んだ様子がうかがえる。表２－１にその様子をまとめた。

表2－1　主要国における評価の特徴

国　名	特徴的な仕組みと評価理念
フランス	・「ギャランター（各セクターからの代表者）」制度（評価における代表民主制） ・社会の中でのディベートに一度委ねた後に意思決定
UK	・「契約」に基づく研究活動と、資金配分に見合った価値基準（value for money）による評価 ・評価結果の資金配分への直接的反映
ドイツ	・下部機関への「信託」と下部機関による自律的運営、3%ルールによる規律の保持 ・支援的評価システム
アメリカ	・多重に設けられたチェックアンドバランスの評価の枠組み ・ピアレビューとメトリックスの発達 ・循環型評価システムの導入
韓国	・一元的統合評価制度 ・Authorization と Appropriation に相当する段階評価

（1）フランス

フランスでは、「ギャランター制度」という他の国には無いやり方が特徴的に実施されている。**ギャランター**とは、社会を構成しているそれぞれのセクターからの代表者（保証人）を意味する。ギャランターシステムによる評価では、各セクターからの代表者によって構成された代表者会議（委員会）で評価の意思決定をすることになる。これは、いわば代表民主制を評価システムに持ち込んだものと理解できる。研究評価のような専門性の高い分野の場合、単なる代表民主制であるわけにもいかず、結果と

してギャランターが自分の信頼できる専門家の意見を携えて参加する。つまりエキスパート付きのギャランター制度ということになる。フランスの場合評価制度はこのようになっている。

CNRS のギャランターによる評価制度

国立科学研究センターCNRS（Centre National de la Recherche Scientifique）は人文社会科学を含む全科学分野を包括するフランス最大の研究連合体であり、教育を本務とする大学の研究機能を高め、補完する目的で設立された。CNRS は 1,200 余りの直轄研究ユニット、混合研究ユニット、連携研究ユニット等の連合体であり、企業との関連も有している。所属研究者数はフルタイムの専従研究者だけで 12,000 人を超える。CNRS は、直接、間接にフランス全体の科学技術研究活動、科学技術政策に大きな影響を及ぼす地位にある。

さて、ギャランター・モデルとは一言でいえば専門家付代表民主制による評価方式である。つまり社会を構成する各界の代表者（ギャランター）から或るパネルにより、代表者が推薦する専門家の支援の下で、評価する方式である。ギャランターは、**それぞれの選出母体で選任され**、担当事務局で編成するものではない。社会全体ないし研究者全体にかかわる大きな枠組みのプライオリティを決める最上位のパネルの場合、内容的な妥当性だけでは判断できないので手続き上の正当性が必要となる。政策評価の最も大きな枠組みの決定には、各主要国とも、このような正当性の視点が必ず加えられている。米国の場合、国家目標や、それから直接導かれる重点分野や優先枠に関しては、大統領選挙によって信任される。逆に言えば大統領候補者への各界からの要望が整理され、候補者毎の政策大綱が提示された上で大統領が選出される。従って、結果として科学技術関連予算の内部構造が 4 年（ないし 8 年）毎に劇的に変化することになる。英国でも、優先分野の順位は社会に開かれた幅広いコンサルテーションプロセスからなるフォーサイト・プログラムによって決められ、各プログラムの優先枠の大きさに反映される。

このような民主性と専門性の調和が、科学技術政策を決定する場合非常に大きな課題となる。そのディレンマを解く鍵は、民主性を優先した正当な手続きにより決定される枠組みの中で、内容的な妥当性を担保するために、専門性を優先した評価システムにより運営される下部機関に権限を委譲していくシステムである。

より具体的に述べると、CNRS は、研究機関及び、研究者個人に対する系統だった評価システムを有している。CNRS の評価は、科学研究委員会 CN（Comite National de la Recherche Scientifique）によって実施される。CN のミッションは以下の通りである。

- 12,000 人余りの研究者、研究活動の評価を行う
- 研究者の採用、昇級の審査を行う
- 1,200 余りの研究機関の創立、閉鎖、更新の評価を行う
- 科学の現状分析と展望を行う

CNには、セクション委員会、科学審議会、部門評議会、学際委員会などがあるが、研究者及び機関評価に携わるのは、セクション委員会と科学審議会である。研究者の評価はセクションの委員が担当する。機関評価の内、新設と閉鎖の審議はセクション委員会と科学審議会が担当するが機関の更新はセクション委員会が受け持つ。

1) セクション委員会

CNRSの研究者が携わる全科学研究分野は40のセクションに分割される。この分割は研究分野の動向、発展に合わせて定期的に更新される。12,000余りの研究者はいずれかのセクションに属し、また、1,200余りの研究機関もいずれかのセクション（あるいは複数のセクション）に属す。各研究者、研究機関所属のセクションの委員が評価を担当する。

各セクションの評価委員はそれぞれ21人（総計840人）で、内14人は分野別研究者共同体毎に選出（総計560人）され、7人（総計280人）は研究担当大臣によって任命される。分野別研究者共同体はCNRS関係研究者55,000人を母体としている。このようにセクション評価員の2/3は、フランスの全ての公的研究機関に属する者、及びCNRSと共同研究関係にある私企業に属する者によって選ばれる（ギャランター・モデル）。21人の委員の大部分はCNRSの研究者であるが、時に、企業の専門家及び著名な外国人科学者が選ばれることもある。

セクションは春と秋の年2回、それぞれに3～4日の会合を開く。秋の会合は機関と研究者の評価に当てられ、春の会合は研究者の昇進、新規採用等が審議される。この会合では各セクションの分野動向、国内の研究状況、国際レベルにおけるCNRSの位置等も合わせて議論される。

委員の任期は4年、2期まで務めることができる。

評価委員は、このように、CNRS内部のみにとどまらず、広くフランスの公的研究に携わる人々によって選ばれる。

2) 科学審議会CS

科学審議会CS（Conseil Scientifique）はCNRS全体の方向、組織などについて検討、提言するほか、研究機関の創立、閉鎖の評価に携わる。

CSは37人の委員により構成され、うち11人はCNRS関係研究者55,000人による選出、他の26人は研究担当大臣の任命による。任命される26人の委員のうち、8人は外国人（うち5人は欧州連合関連国より選出）、また、3人は経済界からの専門家を選ぶことが決められている。この8人の外国人委員の指名には、選出された11人の委員も参加することができるが、決定は研究担当大臣による。任期は4年。

CNRS外の人材を評価作業に取り入れる試みは、2001年に入って決定された。改正後、第一回選挙が10月に行われ、外部の種々の専門家が参加する評価活動は2001年末以降に実施されている。

また、フランスに関しもう一つの特色ある方式がある。フランスでは大学や研究機

関の機関評価を行うための独立評価機関が設定されている。この独立した第三者機関は、例えば大学の場合だと6年に1回程度の割合で順次大学や研究機関を巡り機関評価を行うが、その際評価専門家が大学に派遣されて、3ヶ月から4〜5ヶ月に亘って調査をし、その調査結果を報告書としてまとめる。ここまでは他の国にも類似の制度があるが、フランスの特色はその先にあり、報告内容は社会に向かって発せられる。もっとも大学評価に関するこのやり方はオランダでも同様であるが、評価結果を直接リソースの配分に反映させるのではなく、社会の中で一度議論を深め、そのメカニズムの中でリソースの配分にまで影響が及ぶというやり方を採っている。パブリックな議論を重視しその過程に結論を委ねると言う意味で、やはり民主主義に対する深い信頼感を見て取れる。

これに対してイギリスの場合には、研究機関の評価機構や大学の評価機構の評価結果がそれぞれ直接研究費や研究運営費の額の増減に反映される仕組みになっている。

（2）イギリス

イギリスの特色は「UKモデル」と呼ばれる方式にある。UK **モデル**とは、研究開発の管理機構の内部に評価の専門家を職員として擁する方式のことである。彼等が評価作業のマネジメントを担い、ときには評価のパネリストとして評価そのものに携わることもある。逆の言い方をすれば、そのような評価を担える人材が既に多く養成されていて、彼らが行政機関の内部にまで浸透していると言える。

これに対しフランスやイタリアでは評価人材を育成する高等教育の体系的な制度が発達していなかったため、実際に評価業務に就いて OJT により知識を身につけていくことが普通であり、評価人材の集積はまだ十分ではない。

また、別の特色としてイギリスの場合、評価の前提としてコントラクトつまり「契約概念」の重視をあげることが出来る。例えば、資金提供機関と業務実施機関との間に資金提供に伴い業務が委託されたとする。その契約内容は委託された業務を買い取る形式になっていて、value for money 資金に見合った価値が評価のキーワードとなる。このようなコントラクトを重視するシステムの下では、評価結果が直接リソースの配分にまで反映されても不思議ではない。公的資金が不足してきた段階で、その少ない予算をより多くの価値を生み出すところに配分するためにこのような評価制度が整備されて来た。

ところで最近の英国では、評価結果を資金配分に直接反映させる方式が機能しなくなってきた。それはある種の特殊事情にもよるが、資金総額が増加しない状況下で、

評価を絶対評価制度として運用したため、各大学が其々のパフォーマンスを評価基準に適合させるよう努力したことにより、高得点を獲得する大学が増加し、資金配分額が高位に平準化してきたことによる。

（3）ドイツ

これに対してドイツの場合には、コントラクトに対応する統合的な評価理念はトラスト（**信託**）であり、上部機関が下部機関を信頼して予算配分を行い、下部機関は自己評価を中心にしてそれを運営して期待に応えるという「信託概念」で関係づけられている。従って、基礎研究の場合では、研究者の self-organization が重んじられていて、そういう自律的な研究運営が最大限認められるような形になっている。評価も従って、支援的な評価、サポーティブな評価システムが良いと理解されている。

機関評価の場合でも、上部機関が委嘱した評価専門家パネルが下部機関であるところの研究機関に対しヒアリング調査等を行って評価を進めるが、その結果は上部機関つまり委嘱元に報告され、上部機関はその評価結果である専門家の意見を重視し、その下部の研究機関ができるだけパフォーマンスの上がるような支援的な手を打つという関係になっている。

このようなトラストを中心にして運営されるシステムでは、研究者の自発性が要求される基礎的な研究領域で、特に有効であるが、社会経済的な効果を上げるとか、あるいは大規模な予算を投入して、目標通りの結果を上げることが期待されているような領域においては必ずしもうまくは機能しない。特に最近、ドイツの大型研究は殆ど失敗したと言われているが、それは評価制度自身が大型研究にあまりなじまない理念を掲げていたことも一因ではないかと思われる。

（4）アメリカ

このようなヨーロッパの特色ある国々に対しアメリカでは、まず、ヨーロッパの評価論の研究者たちに対応する研究開発評価研究者はそれほど多くはなく、むしろ遙かに実務的な場でエキスパートが育っていると言える。つまり評価が実務的専門家により担われている点に特色がある。アメリカの行政システムではいわゆる**チェックアンドバランス**のメカニズムが普及していて、個々の行政機構はそれぞれ独自の政策展開を試みることが許されているが、それを互いにチェックする機構が別に用意されている。例えば大統領府は行政機構の統合的な政策をつくる頂点にあり、大統領府への政策提言は当然のことながら行政機構から上がって来るが、それと同時に行政機構を介

さないで、社会から、つまり民間の有識者から直接大統領への提言がもたらされる、そういうメカニズムももう一方で持っている。そしてまた、このような行政の側の政策意図は、議会によってチェックされ、大統領府でまとめる予算案は、通常、議会によって非常に大きく修正を受けることとなる。議会の中にも内部でのチェック機構がある。もちろん上院と下院という大きな構図がそれに当たるし、また予算を決めていくメカニズムで考えると、Authorization の委員会と Appropriation の委員会に機能が2つに分けてある。Authorization の委員会でまずその予算の支出項目の是非が問われ、それが是とされたものについて、今度は Appropriation の委員会で予算額が議論されるという。このようなチェックアンドバランスの枠組みの下で、個別の評価は、担当部局毎にそれぞれ独自に行われる。

1993 年に成立した GPRA（Government Performance and Results Act：政府業績成果法）によれば、各機関毎に 5 年程度の Strategic Planning を立てその機関がつくった Strategic Planning の実績を毎年予算当局や議会に報告することが義務づけられている。このような、評価結果の循環的な反映メカニズムの中で、年々実績を改善していく方向にマネジメントや目標が微修正される。この年度毎の実績評価の結果は、次年度の予算要求に際して、参考資料として OMB に提出されるが、少なくともクリントン政権の下では、OMB はこの自己評価の結果を直接次年度の予算に反映させるということは意図していなかった。しかし、ブッシュ政権に変わってから、自己評価結果は OMB の査定の非常に重要な資料に位置付けられることになった。現在はパイロット・プログラムに関しての自己評価結果が予算に厳しく反映される情勢になっていて、そうした試みが予算全体に広がるかどうか注目されている。

このように各機関毎に行われる自己評価は、より上位の機関、あるいは予算過程のより下流側の機関によって再検討され、その結果一元的に OMB が予算額を決めるのではなく、議会がそれをもう一度見直す。その際も議会の分科会で多元的に見直すというように、いわばチェックアンドバランスのメカニズムの中で様々な意見が反映される仕組みになっている。

（5）韓国

3 年前に始まった韓国の評価制度は、米国と対照的で、いわば**一元的**、**統合的**に評価を行う方式である。研究開発の全プログラムとプロジェクトに関して、毎年 6－8 ヶ月かけてその次の年度計画の評価を行い、予算額まで決定する。2 段階の評価が統合的に行われる方式で、Authorization と Appropriation に相当する評価メカニズム

が用意されている。公的資金による全部のプロジェクトとプログラムを同一の評価システムの下で統合的に行うところに特色がある。

Authorization のプロセスでは大統領の科学技術諮問会議が中心になって、そこにいわゆる Eminent Committee を形成し、著名な研究者や研究管理者たちによって提案されたプロジェクトが採りあげるに値するかどうかの価値判断が行われる。第2段階の予算額を決めるいわば査定のプロセスでは、韓国科学技術評価院 Korea Institute of Science and Technology Evaluation and Planning （KISTEP）が中心になり、技術専門家パネルを使いながら予算額を最終的に決める。従って KISTEP は研究機関からみると大変怖い存在で、その反動として、議会で行われる KISTEP 自体の予算を決めるための査問会議では、研究機関からの情報をもとにした厳しい査定を受ける運命にある。こういう大きな摩擦を伴いながらも、それを統合的にやり抜いていくというのが韓国の大統領制の下での統合的な評価制度である。

このような国による違いについては、例えば Kuhlmann や Luukkonen もそれぞれ独自の視点から分析をしている。巻末の文献（参考資料４）を参照されたい。

２．２　我が国の状況に適合した理念

我が国の評価システムはどのような統合的理念の下で展開されるべきであろうか。この問題は実は非常に重要で、我が国の現在の状況や組織風土に適合しない理念を採用したとしてもそれは定着しないであろうし、またもし明確な統合的理念を定めないまま評価システムを設定した場合には、そのシステムの内部にお互いに矛盾するメカニズムを含んでしまうおそれがあるからである。厳に慎まなくてはならないことは、外国で実施されている評価制度の部分部分を取り寄せてきて、それらをいわばパッチワークのように貼り合わせるといったやり方である。

それでは、日本の現在の状況と社会システムあるいは組織風土に適合する理念とはどのようなものであろうか。この問題を考えるために、多少日本の研究開発評価制度が展開されてきた歴史について見てみる必要がある。

研究開発評価の必要性は民間企業においてまず始まり、殆ど同時期に公的資金の評価に関してもその必要性が起こったと言ってよい。民間企業の中央研究所の設立ブームは 50 年代の末から 60 年代の中頃にかけてであり、研究開発評価の実務的な必要性がその頃初めて認識されたと言ってよい。公的資金による研究開発評価への本格的な

取り組みは、大型プロジェクト制度等の研究助成制度が設定されてからであると理解出来るとすれば、少し遅れてはいるが、公的資金に関してもほぼ同じ頃にプロジェクト評価の必要性が生じたと言ってよい。それ以前は産業技術政策の中心は技術導入と産業構造の整備にあって、本格的に新規の研究開発課題に取り組む体制にはなっていなかった。

さらに、その後の研究開発評価制度の展開に関して、まず民間企業の側から見てみよう。図２－１にあるように、まず、導入された評価法は Scientific Management（科学的マネジメント）を強化する方向に向かうが、オイルショックの混乱期を経て、その後に到達した方式は Human Relations Management（人間関係マネジメント）の側になっていた。各社で多少のバラエティはあるが、80 年代を通して展開された方法は、Human Relations Management を強化する方向であった。

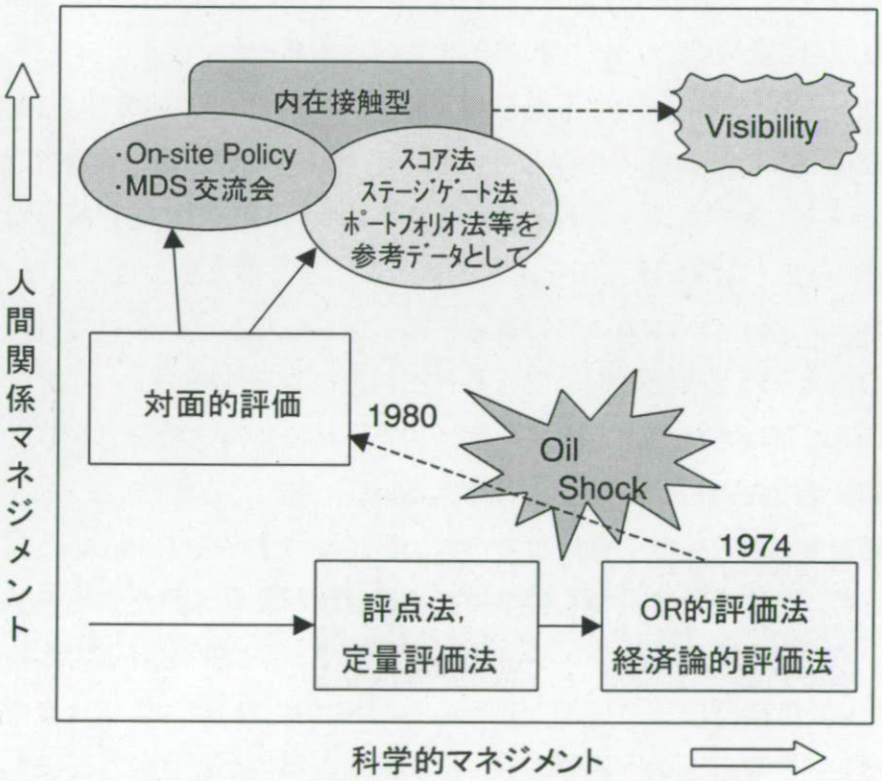

図2－1　我が国の民間企業における研究開発評価法の展開

例えば評点法を用いるにしても、その評価の結果は参考資料としての役割にとどめ、また、ポートフォリオやステージゲート法を用いるにしても、決してその位置や評点によって一義的に評価を定めるのではなく、ポートフォリオの位置を示す評点の構成要素の値や、ステージゲートの評価を構成する値そのものを素材として、先ずその妥

当性に関し評価者と被評価者が対話を交わし、プロジェクトの状況把握を巡って互いに意思疎通を図ることを目的とする。このような相互検討プロセスを経て検討結果が評価結果として共有され定まる。

公的機関の側の研究評価についても、似た状況が展開されてきた（表２－２）。71年に出された、科学技術会議の５号答申では日本が直面している課題を解決するための、新たな科学技術領域として、「ソフトサイエンス」の必要性について強調している。そしてこの時期に第１次のシンクタンク・ブームが起こり、研究開発評価法の学習も本格的に始まった。しかしこの時期に学習した評価法は、典型的にはOR的評価法のような柔軟性のない硬い評価法が中心であり、オイルショックという環境条件の激変に、そのような方式では追随できなかったばかりか、このような硬い方法が日本の風土にも適合せず研究開発評価そのものも定着しないまま、評価そのものが一時見捨てられることになった。２度のオイルショックの後、85年には政府をあげて本格的な行政改革に取り組むこととなった。いわゆる土光臨調である。

それを受けて、科学技術会議に下部組織を設け、３年かけて「研究開発評価指針」をまとめ、国立研究所にその導入を促した。しかし、これもまた殆ど定着しなかった。その原因は、研究開発マネジメントが成熟していない当時の国立研究所の環境条件の下で、評価法のみを、しかも評点法を定量的評価法として導入しようとした点にある。しかしながら、国立研究所はテーマの選定に関し何もしなかったわけではなく、その国立研究所の担当分野と規模に応じて、それぞれ独自のヒューマンサイド・マネジメント・システムを発展させて、定性的な評価法をそれに組み込む努力をしていた。この傾向は民間企業で見たものと同じものである。

多様な評価対象に取り組むようになるのは最近になってからで、95年に科学技術基本法が制定され、翌年それに基づく科学技術基本計画が策定されたことを契機としている。その中で、研究開発マネジメントの目玉として、研究開発評価の本格的導入が唱われ、新たに研究開発評価の指針、いわゆる「大綱的指針」が97年に策定された。ここでは主としてプロジェクト評価の他に機関評価の導入が促された。また、行政全体に対する評価の取り組みは97年から始まり、いわゆる行政改革の一環として、策定されることになった。つまり、研究開発に関してはそれに先行して評価への取り組みが導入されたことになる。そしてまた、この行政改革の一環として、国立研究所の多くが独立行政法人となり、その独立行政法人の実績評価が法律によって義務づけられ、2001年度から実施に移された。97年度から実施された大綱的指針に基づく機関

表2－2 我が国における公的研究開発評価制度の展開

65 農林省研究所レビュー
66 通産省大型プロジェクト評価
71 科学技術会議5号答申（ソフトサイエンスの振興）
84 科学技術会議11号答申：研究評価の充実
85 臨時行政改革推進審議会答申（科学技術政策大綱提起）
85 科学技術会議政策委員会・研究評価指針策定委員会の設置
86 「研究評価に関する基本的考え方」、「研究評価のための指針」（研究評価技術策定委員会）
87 科学技術会議政策委員会：大規模プロジェクト評価の検討の進め方
87 科学技術会議13号答申：国研問題
89、90 厚生省評価マニュアル
92 科学技術会議19号答申（ソフト系科学技術の研究開発基本計画）

---

95 科学技術基本法公布・施行
96 （第1期）科学技術基本計画（閣議決定）
97 国の研究開発全般に共通する評価の実施方法の在り方についての大綱的指針（内閣総理大臣決定）
97 郵政省・農水省・通産省・科学技術庁・文部省などの評価指針
98 厚生省・環境庁・運輸省・防衛庁など評価の指針
98 中央省庁等改革基本法（施行）
98 科学技術会議による評価実施状況のフォローアップ（以降毎年実施）
99 各省庁設置法等からなる中央省庁改革関連法成立
99 独立行政法人通則法（公布）、01（施行）
99 経済産業省における政策立案・評価指針など

00.04 産業技術力強化法第10条：資金重点化
00.12 経済構造の変革と創造のための行動計画
（第3回フォローアップ：閣議決定）
01.01 新府省体制発足
01.03（第2期）科学技術基本計画（閣議決定）
01.04 独立行政法人発足
01.06 行政機関が行う政策の評価に関する法律（政策評価法）成立（02.04施行）
01.11 国の研究開発評価に関する大綱的指針（内閣総理大臣決定）
02.03 総務省情報通信研究評価実施指針
02.03 防衛庁研究開発評価指針
02.04 経済産業省技術評価指針
02.05 文部科学省における研究及び開発に関する評価指針
02.04 経済産業省における政策評価基本計画など

農水省	01.04 農林水産省における研究・技術開発の政策評価に関する指針（さらに改定検討）
環境省	01.06 環境省研究開発評価指針（02.04改定予定）
厚生労働省	厚生科学研究に係る評価の実施方法に関する指針』（01.12以降改定検討）
国土交通省	国土交通省研究開発評価指針（仮称）（02.04策定予定）
警察庁	科学警察研究所研究評価実施要領（02.04目途）
財務省	未定（02.06目途に策定）

評価の多くでは、評価パネルの構成が“評価対象である評価機関自らが委嘱した評価者からなる”という評価システムであり、いわば仲間内の評価で終わってしまう危険性をはらんでいた。これに対し、独立行政法人の評価に関しては、所管大臣が 3～5 年の中期目標を独立行政法人に与え、機関の側はその実施計画を立て、その実績評価を大臣が任命する評価委員会によって行う、という方式になった。この場合にはドイツのシステム、つまり上部機関が任命した評価委員によって下部機関が評価され、それが上部機関に報告されるという仕組みと同じであるが、アメリカで実施されている GPRA と異なり、研究機関の側が目標を定め戦略的計画をつくるのではなく、所管大臣によって 3～5 年という比較的短期の目標が与えられることになっている。しかしながら、実際には中期目標の原案は主として研究機関の側で立てられ、それを上部機関に持ち込んだ後、上部機関の意向も踏まえて、所管大臣から正式に目標として与えられるという過程をとっている。ここで実施されている前段の部分は、法律では決められていないが、このプロセスを付け加えることによって、評価者と被評価者が一体となって政策やその目標を立て、評価を行うという方式に近づくことになる。

つまり、70 年代の後半から 80 年代にかけて、民間企業で展開してきた方式、これは評価者と被評価者が一体となってその機関が目指す共通の目標に向かって、研究や事業を展開するために、お互いに支援的に評価をしていくというやり方が、「大綱的指針」においても取り入れられ、そしてまた、独立行政法人の評価においても実質的にはそのようなメカニズムで運用されるようになっていると言える。

このような、評価者と被評価者が一体となって評価に取り組む支援的な評価システムであるところに、日本の社会や組織に定着しつつある評価制度の原理を見ることが出来る。しかし、このような方式は、評価者と被評価者が共通の目的を容易に描き得たキャッチアップの時代には有効であっても、フロントグループの一員として固有の課題を認識し、それらに対して戦略的に取り組まねばならない状況下にあってはそのままでは機能しないと思われる。

まず、機関や政策が目指す適切な目標を明確に立てておく必要があり、そのための戦略や計画の体系が整備される必要がある。そのうえで、公的資金に依存する場合では、少なくとも出資者である国民に対しての情報開示や、運営の透明性が確保される必要性がある。つまり visibility が確保されることが、仲間内の評価に終わらないために重要であるし、また民間企業の場合でも研究開発部門以外からも理解出来るような、そういう透明性 visibility が必要になる。

現在展開されている評価システムの多くは、このメカニズムを欠いたままであって、

そこに重大な欠陥があると思われる。評価者と被評価者が一体となって取り組む方式が展開された背景は、終身雇用制度にある。この方式が、終身雇用制度の下で研究者の能力を最大限に発揮させることにおいて優れていても、このような支援的な評価システムだけでは仲間内のなれ合いになってしまう危険性がある。自己満足に陥らないために他部門、他のセクターからの批判を受け入れる方式である必要がある。

結論として日本の評価システムの背景的な理念をどのように定めるべきかという問いに関しては、これはいわば参加型の、評価者と被評価者が一体となって評価に当たる inclusive（システムの内部に包含する：「**内在的**」）な評価制度であって、評価者が被評価者を支援し伸ばすことを本旨とする支援的な評価制度である。一方でこの方式が内包する欠陥を補うための、visibility（運営の内実が明示的である：「**明示性**」）の確保であるとか、あるいは、上部機関や第三者機関によるチェックであるといったような、第三者からのフィードバックがかかるメカニズムを補う必要がある。そのような補助装置が付けられるならば、サポーティブな評価制度は研究者の自発性を活かした非常に望ましい制度であると考えられる。このような評価理念の下で、その効果を活かしながら、研究者にも、また出資者であるところの国民の側にも信頼される評価制度として、このような理念に裏打ちされたシステムを定着させていくことが期待される。

第３章　評価の仕組みと推進体制

３．１　評価推進体制の全体像

まず、研究開発評価の推進体制全体についてその概要を理解する必要がある。ここでは研究開発評価の推進体制の全体像を、２つのレベルに分けて説明する。第一は、研究開発評価を担う組織の外側に配置されるべき様々な体制であり、第二は研究開発評価組織の内部で培われていなくてはいけない仕組みについてである。

通常「評価」といえば第二のレベルの内容しか取り扱われないことが多いが、評価活動と無縁であった者にとっては、この内容の理解だけでは不十分である。その理由は、我が国の教育体系のあり方に根差していて、我が国では第一のレベルに相当する内容が高等教育機関のいずれの学科においてもほとんど取り扱われることが無いからである。ちなみに海外の研究評価に関する研修コースのカリキュラム内容は第二のレベルに特化しているが、これは当該国における第一レベルの教育内容の普及を前提としている。従って我が国で勉学した者が予備知識無しに海外の研修コースを受講しても、そのコースがめざす研修内容の本質を理解することは困難であろう。

「評価」と「意思決定」とは区別されなくてはならない。**評価**の役割は、評価すべき対象が置かれている状況や対象の価値を明晰にすることである。その明晰にされた知見を用い次の行為に移るべきかどうかの判断を下すのが**意思決定**である。評価が明晰であるかどうかの判断（評価活動における評価の決定）とこれは別物である。

さて、評価の観点から見た場合、行政組織は、３つの機能を担う内部構造に分類できる（図３－１）。第一は、意思決定をする査定部門である。通常これは官房に集約されている。第２は、研究開発評価を担う部署であり、これは２つに分けられる。通常、「評価課」と呼ばれている部署は、省全体の評価体制や評価システムを整備し、全体の評価結果をとりまとめたり新しい評価制度を導入するための様々な検討を行い、評価システムの支援部署と位置付けられている。また、この部署では実際の評価を担当することがあっても、省全体にまたがる横断的な評価項目を担当し、例えば各課や個別の政策担当部署で個別に展開された既存のプロジェクト群の中間評価や事後評価を取りまとめて行う場合等である（官房に「政策」や「独立行政法人」等のより高次の横断的評価項目を担当する部署が置かれていることも一般に見られる現象である。し

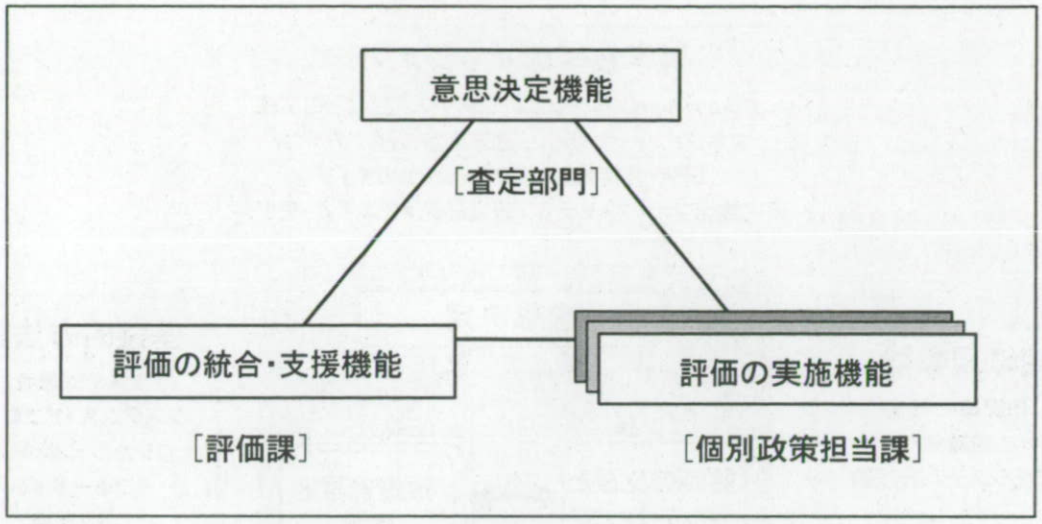

図3－1　行政組織内部の評価関連機能

かし研究開発に関わる事項については、科学技術の「評価課」が評価手法に関し十分に支援すべきであろう)。

これに対して、大部分の行政組織の個別政策担当課は所掌分野の政策を形成し、そしてまたそれを執行していくという役割を担う部署と考えられるが、このような部署は政策を形成する際にも、またある制度のもとで研究開発プロジェクトを展開する際にも、自らが先ず事前評価を行わなくてはならない。従ってそれぞれの担当部署や課は、評価の実施機能を備えそれを担うべき組織である。個別の政策担当課はこのように評価スキルを十分に修得する必要があり、その実現のためには本格的な研修態勢を整備する必要がある。

このように、関連する機能の観点から評価についてみると、評価の支援機能と評価の実施機能、そしてさらに評価結果を踏まえて意思決定を行う査定機能という３つの機能に分けることができる。

さてこのような機能を担う組織で評価が行われ、またその結果が利用されるに際し、研究開発評価の外部装置として用意されなくてはならないことは、少なくとも、図３－２に示すように４点ある。第一は、評価論そのものであり、研究開発評価論という研究開発に特化した評価論に比し、より一般的な、評価一般に関わる仕組みについてである。第二は、評価が研究開発に関わることから、研究開発や科学技術のマネジメントをどのように執り行うべきかということに関してのワンセットの知識である。三番目は研究開発を評価すると言うこと自体、組織に関わる現象を取り扱うことである

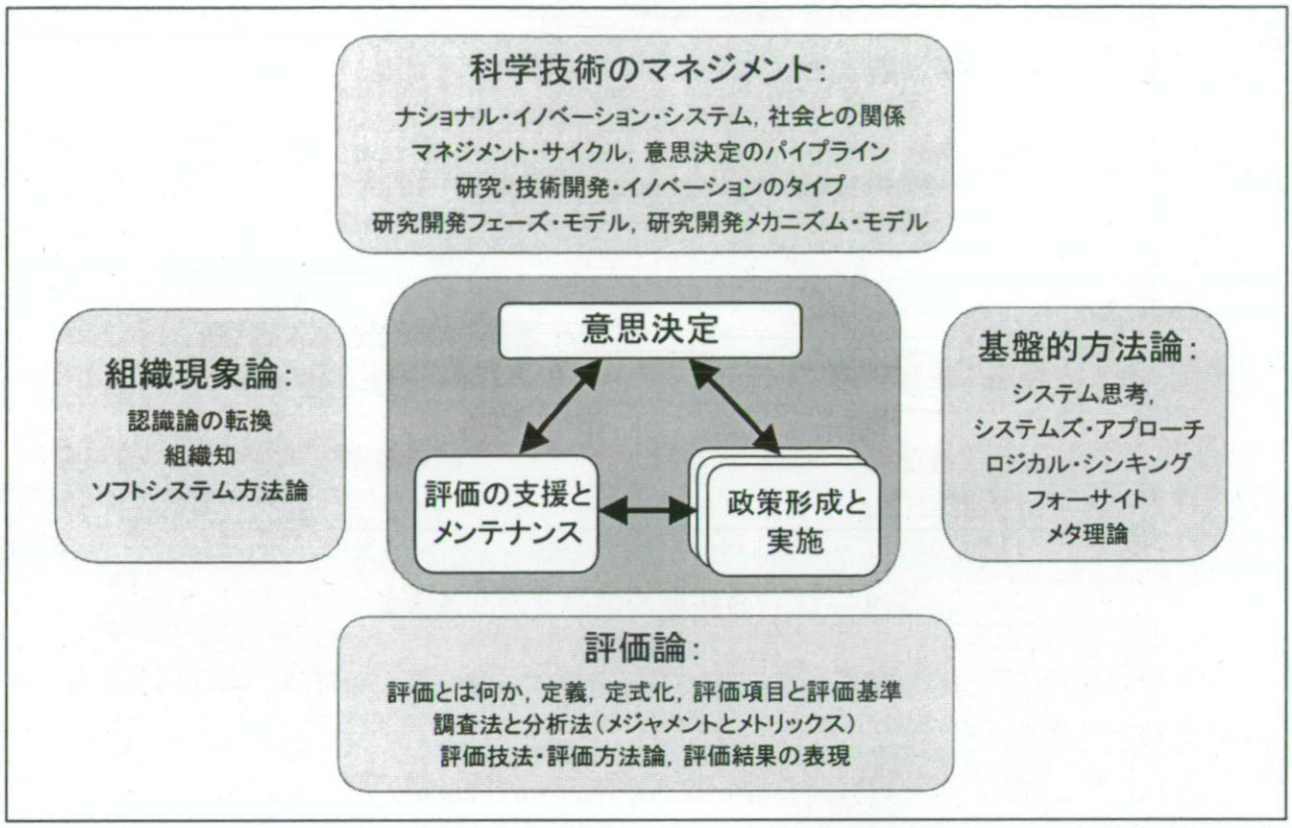

図3－2 研究開発評価の外的枠組み

から、そのような「組織現象」を扱うための基礎的な知識である。そして最後の四番目には、例えばシステム論のような、研究開発評価法を運用するための思考世界に関わるより基盤的な知識である。

また、研究開発評価の内部構造はどのように整理すべきであろうか。図３－３にまとめたように、まず第一には研究開発評価システム全体を貫く評価理念が必要であり、これがもっとも重要である。2 番目には評価の対象の側をどのように捉えるのかということであり、また 3 番目には、その対象に関わって実際に評価を行うための、データをどのように測定・収集し分析するのか（メジャメントとメトリックス)、そしてその結果をどのようにして評価に結びつけて行くのかという評価の方法論についてである。最後に 4 番目としては、この対象の理解や評価法を担う専門家たち、つまり人材のあり方や評価作業の分担の問題である。

以下では、研究開発評価の外側にある問題は巻末の「参考資料１及び２」の項で詳しく説明するとして、研究開発評価の内部に関わる問題に関して次項以下で詳しく見ていくことにする。

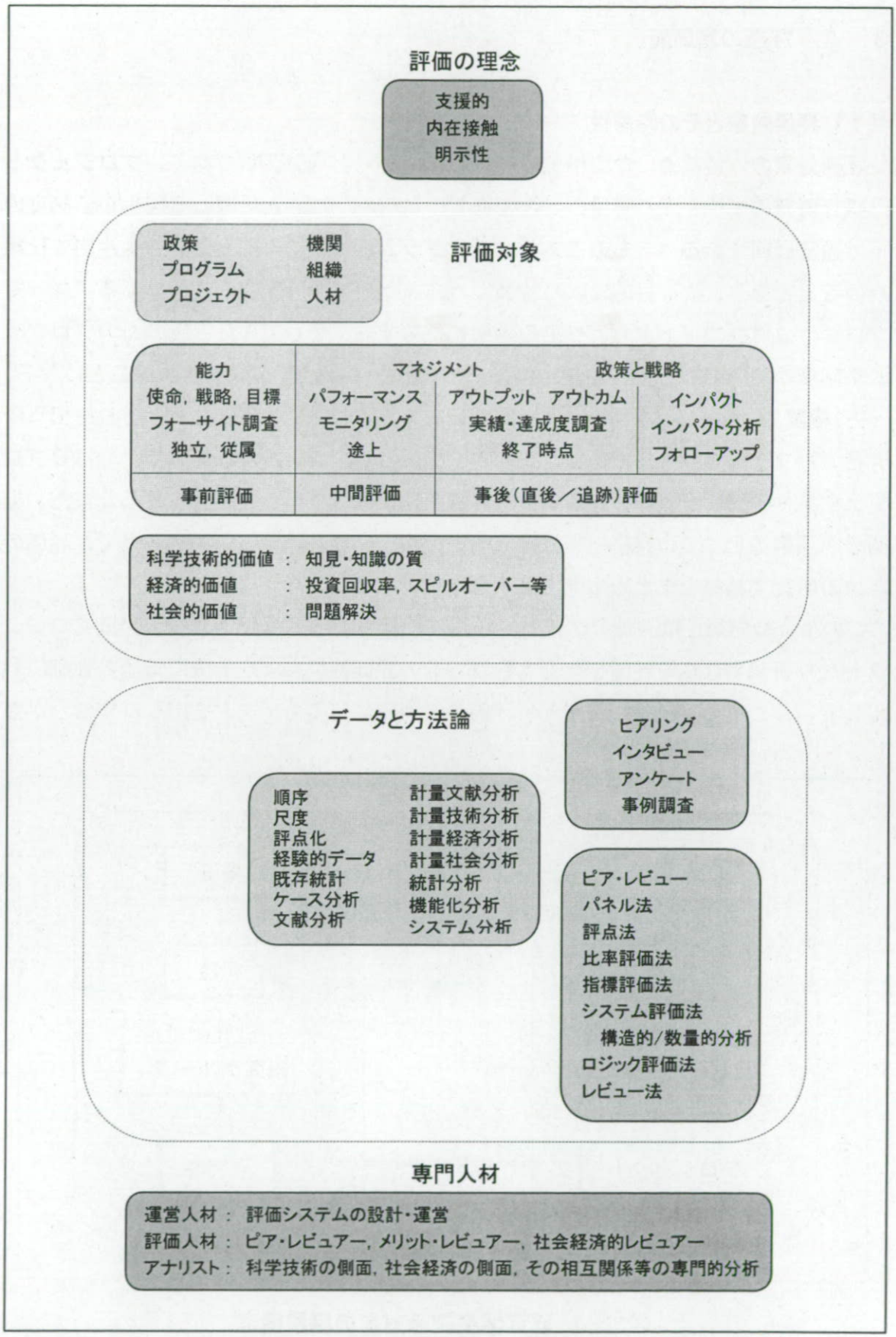

図3－3 研究開発評価の内部構造

３．２　評価の諸局面

（１）評価対象とその階層性

評価対象の一番基礎になる単位は「プロジェクト」と「研究者」である。**プロジェクト**は研究開発の単位となる事業のことであり、「プログラム」と呼ばれる何らかの制度の下で通常展開されるべきものである。**プログラム**とは、プロジェクトを生み出す仕組みのことであり、政策目的により特徴づけられた研究開発制度のことである。従ってプログラムはプロジェクトより上の階層に位置する。そしてまた、幾つかのプログラムや制度が組み合わさり、より上位にある「**施策**」や「**政策**」が展開されることになる。一方「**機関**」の側をみると、その機関において実施されるプロジェクトとそれを担う研究者のパフォーマンスを総合したものが、機関としての実績になるので、やはりプロジェクトや研究者が機関の評価単位や評価基盤になっていると考えられる。また、機関の内部構造としての業務担当組織（研究グループ等）はその中間にあって、評価のための研究業務等をまとめる際に重要な単位となる（図３－４）。

このような階層性は評価のコストを縮減する観点からも重要である。一度プロジェクトなり研究者なりを評価したとすれば、その評価結果をより上位の階層の評価に利用していくことによって、研究者が評価の階層毎に何度もデータを出し直すことが無

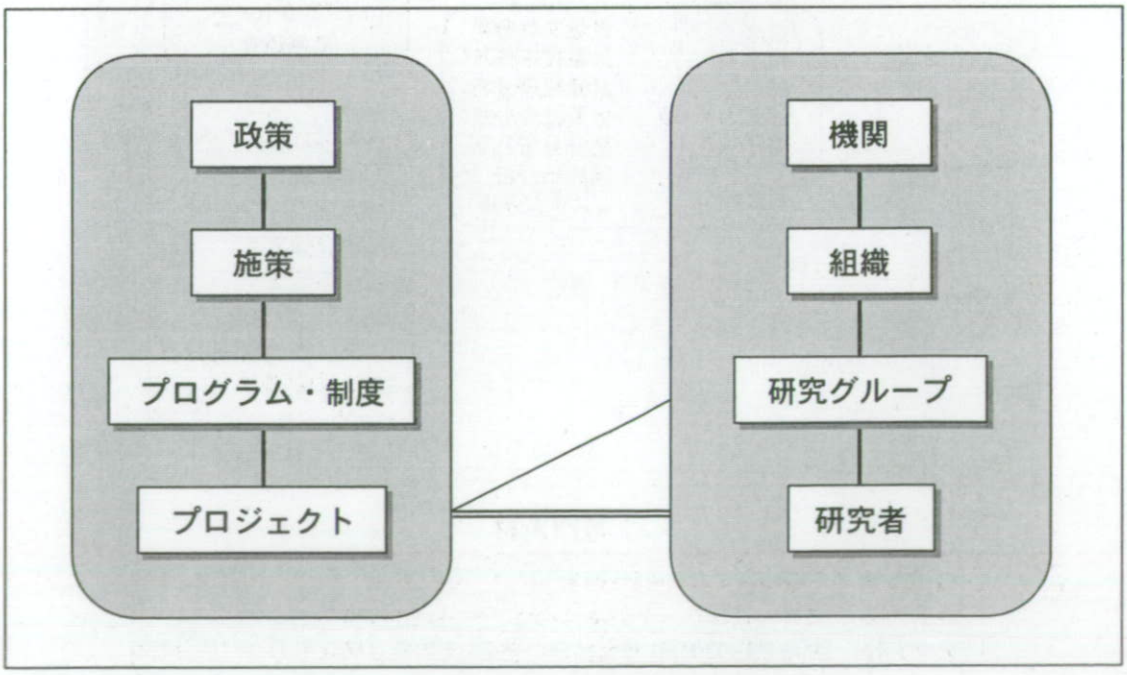

図3－4　研究開発評価対象の階層構造

いようにすることができる。そのためにも、この階層性を踏まえた評価データベースを作る必要がある。とはいえ、各階層毎に階層固有の評価情報を付加する必要が通常生ずる。しかしその様な情報は当該階層の評価担当者が作成すべきもので、研究者レベルまでその都度遡及すべきではない（詳しくは後出の図３－８、９を参照のこと）。

プロジェクトに関しては、このような階層構造の中で、しかしながら２種類の異なる性格を持ったカテゴリーが存在する。第一はプログラムのもとで策定されるプロジェクト（**従属プロジェクト**）であり（図３－５）、第二はプログラムを設定することなくプロジェクトを独立に立てる場合（**独立プロジェクト**）である（図３－６）。従来日本の行政機関の中ではプログラム概念が薄弱であったために、プログラムのもとで打ち出されるプロジェクトは多くない。多くのプロジェクトは独立に政策担当課の司々でその都度構想され、個別に打ち出されてきた。この「独立プロジェクト」と、プログラムに基づいて展開される「従属プロジェクト」では、図３－５、６に示したように、評価のコストからみると非常に大きな差がある。次項で述べる理由により独立プロジェクトの場合には、評価コストが非常にかかるため、独立プロジェクトを可能な限り少なくして、研究開発制度をすべてプログラム化する方向が、世界の潮流になっている。例えばドイツでは４年前に社民党連立政権に移行した際、BMBF のすべての

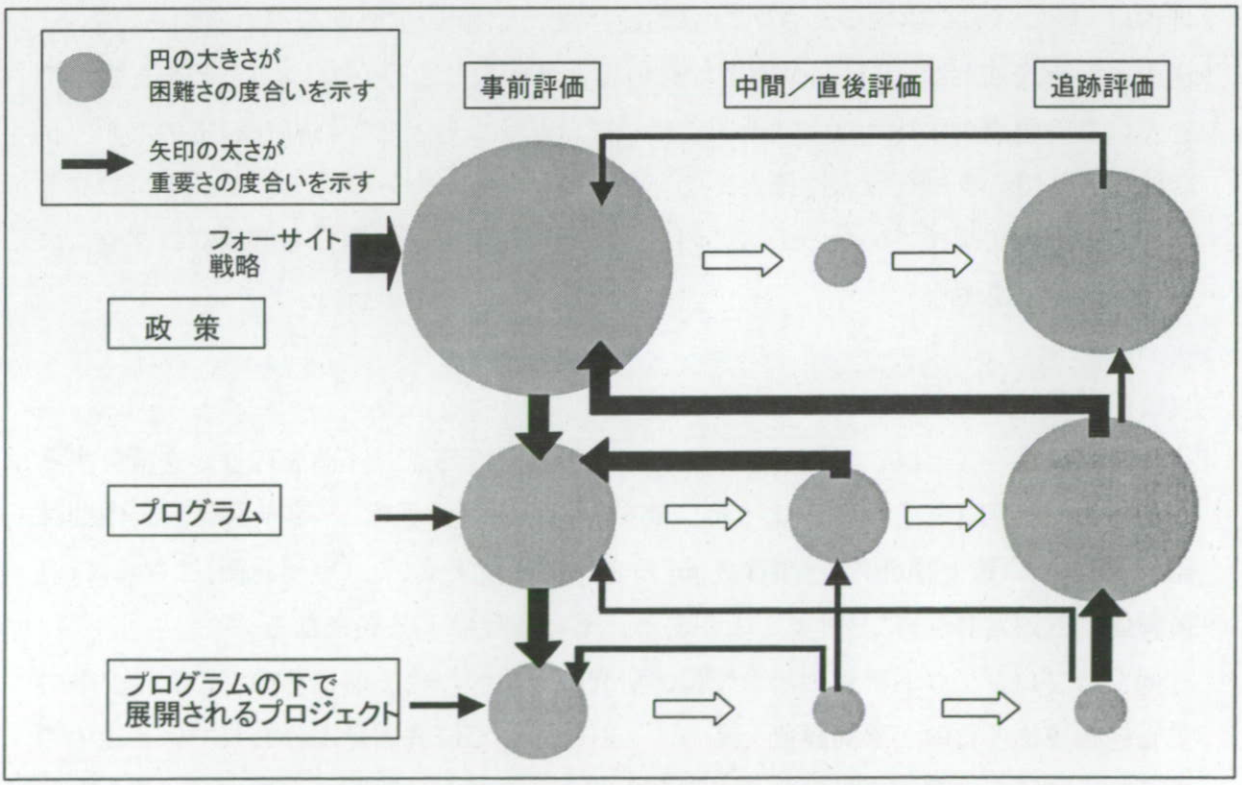

図3－5 評価の階層性と評価の困難さ(1)

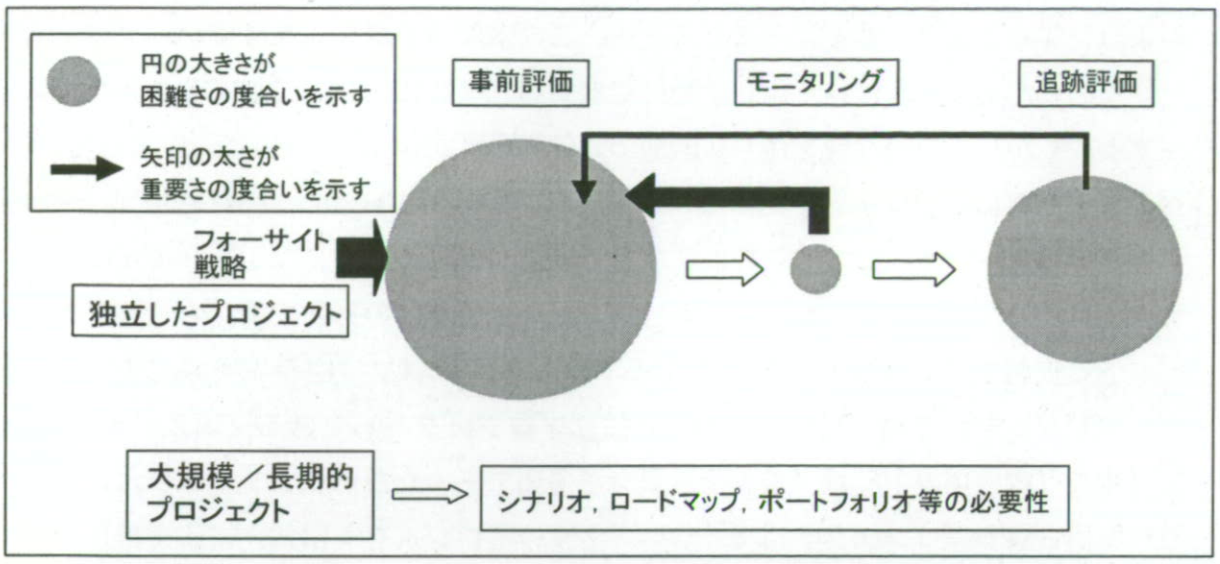

図3－6　評価の階層性と評価の困難さ(2)

研究開発制度はプログラム化された。このプログラム化の意味について次に述べる。

プログラムの定義として、“政策目的によって特徴づけられた研究開発制度”であると上で記した。ところで、政策目的の他に、研究開発制度を特徴付ける属性としては、例えば科学技術領域や研究開発手法等を挙げることができる。もちろんこれらの属性を定義の中に付加してもよいが、属性の中で「政策目的」だけを定義の必要条件とした理由は、後で詳しく述べるように、「目的適合性」が評価の主要部を占めるからである。逆の言い方をすれば、“目的が明確でない場合評価のしようがない”ことを主張している。この観点から言えば、我が国には“プログラムもどき”は存在していたが、正統的な意味でのプログラムはほとんど存在していなかったと言える。ここでプログラムの例をあえて挙げていないのはこのような理由による。しかしもはやプログラムの概念は明確であろう。

（2）評価のフェーズ

評価のフェーズとは、評価対象の時間的経過のどの時点で評価を行うかを区分する概念である。図３－７に示すように、評価のフェーズは通常、「事前評価」、「中間評価」（ないし「途上評価」）、「事後評価」の３つに区分するが、「事後評価」はさらに「直後評価」と「追跡評価」に分けることがある。その場合は４分類となる。

例えばプロジェクト評価の場合では、研究開発が行われる前にそのプロジェクトの妥当性を評価するのが**事前評価**であり、これに対して研究開発が行われている途中の段階での評価を**中間評価**ないし**途上評価**と言う。また、研究開発のプロジェクトが終

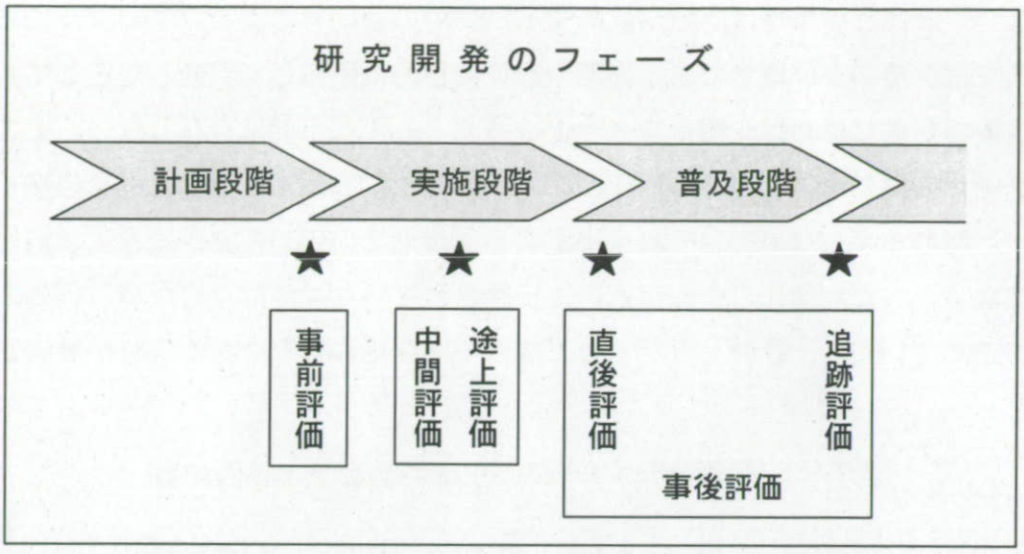

図3－7　研究開発のフェーズと評価実施時期の関係

了した後で行われる評価のことを**事後評価**というが、事後評価は研究開発が終わった時点で評価を行う**直後評価**と研究開発が終わってしばらく、例えば5年ぐらい経過した後で行う評価を**追跡評価**と言う。

このような評価の時点と階層の異なる評価対象との組み合わせの種類によって評価として取り組むべき内容が大きく異なる（表3－1）。

まず、事前評価の場合を考える。独立したプロジェクトや新たに設定するプログラムや政策に関しては、それらのターゲットや目的あるいは大目標や使命等の妥当性を評価することになるが、従属的なプロジェクトの場合では、設定されているプログラムの目的にその提案されたプロジェクトの目的が適合しているかどうかを評価すればよいので、この場合にはプロジェクトの目的自体の妥当性を問う必要がない（図3－5、6）。この差は実は大変大きい。

目的や目標自体を評価するには、それによって実現されるであろう成果の「費用対効果」を予測し、他の代替的な目的に比しそれが優れているであろうことを明らかにする必要がある。この作業を何らの制約的な枠組み無しに完遂することは論理的に考えてもほとんど不可能であろう。従って通常上位の目的をまず定めその枠内に下位の目的があるかどうか、或いは提案された下位の目的が上位の目的に適合しているかどうかを判断し、下位の目的自体の妥当性を改めて問うことはしない。さてその場合、上位の目的はどのようにして定めればよいであろうか。

ここでは結論のみを述べておこう。詳しくは巻末の参考資料1を参照されたい。「目的（上位の）」を評価するためには、その前提として「計画」が立てられていなくては

ならない。計画の妥当性は計画が依拠する「戦略」の適切さによる。プログラムの目的については、その上位概念である施策や政策の目的に照らして判断したとしても、施策や政策の目的にはもはや照合すべき上位概念としての目的は存在しない。そこで戦略や計画が必要になる。ところで、このような最上位の政策目標は、社会の多くのセクターに関わる事象なので、それを決定するプロセスの正当性（手続き上の適性）が問題になる。つまり国民の意思が適切に反映されていなくてはならない。評価の問題を突き詰めていくと、このようなメカニズムで、評価の外部にある戦略や計画を立て

表3－1　評価カテゴリー毎の主な評価目的と評価内容

<table>
<tr><th rowspan="2"></th><th rowspan="2">事前評価</th><th rowspan="2">（途上評価）
中間評価</th><th colspan="2">事後評価</th></tr>
<tr><th>直後評価</th><th>追跡評価</th></tr>
<tr><td>従属プロジェクト</td><td>目的の適合性
［上位の枠組みとの照合］</td><td rowspan="5">目的や計画に対する到達度や達成度
［実績の評価，つまりパフォーマンス（研究の内容：「費用対効果」，「実績内容」，「コストの内容」）とマネジメント（研究開発の運営のあり方）の評価］
［対目的比，対計画比の他に、対類似他者比，対チャンピオンデータ比，対過去比等の採用］
［評価結果の計画や設定目標あるいは次期計画や上位の目標等の見直しへの反映］</td><td rowspan="2">通常「プログラム」の追跡評価の中で行われる</td></tr>
<tr><td>プログラム
政策・施策
独立プロジェクト</td><td>目的や目標自体の妥当性
［代替性等を含めた調査分析を前提とする］</td><td rowspan="3">成果の最終確認と設定された当初目標の妥当性
［成果全体（直接/間接，目的内/目的外），つまりアウトプット，アウトカム，インパクト，副次効果等の分析に基づく最終評価］
［目的や計画等の見直し材料に基づく上位目標や次期計画等の見直し］
［事前，中間，直後評価のあり方に対する評価］</td></tr>
<tr><td>最上位の政策</td><td>目的や目標自体の妥当性
決定プロセスの正当性
［フォーサイト等の社会に開かれた評価システム］</td></tr>
<tr><td rowspan="2">機関</td><td>機関の新設は、「政策」や「施策」課題に類する</td></tr>
<tr><td>機関内で発足させるプロジェクトの事前評価は「従属プロジェクト」の事前評価と同一</td><td>機関内で発足させたプロジェクトの追跡評価は「従属プロジェクト」の追跡評価に類する</td></tr>
</table>

るシステムと関わることになり、評価のみで閉じている思考では完結しない。

さて、次に中間的な評価と直後評価、つまり途上評価と直後評価についてみてみよう。評価で取り扱う内容としてはこの両者はほぼ共通しているので、それらをまとめて考える。途上評価の場合には、あるいは直後評価の場合でも、事前に設定された目標なり到達点に対して、実績として結果が現実にどこまで到達しているか、つまり達成度とか到達度とかの実績を評価することになる。これはモニタリングに相当し、その進み具合を観察することは、それほど困難ではない。事前に設定されている目標に対して結果がどのような位置にあるかを照合すればよいので、この作業は基本的にそれほど難しくはない。

また、この場合の評価内容のカテゴリーには、まず実績の内容として、「費用対効果」つまりコストパフォーマンス、さらに詳細には効果である「実績内容」と、「コストの内部構造」の 3 種類である。またもう一方で、研究開発の運営の仕方、つまりマネジメントの実績についても問われなくてはいけない。プロジェクトやプログラムだけではなく機関評価についても上記のことは基本的にあてはまる。異なる点は、機関評価の場合、評価の視点が、「対目的比」だけではなくそれをブレークダウンした「対計画比」に重点が移動する程度である。なお、目的や計画が照合出来るほど明確に定められていない場合（本来そうあってはならないが）、評価の視点として、補助的に「対類似他者比」（いわゆるベンチマーク等）や「対チャンピオンデータ比」、あるいは「対過去比」（対前年比等）等を援用することになる。

一方、途上評価と直後評価では、多少異なる点もある。途上（中間）評価の目的は、研究開発計画をそのまま進行させてよいかどうか（中止を含め）についての状況を明晰にすることであるのに対して、直後評価では達成状況を明確にし次期計画への参考に資する点にある。ただし評価時点が終了直後である場合、タイムラグをもって出現する成果（アウトカムやインパクト）に関する情報はこの時点では十分には顕在化していないので、直後の評価情報は最終的な情報とはいえない。従って設定した目的自体の適否について言及出来ることは限られている。この点に関しては次項の追跡評価の方がはるかに重要である。さらに、同種の問題であるが、途上の方が直後より利用できる情報が限られ、評価結果の信頼性に関し多少の違いがあるといえる。

これらに対して、3 番目のカテゴリーである追跡評価では、波及的な効果等も踏まえて評価できるので、本来的な意味でのフォローアップが可能となる。波及的な効果等の内容に関しては、次節において詳しく説明するが、いわゆる直接的な成果であるアウトプットやアウトカムの他に、間接的な効果としてのインパクトまでを含む。ま

た、目的内のこれらの成果の他に、目的外のいわば図らずも挙げることが出来た類の成果も分析する必要がある。これは、直接研究開発に携わった人以外の部分も含めて調査することを意味していて、その調査対象が広がるために、中間（途上）及び直後評価よりはるかに困難な分析が必要になる。

追跡評価の目的は、当初設定した目標や、さらに上位の目的が適切であったかどうかを評価することにあり、主としてプログラムの追跡評価が重要で、その上位の政策や戦略それ自体を見直す機会を与えることになる。また、独立プロジェクト評価であっても、その事前評価のあり方を見直す契機ともなる。

ところで、このような当初設定された目的の是非を評価する際に、その設定された時点で設定された目的が正しかったかどうかを評価すべきとする考え方と、追跡評価時点でどのような方向に見直せばよいかという、いわば見直す現在時点の状況に合わせて評価すべきという、2 つの考え方がある。この適否は、評価の目的に依存するわけで、その追跡評価の結果を政策の見直し、あるいはプログラムの設定の見直し等に繋げていくとか、事前評価のあり方を見直す等のためには当然、その追跡評価をする時点での状況に合わせて評価されなくてはいけない。しかし、その当初設定した政策形成者のパフォーマンスを評価する場合であるならば、その政策設定時点に照準を合わせて、その時点として正しい判断であったかどうかということが評価されてもおかしくない。但し、このようなケースはあまり重要ではなく、国民の側からみるならば、その政策の見直しとかプログラムの見直しのために追跡評価が行われることが重要である。

また、マネジメントに関わる内容であるが、追跡評価の時点で、事前評価や中間評価あるいは直後評価が適切になされていたかどうかといういわば「評価の評価」を行う必要性もある。評価委員も自分の言ったことが後で評価されるという意識を持ち、緊張感をもって評価に望むことが期待される。中間評価や直後評価においても信頼性の差はあるものの同種の効果を期待できる。

（3）評価の内容

ここでは評価の内容について詳しくみてみよう。既に述べたように、評価の内容を区分すると基本的には 4 種類になる。第一は目標、目的、使命（ミッション）、そしてビジョンや戦略さらには計画等の「目的系」の事象であり、第二はインプットの側の「コスト系」、そして第三はアウトプットの側の「成果系」、最後は内容そのものではなくやり方に関する「マネジメント系」である。また、これらには全て将来に関する側と

過去に関する側そして現在の事象とがあり、将来の側（事前評価）の評価では評価内容の予測値やリスクの大きさ等を計ることが中心であり、過去の側（事後評価）では実績や到達度が問われることになる。また現状に関しては（途上評価）過去の実績のほかに現在の状態つまりパフォーマンスとポテンシャルが問題となる。以下に、順次詳しくみていこう。なお、費用対効果のように上に述べた4種の評価要素を組み合わせることによって導かれる量についてはここでは十分には触れない。

○目的系

目的系の評価は、基本的にはその目指すものの妥当性をもって行う。妥当性を判断する指標としては、主として費用対効果を用いるが、その他に費用それ自身及び効果それ自身の質を問うこともある。つまり次に述べるコスト系と成果系の情報を動員することになるが、特に将来に関する目的系の評価を信頼できる形で実施することは極めて困難であることは容易に理解できる。そこで、既に述べたように欧米では、通常目的系の階層構造を活かした簡便な方式を採用できる枠組みの中で運用するように努めている。

使命や目的や目標（ターゲット）は、階層構造を成している。評価すべきターゲットの適否は、上位の目的の中でそれが占める位置に照らして判断することになり、またその目的自身はさらにその上位の使命 mission に照らしてその位置付けを評価することになる。このような枠組みの中で、採り上げたターゲットの妥当性が上位概念に照らして適切であるかどうかが第一の評価の視点である。この場合には、評価すべき目的と上位の目的との包摂関係をチェックし下位目的の内容にまで深く立ち入る必要はない。さらには2番目として、そのようなターゲットの他に代替的なそれに代わるターゲットと比較して採り上げるものが適切であるかどうかという代替性の視点からの評価がある。また、3番目にはそのターゲットをもし採り上げないとすれば、つまり実行しないとすればどのようなことが起こるであろうか、ということを想定して評価を行う非実行性の視点からの評価がある。

先にも述べたが、一般に事前評価の場合には、その使命、目的あるいはそのターゲットを明確に評価することは非常に困難であり、つまり将来にわたった予測をした上で評価しなくてはならないわけで、何らかの不確実性を必ず含むことになる。その意味でこの評価を厳密に実行することは非常に困難である。従って、目的系の階層構造が十分には整備されていない場合等では、評価可能な部分は最大限評価するとしても、不確実性 uncertainty を含んだまま目指す内容を実施することになる。その際、実施しながら目的自身を随時見直していくという、学習的な評価のあり方を採るのが一般

的で、特に最上位レベルにある目標に対してはモニタリングを定期的に行い、事前に評価したら最後まで見直さないという態度は避けるべきである。米国の GPRA ではこの循環的な評価方式を採用している（13 ページ参照）。

また、目標はそのレベルで定められた計画や戦略との整合性が必要である。そして最上位の目標については、既に述べたように社会に開かれた意思決定の場でオーソライズされる必要がある。この階層的関係を図３－８、９に示す。ここでは行政府内部の階層的な組織構造を想定し、最上位のレベルでは政治過程による民意の反映を意図している。しかしこの形態についてはこの他に様々な有効なメカニズムが存在し、また提案されてもいる（参考資料１）。

独立プロジェクトのように、このような階層関係や戦略計画が定められていない場合には、まず戦略やロードマップを設定する必要がある。これはもはや通常の意味での「評価」の枠組みをこえている。その詳細については参考資料１を参照されたい。

○コスト系

我が国では従来「成果系」の評価を中心にして行い、「コスト系」については評価においてほとんど触れることがなかった。しかし、費用対効果をみるうえでも「コスト系」

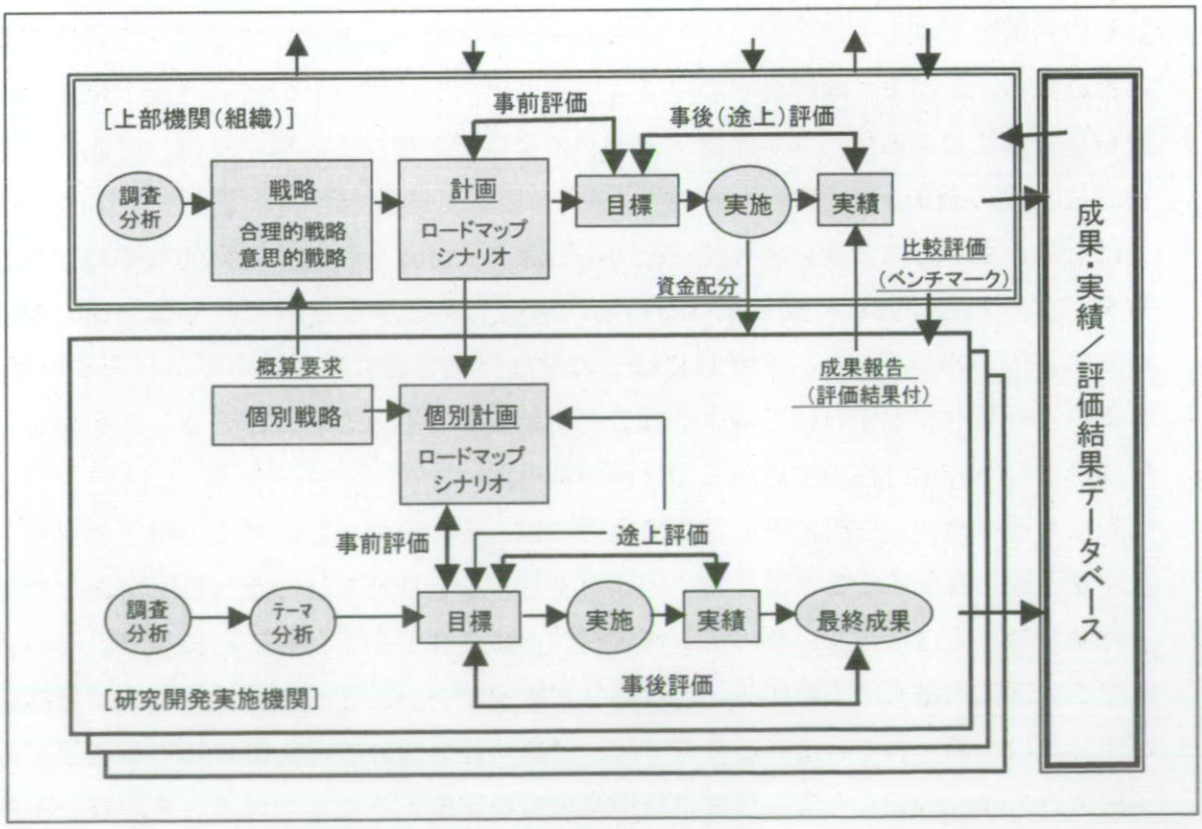

図3－8　階層的な組織構造間での評価のあり方—戦略，計画と評価

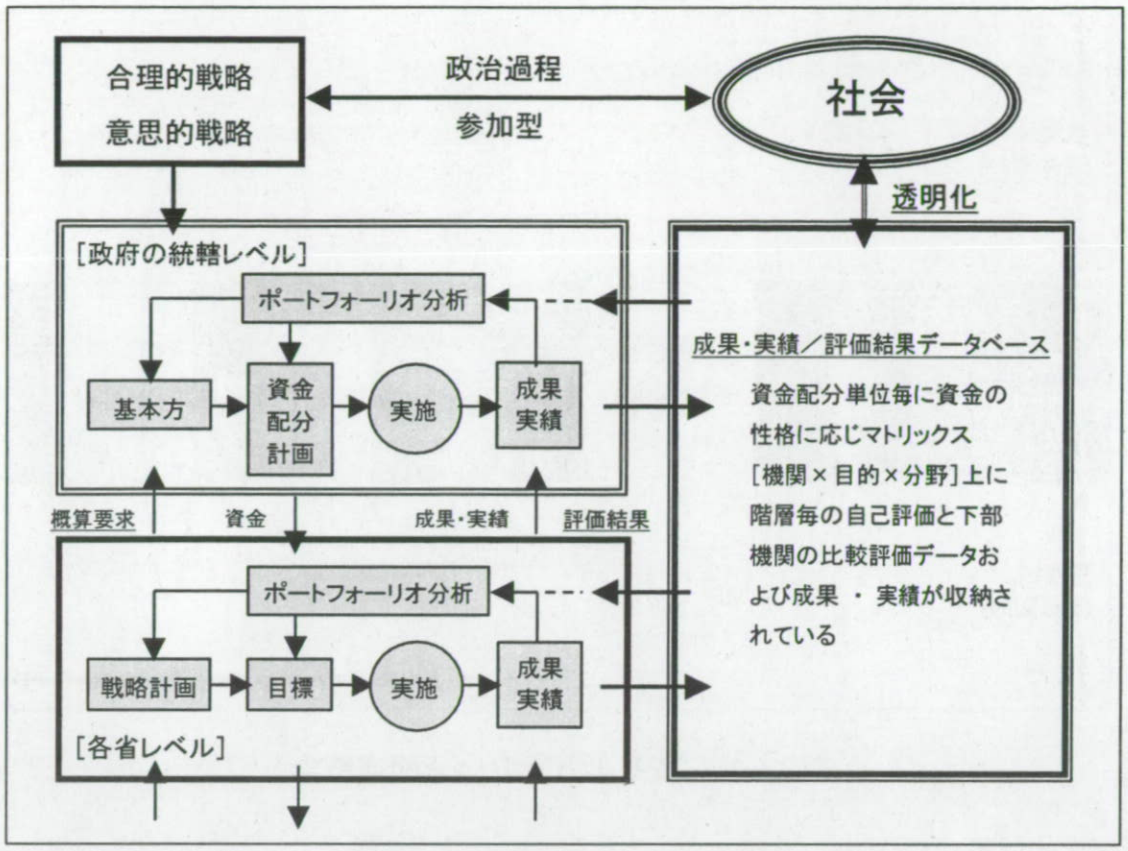

図3－9 階層的な組織構造間での評価のあり方
—最上位レベルでの戦略，計画と評価

の評価は欠かせない。

コスト系の構成要素は、研究開発に関わるリソース全てから成り、経費の他に人材、原材料、設備、施設、そして知的財産権のような無体財も含まれる。これらの評価に際しては、数量的な観点と質的な観点とがある。数量の実態を示す絶対値を把握したとしても、それだけでは評価にならない。対目標や計画比、対前年比、対類似計画比等のベンチマーク等何らかの比較の視点を導入することが必要である。また、質的には、通常これらの指標の内部構造の分析等が行われる。例えば、経費の中で管理部門が占める割合であるとか、導入した設備のマシンタイムの利用率等である。

○成果系

まず研究開発成果の内容をどのように区分しているかについて述べる。図３－１０に示したように、中間、事後評価の場合、まず目的として設定した範囲内で現れた成果（intended：「**意図的成果**」）と、目的外であり意図しなかった成果（unintended：「**非意図的成果**」）とに分ける。非意図的成果は基礎的な研究等においてはしばしば重要な成果になる場合がある。そのように目的として意図した部分の成果と意図しない

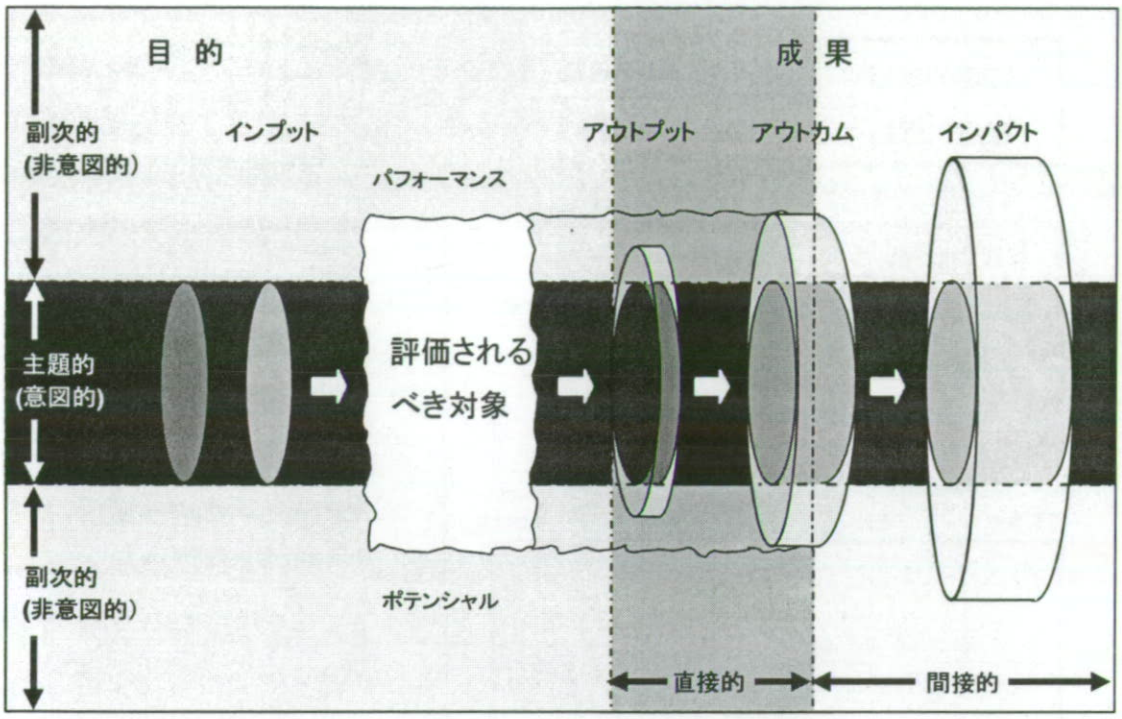

図3－10 成果のカテゴリーと関連概念

「**副次的成果**」に分けることができる。さらにもう一つの軸として、「直接的な成果」と、「間接的な成果」に分類する。この区分は評価論の研究者によりかなり定義に乱れがあるが、標準的には以下の定義を用いる。**直接的な成果**とは、研究開発に携わった人（共同研究者等も含む）が直接関わって挙げた成果と捉え、それ以外の部分は**間接的な成果**と考える。例えば研究開発成果を他の研究者が利用して挙げた成果は間接的成果である。しかし例えばBETA法の場合には、直接的な成果とは、科学技術に関わる成果であると定義し、経済的な成果は間接的な成果だと分類する。経済学を背景にした研究者にはこのような定義が馴染むようであるが、研究開発の目的が科学技術的な成果を挙げることに局在化しているわけではないので、この定義は一面的である。また、テクノポリスの定義では意図したものの直接的な成果とは、目的に設定した範囲の中でインパクトまで含めて捉える。この場合は従って直接的成果の範囲がかなり広がることになる。そしてそのインパクトの後に現れる成果、つまり目的として設定しなかった範囲の成果については間接的な成果であると考える。この場合はむしろインパクトの定義が一般と異なり、限定的に使われている。これらの定義の妥当性は次章で述べる評価法における利便性ないし有用性によって判断されるべきである。ここで推奨する定義の有用性は、後で分かるように、成果の収集が研究者をベースにして行う作

業であるからである。また、予期しなかった副次的な成果の部分についても同じように直接的な成果と間接的な成果とに分けることができる。

上で、研究開発に携わった担当者及びその共同研究者の成果を直接的な成果であると定義した。次にさらにもう一つの分類基準を導入する。それは、「アウトプット」、「アウトカム」、「インパクト」であり、成果の実質に関わる分類基準である。**アウトプット**は成果を形態的に特定した状態のことであり、**アウトカム**はそれを価値ないし意味の体系に置き換えて測定したときのことであると理解できる。例えば通常の研究開発の場合、典型的なアウトプットは学術論文や特許、あるいはその研究に関する口頭発表の数等であるが、アウトカムは例えばその研究を活かして造られる製品や、改良されたプロセスである。別の例でいうならば、例えば政策研究の場合、今のように政策研究の報告書や論文はアウトプットであるが、アウトカムは、その研究の結果がどのように政策形成に活かされたかに関わり、例えば新しく生まれた政策がアウトカムに相当する。これらに対して次に波及的な効果として、その政策によって社会に何が生み出されてきたか、どのような影響が出てきたか等に関する成果も有り、これが**インパクト**である。製品の場合でいうならば、その製造されることになった製品が、さらにどのような経済効果を引き起こし、あるいは社会的にどのような効果をもたらしたかがインパクトに相当する。なお、インパクトの語は、特に欧州では波及的部分だけではなく、成果全体を意味する場合にもしばしば用いられる。

このように、研究開発の場合アウトプットは直後評価においてもほぼ捕捉できるが、アウトカムはまだ部分的にしか顕在化していないであろうし、インパクトにいたってはほとんどその様子を把握できないであろう。本格的な事後評価は追跡評価でしか行えないことが理解できる。

また、このようなアウトプット、アウトカム、インパクトという分類の他に、「アディショナリティ」や「スピルオーバー」等の概念があるが、これらについては評価法に関連し次章で説明する。なお将来もたらされるであろう成果についても、同様の定義を適用するが、予測される成果の内容は、定義の性格上ほぼ意図的かつ直接的な領域に限られる。

次に、成果と実績について、内容的側面から見直してみる。

第一は科学技術としての価値、2 番目は経済的価値、3 番目はそれ以外の価値で一般的には社会的価値と言われている。言葉を換えて言えば、科学技術内在的価値と科学技術外在的価値にまず分類し、科学技術外在的価値の中で重要なのは経済的な価値であるからその部分を括り出し、それ以外のものを社会的な価値に代表させて社会的

価値と呼んでいる。このように分類すると、全てがいずれかの価値に分類され尽くされることになる。

科学技術の内在的価値、つまり**科学技術的価値**は、どのような科学技術的知識あるいは知が見いだされたかに相当し、知の質に関わる評価を行うことになる。また**経済的価値**は、投資効率 return of investment や経済波及効果 spillover のような、経済的な効果全般にわたる様々な視点での価値の測り方に関わる。**社会的価値**は、一般には problem solving 問題解決がどれだけ行われたかによって測られる。これらの評価の方法をどのように行うべきかについては章を改めて説明する。

最後に、成果系においてもコスト系と同様、その内容が把握されただけでは評価の入り口に立ったにすぎない。成果が数量的ならびに質的に把握された上で、対計画比や対コスト比（費用対効果）等の多様な比較を行う必要がある。

○マネジメント系

マネジメントの評価としては、第一に費用対効果等の実績やパフォーマンスとそこで採用されたマネジメントのあり方との関連においてその適否が議論されなくてはならない。また第 2 には、コストパフォーマンス等が通常直接的ないし短期的な成果の反映であるので、間接的ないし長期的な視点からマネジメントの適性を論ずるとすれば、その研究開発に関与した研究者や関連組織の満足度やパフォーマンス等とマネジメントとの関連について評価されなくてはならない。一般にマネジメントの効果は徐々に顕在化してくるものなので、このような何らかの効果との関連において評価することが不可能な場合が多い。そこで第 3 には、マネジメント上の工夫の有無やその妥当性について、論理整合的に評価する場合もある。第 4 には、マネジメントが多くの場合、特定の対象部分やマネジメント方式に局在的に関係しているわけではなく、評価体系全体に関わる広がりがあるので、全体のパフォーマンスとの相関を評価する必要がある。

いずれにしても、これらの評価においては、マネジメントと実績との因果関係がそれほど単純ではなく、明確にし難い面が多いので、実証的には多くの場合ベンチマーキングや比較の手法をとる。つまり似た課題について、マネジメントの方式が異なる取り組みをしたときに、どのようなパフォーマンスをもたらしているかについて比較分析し、マネジメントの適否に迫る方式である。このようなベンチマーキングや比較評価法は、単純な評価法ではないので、専門のエキスパートにその作業を依存せざるを得ない。

3．3　評価人材

評価を担う人材を分類すると、主には次の3種類になる。第一は、評価のマネジメントを担当する人材であり、2番目には社会経済性を評価する評価パネリストと、社会経済性を分析する専門家、エキスパートである。この両者は同一であることが多い。また3番目には科学技術の質を評価する科学技術ピアレビューアである。**ピア**とは仲間や同業者のことであり、**科学技術ピアレビューア**は科学技術のプロポーザルの内容とその専門性を共有する者であることを意味する。社会経済性分野では、社会経済効果の現れ方を専門とする一般研究者がほとんど存在しないので、社会経済性の質が計れるとする「**社会経済性ピアレビューア**」なる語は、実態的には意味がない空語である。わずかにベンチャーキャピタリストのような実務的専門家がそれに該当する。

評価マネジメント人材は、行政機関の内部や実際に評価を実施する組織の内部において欠かすことが出来ない存在である。また行政機関やファンディング機関さらには研究機関内部の評価支援部署にもそのエキスパートが必須である。

社会経済性の評価は、直感的には行い難く、評価の前段となる調査分析を深めておく必要がある。従って、経済性あるいは社会性に関しての調査や分析を十分に行える**社会経済性エキスパート**が必要になる。このような専門家は行政機関内部には雇用しきれないので、その外部の評価支援機関、例えばシンクタンク等にそのような専門家が集積される必要がある。また、社会経済性評価のエキスパートに関しては、人文社会科学の学者がそれに適しているわけではなく、学問的な専門家ではなく実務的な専門家によってこれは担われるべきである。特に実態を調査し、広く情報を集めてそれらを分析する、そのような実務的専門家が必要である。

これに対して科学技術の質を測る評価者達は、まず科学技術の研究者である必要があるが、それだけではなく、その研究者の中でも高い見識をもち、広い研究分野を見渡せる能力、そしてどこに真の研究フロンティアがあるか等を見分けることが出来る研究者であるべきである。従来の日本の研究評価システムでは、この科学技術ピアに相当する人材として、その見識や能力ではなく単に高いポストに在る人物や著名な研究者を充てる傾向が強く、さらにその上彼らに過度の期待を込めて、社会経済性の評価や、あるいは評価システム全体についてのアドバイスを受けることまでを期待していた。これは明らかにミスキャストでありまた過剰な期待でもある。それぞれの専門家を至急養成する以外にこの窮状を解決する方策は無い。また、これ無くしては正当な評価が根付くこともない。この点に関しても、従来日本の評価組織の中では軽視さ

れてきた。

評価の観点からこの3種類の評価人材を比較したとき、科学技術ピアレビューアは本来研究者であるため、評価パネルを構成したとき、評価のための一般的なガイダンスを受けるにしても評価のための専門的な研修や研鑽を彼らに義務づけることは難しい。むしろ、科学技術的な質の評価に彼らの役割を限定することの方が重要である。それに対して社会経済性評価のエキスパートは、非常に広い社会経済的な分析対象をこなすだけのスキルを身につける必要があり、多くの事例を対象としたケーススタディを重ねながら、そのようなスキルを身につけていく、長い研修・研鑽のプロセスが必要である。そのような専門家は、逆にそれなりの評価がなされなくてはいけない。さらに、評価マネジメント人材は、評価全般についての枠組みを深く理解した上で、評価対象の特質を活かした評価システムを設計し、そしてまたそれを運用していく責任がある。このような専門家を養成することも、また時間のかかる仕事である。そのためには、それなりの研修コースを設け、研鑽を積む必要がある。現在各省庁、あるいは研究機関で展開され始めた研究開発評価に対し、戸惑いと混乱がみられるのは、そのような研鑽を積んでいない、いわば評価の素人が設計した評価システムによって評価が運営されざるを得ない状況であるにもかかわらず、評価が広く義務づけられていることによる。このような矛盾や、評価の稚拙さから一刻も早く抜け出し、評価の質的改善をしていくことが必要である。

評価において最も重要なことは、評価結果に対して、研究者の側も評価の運営を担当するマネジメントの側にとっても、その結果が信頼できるものであることである。特に被評価者である研究者が評価結果を受け入れ、またそれを契機として自らを見直していくことになるような信頼性のある評価が行われることである。またもう一方では、研究者に過剰な評価負担をかけることがないように配慮しなくてはならない。このような多様な課題に対処できる専門家を養成するための研修コースの設定が今急務として我々の前にある。

3．4　評価推進体制のまとめと課題

評価推進体制の諸局面を個別に概観してきたが本章の最後に評価対象別に特色と取り組むべき課題についてまとめておく。より具体的には第6章を参照のこと。

（1）プログラムのもとで展開されるプロジェクト（従属プロジェクト）

- 評価の意思決定の責任は評価担当部署がとる
- プロジェクトの選定作業は、中立性を確保し、政策意図を反映させるため、通常、評価マネジメントは政策当局内部で行う。また、専門性を導入するために評価パネルを設ける
- プログラムのミッションに照らして評価する
- プロジェクトの選択（事前評価）が主要な課題である
- 科学技術外在的な価値に関わる評価に格段の工夫が必要
- 評価システムの設計とパネリストの選定にアドバイスが必要な場合がある

（2）プログラムに基づかないプロジェクト（独立プロジェクト）

- プロジェクトの目的を、それに照らして評価ができるように、独自に具体的に規定する必要がある
- プロジェクト独自の目的の妥当性を判断するための背景調査を深めることが主要課題（我が国では通常行われていないが）
- 本来なら、事前評価を綿密に行うべきだが、従属プロジェクト型に比しコストが格段にかかる
- プログラムの形式の下で構想されても、そのプログラムで目的や内容を明確に規定していない場合のプロジェクトは、この区分に属する（「大型プロジェクト」等）
- 我が国の場合、プロジェクト形成のための組織的な仕掛けや仕組みに特色があり、プロジェクト対象そのものに対する事前評価を本格的に行ってこなかった
- 査定部署や財政当局向けの表面上の目的と、実施上の本来的な目的とがずれていて、二重構造になっていることが多かった

（3）プログラム

- プログラムの事前評価を詳細に行うことは、かなり困難。従って、循環型（例えばGPRA）ないし、途上（中間）評価で見直すことが多い（欧米の場合）
- 事前評価は政策形成と一体的に行われることが多く、プログラム内容を詰めること自体に何らかの評価が組み込まれている。その際、事前調査等の支援が必要となる場合があり、またもし本格的なアセスメント（事前評価）にかけるならそのための支援も必要で、大規模なプログラムの場合欧米で

は複数の機関による支援が行われる（例えばITSプログラム）

- 見直しは、自己評価だけではなく、外部評価を加えるべき（独善に陥らないために）
- 外部評価は、外部パネル、上位機関、独立評価機関、監査／監察機関、議会等のイニシアティブによっておこなわれるべき
- 中間、事後評価は、そのプログラムで実施されたプロジェクト全体の見直しと、プログラムの設定内容の妥当性とを対象とするため、外部の調査分析機関（シンクタンクやコンサルタント等）の支援を通常必要とする
- 成果内容が明確になった時点で本格的な追跡評価を行い、当初に設定されたプログラムの目的を見直し、当該プログラムを包摂している政策までを見直す契機を与えることが最も重要

（4）施策と狭義の政策（制度の複合体）

- 複合化されたプログラムや多数のプログラムが評価の対象となるので事前、事後を問わず一層実施が困難
- 施策や政策の形成過程の様々な局面で、意思決定のための評価支援が必要
- 継続的に検討を深めるための専門の「政策研究」機関が必要（シンクタンクでは対応困難）
- そこでは、日常的に、制度や政策のアセスメント（事前評価）に資する状況分析やオールターナティブの抽出、中間・事後評価のためのインパクト分析、横断的な比較分析とベストプラクティスの検出、行政に対する遺漏分析、等の支援を行う（例えば米国の場合のRANDとSTPI）
- 分析に必要なデータベースを整備し維持することが欠かせない

（5）全体に関わる総合政策

- 広く開かれたアドバイスを吸収するシステム（フォーサイト、サミット・フォーラム、選挙過程のオープンアドバイス、セクター代表のギャランター制度等）が必要
- ラショナルな支援を行うための知的支援機関（部署）や知的データベースの整備が必要
- 「政策研究」機関を中心にした外部機関の機能を活用する

第４章　評価の方法

４．１　評価論

研究開発評価の特殊性に入る前に、評価一般についての論理をまとめておこう。研究開発評価は全て応用問題であると既に述べた。評価対象と評価局面に適合した評価システムを設計し、またそれを運用するためには、評価の基礎力が必要である。

（１）評価の枠組み

○評価と意思決定

「評価」と「意思決定」とは同一ではなく、両者を明確に区別しておくべきである。評価結果は、可能な限り対象の状況を明らかにし、その限りにおいて対象の価値を明確にしたものであるが、このこととは独立に意思決定者には「かくありたい」とか「かくあるべし」という思いがあり、その思いを反映させて意思決定がなされる。従って評価結果は意思決定のための重要な素材ではあるが、意思決定そのものではない。

○外在的評価と内在的評価

評価の枠組みとしては先にも触れたように、「外在的評価」と「内在的評価」がある。**外在的評価**は、評価者が評価対象である被評価者等を外に置いてそれを外部から眺め、評価者の価値観の下で行われる評価である。それに対して**内在的評価**は、評価者も被評価者も同じシステムの中に組み込まれ、価値を共有する関係のもとで評価を行う方式である。意思的な人間（医学や生理学の対象としての人間ではない）や組織は自然そのものとは異なりその内部に普遍的内在原理を備えていない。つまりＡさんで成り立つことがそのままＢさんに適用できるとは限らないし、またある時Ａさんに適用できたとしても別の時点では同じＡさんにも適用できないのが一般的である。このような認識論に立つ場合、外在的評価によって人間や組織を計ることには原理的な危うさがある。日本では、内在的評価が卓越しているところに特色があることを既にのべた。

○明示的評価と暗黙的評価

また、評価は、「明示的評価」と「暗黙的評価」に分けることも出来る。**明示的評価**は、評価者が被評価者に対して評価方式と評価結果を開示し、それを共有して行う評価方式であり、また**暗黙的評価**は評価方式または評価結果あるいはその両者を被評価者に

伝えない評価方式である。日本の多くの組織では、上司が組織内部の構成者を評価する評価のあり方として、暗黙的に評価が実施され、評価結果の中身が開示されないまま、処遇に反映されることが多かった。これはまさに闇討ちであり、少なくとも明示性の確保が必要であることが、80年代から議論されてきた。欧米の場合には、外在的で明示的な評価が一般的であり、その評価結果については評価者と被評価者の両方がサインをして、その上司に結果が報告されるのが通常である。被評価者が評価結果に不服な場合サインを拒否し、評価者の上司に訴える道も用意されていて、この場合評価者がその上司によって評価されることになる。このような裁判制度にも似た修正メカニズムを整備したうえで外在的評価方式を採用するのも一つの道である。しかし、外在的な評価をしながら暗黙的であるのは非常に問題がある。内在的評価の枠組みの中で暗黙的である場合は、共通の目標を持っている仲間内同士で暗黙的であることに相当し、外在的暗黙的評価に比べれば信頼関係が損なわれる程度は小さいといえる。しかし、よりよい方式としては、内在的ではあってもが明示性が確保されている方式である。

（2）評価の原理

○評価の定義

評価 evaluation とは、対象の価値を明確にすることである。そのためには対象の属性や挙動に係る状況を明らかにし、対象の明示性を確保することがまず必要である。評価対象の状況は評価値の集合として表現できる。この評価値の集合に評価要素毎の重要度（これは価値観に相当する）を乗じたものが評価結果としての価値である（図４－１）。評価対象（例えば研究開発機関）は評価単位である評価要素（例えば研究グループ）にまで分解され、この評価要素の集合が「**評価対象系**」を構成する。一方、評価目的をブレークダウンして評価目的に適った評価項目群を導出し、評価項目毎に評価基準を定める。この総体が「**評価体系**」である。

○評価の定式化

人間の認識は、評価する場面を限定すればその限定の度合いに応じてより信頼性のある判断や評価が可能になると考えられる。従って、評価を行うときには漠然と全体を眺めるのではなく、評価対象を要素に分解して要素毎に評価を行い、それらの評価結果をまた再構成して全体の評価に繋げるというやり方を採る。

評価を定式化すると図４－１に示したような式になる。評価対象系と評価体系という２つの軸から成る**評価値マトリックス**は具体的には評価要素群と評価項目群を両軸

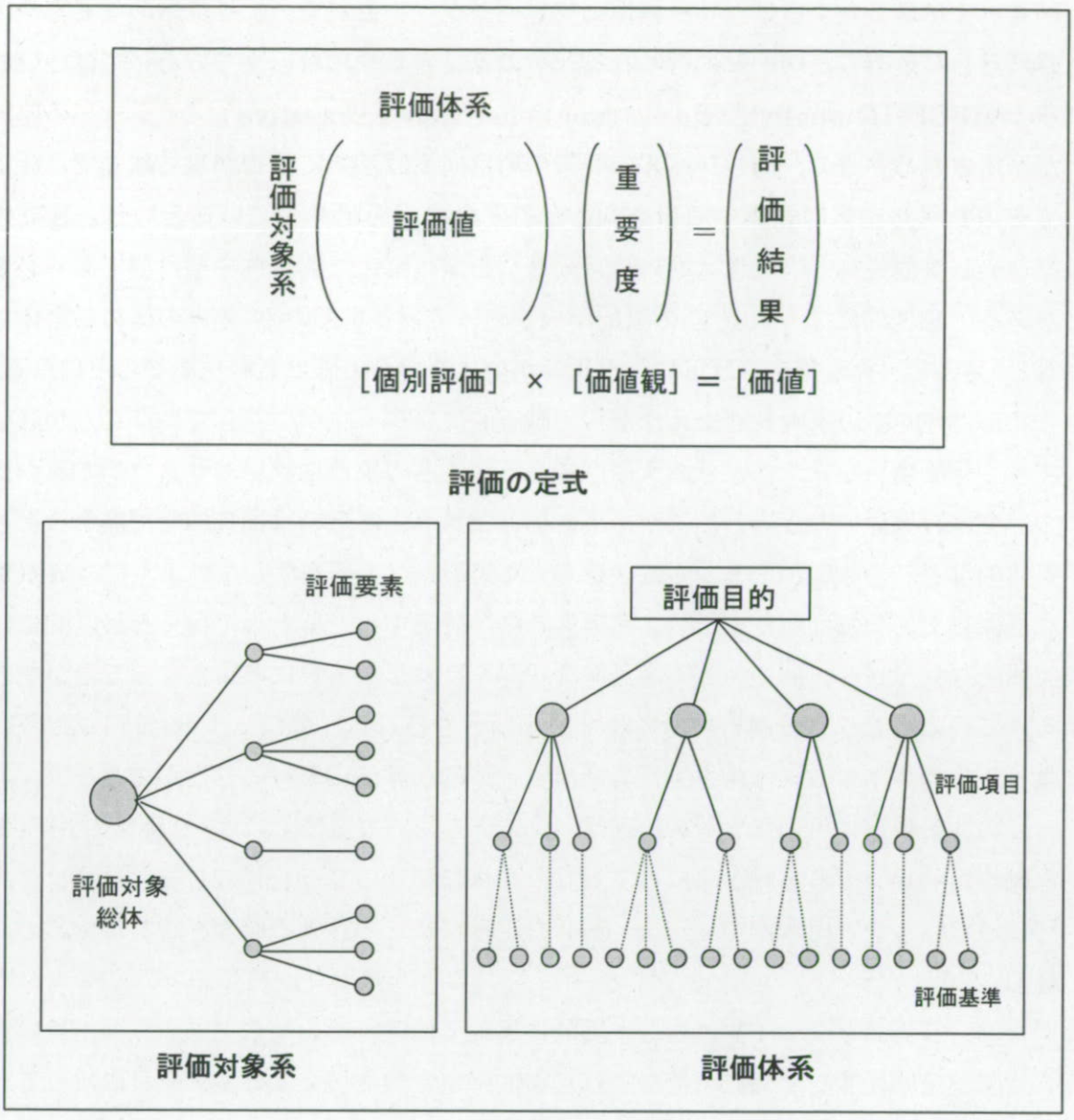

図 4-1 評価の定式化

とする行列である。行列の各要素の値は、評価要素毎及び評価項目毎に定められた評価基準に従って評定された、それぞれの評価値から成る。さらに、評価要素毎に評価項目のどれをどれだけ重視するかに関してのウェイト付けを行う。この各項目毎の重要度もまた行列形式で表現でき、**ウェイトマトリックス**と呼ぶ。従って、評価値のマトリックスにウェイト付けのマトリックスを掛け合わせたものが、評価要素毎に明らかにされた評価結果であり、その全体が評価目的からみた評価対象の価値である。

評価要素や評価項目は通常階層的になっていて、大きな枠組みの要素や項目から、

図４－１に示したようにツリー構造にブレークダウンされて、より具体的な要素や項目に落とし込むことが一般に行われる。その際、要素や項目の立て方としては、いわゆる MECE（Mutually Exclusive and Collectively Exhaustive：ミッシー）の原則が適用されるべきで、同じレベルの要素や項目同士は互いに重複が無く独立で、そしてそのレベル全体の要素や項目を合わせると全局面を尽くしているという、独立性（ないし排他性）と網羅性（ないし完備性）を兼ね備えた要素群や項目群である必要がある。論理的には、MECE の原則が成り立っている場合のみ「要素の総和が全体に等しい」ので、要素毎や項目毎に行う思考が全体を捉える際に意味をもつことになる。

この MECE の原則を踏まえながら、階層的にブレークダウンしていくことが評価要素や評価項目を立てるときの大原則であるが、どのようにブレークダウンするかについては、概念の大きいところから小さいところへ、あるいは抽象的な把握からより具体的なところへ等、この展開軸自身もまた階層的な矛盾を生じないように、MECE な展開軸であることが望ましい。通常何階層か評価項目を落として行くと最終的には実体的な物理量や、あるいは実測可能な状況を反映した事象に収束させることができる。このようにツリー構造に分解した評価項目を各レベルでウェイト付けしながら、より上位のレベルに評価値を上げていって全体的な評価に繋げることができる。

この間の展開に用いた思考の原理は、システム論の表現によれば、2 種類の MECE な軸を用意し、その 2 軸から成るマトリックスないしツリーを構成し、マトリックス評価法ないし決定樹木法 Decision Tree を組み合わせて評価の原理を説明したことになる。

なお、評価値とウェイト値とは本質的に異なる位置付けのものである。評価値はまさに状況を明示するために可能な限り厳密に評定された値であるが、それに対してウェイトの値は、価値観を直接反映したものであり、これは評価におけるポリシーやストラテジーを反映したものと考えるべきである。もちろん評価値の評定においても主観的判断を排除している訳ではないが、この過程で混入する主観性と、その後で反映させるウェイト付けとでは主観性の意味が異なる。

（３）評価手順の基本設計

機関評価を実施する場合を想定し、評価の流れを考えてみよう。

○評価ドキュメント

評価作業に必要なドキュメント（文書資料）は、基本的には次の 4 種類である。「評価対象体系表」、「評価体系総括表」、「評価個票」、「評価結果総括表」。

評価対象体系表（表４－１）は、評価対象を階層的に評価要素にブレークダウンする論理過程を表の形にまとめたものである。通常、横軸は評価対象の階層構造を示す軸であり、縦軸は評価要素に関するMECEな軸である。

表4－1　評価体系総括表の例（機関評価の場合）

機関全体	部等	研究グループ	業務内容
○○研究所	A研究部	a_1 研究グループ	a_{11} 調査業務 a_{12} 分析業務 a_{13} 研究業務
		a_2 研究グループ	a_{21} 調査業務 a_{22} 研究課題
		a_3 研究グループ	b_{31} 研究プロジェクト
	B研究部	b_1 研究グループ	b_{11} 研究課題 b_{12} 研究プロジェクト
		b_2 調査グループ	b_{21} 調査業務 b_{22} 調査プロジェクト
	～	～	～
	Z研究部	z_1 プロジェクトチーム	z_{11} 開発プロジェクト
		z_2 プロジェクトチーム	z_{21} 開発プロジェクト

評価体系総括表（表４－２）は、評価項目と評価基準の関係を示す表であるが、評価目的から評価項目までのブレークダウンの論理過程を併せて示してあり、この階層構造を横軸の左半分に取り、右半分は評価基準の内容とその評定区分構造を示す欄に充てる。縦軸は評価項目の MECE な軸から成る。なお、評価項目によっては複数のサブ項目から構成される場合もあり、それぞれに対応する評価基準の欄が縦軸上に副次的に配置されることがある。

表4－2　評価体系総括表の例（機関評価）

	評価区分				評価項目	評価基準
○○研究所の途上評価	研究開発実績	科学技術的価値	アウトプット成果	量	成果発表回数 HPアクセス回数	成果発表目標回数に比し実績は HPアクセス目標回数に比し実績は
				質	クオリティ成果割合	クォリティ論文誌と国際招待講演数の対目標比実績は
			アウトカム成果	量		
				質		
			インパクト成果			
		経済的価値	アウトカム成果	量		
				質		
			インパクト成果			
		社会的価値・その他	アウトカム成果	量		
				質		
			インパクト成果			
		副次的成果				
	コスト実績	量的状況				
		質的状況				
	費用対効果実績	量的状況				
		質的状況				
	マネジメント実績	個別マネジメント				
		全体マネジメント			戦略計画	戦略計画は組織的に機能したか
					組織計画	組織再編計画の対計画比達成度、実績は
						再編組織従事者の対計画比満足度は

省　略

評価個票（表４－３）は、評価単位である評価要素毎に用意され、評価者が評定結果を記載するために用いるシートである。基本的には評定結果を記載するための表であるが、場合によっては、その便宜のために、当該評価要素の内容や実績等の概要説明、及び当該評価要素の評価に該当する評価項目と評価基準及び対応する評定区分構造を「評価体系総括表」から抜き出し小表の形で配置しておくこともある。

表4－3　評価個票の例（機関の途上評価の場合）

<table>
<tr><td>対象部署</td><td>部／プロジェクト</td><td colspan="2">研究グループ
プロジェクトチーム</td></tr>
<tr><td>対象業務</td><td></td><td colspan="2"></td></tr>
<tr><td colspan="4">［計画内容］</td></tr>
<tr><td colspan="4">［実績内容］</td></tr>
<tr><td colspan="4">［自己評価］</td></tr>
<tr><td>評価項目</td><td>評価基準</td><td>評定結果</td><td>コメント</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

評価結果総括表（表４－４）は、評定結果を取りまとめるための総括表で、総合評価欄の他に評価要素毎に評定結果を記載できるようになっている。多くの場合該当する評定区分を記載する欄だけではなくコメントを付すためのスペースが用意されている。

表4－4 評価結果の総括表の例(機関の途上評価の場合)

対象部署		対象業務	計画内容	実績内容	評価項目	評価基準	評定結果	コメント
総合評価								
A評価部	a_1	a_{11} a_{12} a_{13}						
	a_2	a_{21} a_{22} a_{23}						

○評価項目、評価基準、評定区分とその構造

評価作業を準備するうえで最も困難なことは、「評価項目」と「評価基準」そして「評定区分」とその構造を具体的に定めることであろう。これは評価体系そのものでもあり、評価の理念や枠組みをはじめ評価目的や評価法等を具体的な手続きのかたちとしてまとめたものに相当する。現在我が国で実施されている公的資金による研究評価制度で、上記の3者が完備している制度はほとんどない。特に評価基準に相当する内容が示されていないことが多い。「評価基準」の用語が用いられていても内容的には「評定区分」であったり、「評価項目」と「評価基準」を混同していたりする。多くの場合では、内容的にみて評価項目と評定区分の構造が示されているにすぎない。このような評価体系に従うとすれば、評価者は暗黙的に自己の評価基準を想定し評定を実施することになる。これでは評価結果が開示されたにしても、評価プロセスの論理が明示されていないので、暗黙的評価であると言わざるを得ない。

評価項目とは評価内容を区分しその区分カテゴリー毎に設定された評価局面を特定するための項目のことである。評価内容は、評価目的をブレークダウンして得られ

る。具体的には、評価目的に照らして評価対象の状況を把握するために必要とされる事項を特定する作業に相当する。評価項目は互いに MECE な関係で構成されているべきものである。即ち項目間に重複が無くかつ項目全てを合わせると全評価局面を構成していなくてはならない。評価項目の完全な組の構成例を図４－２に示す。また、代表的な評価目的に適合した評価項目のセットの例を評価基準と共に表４－５に示した。

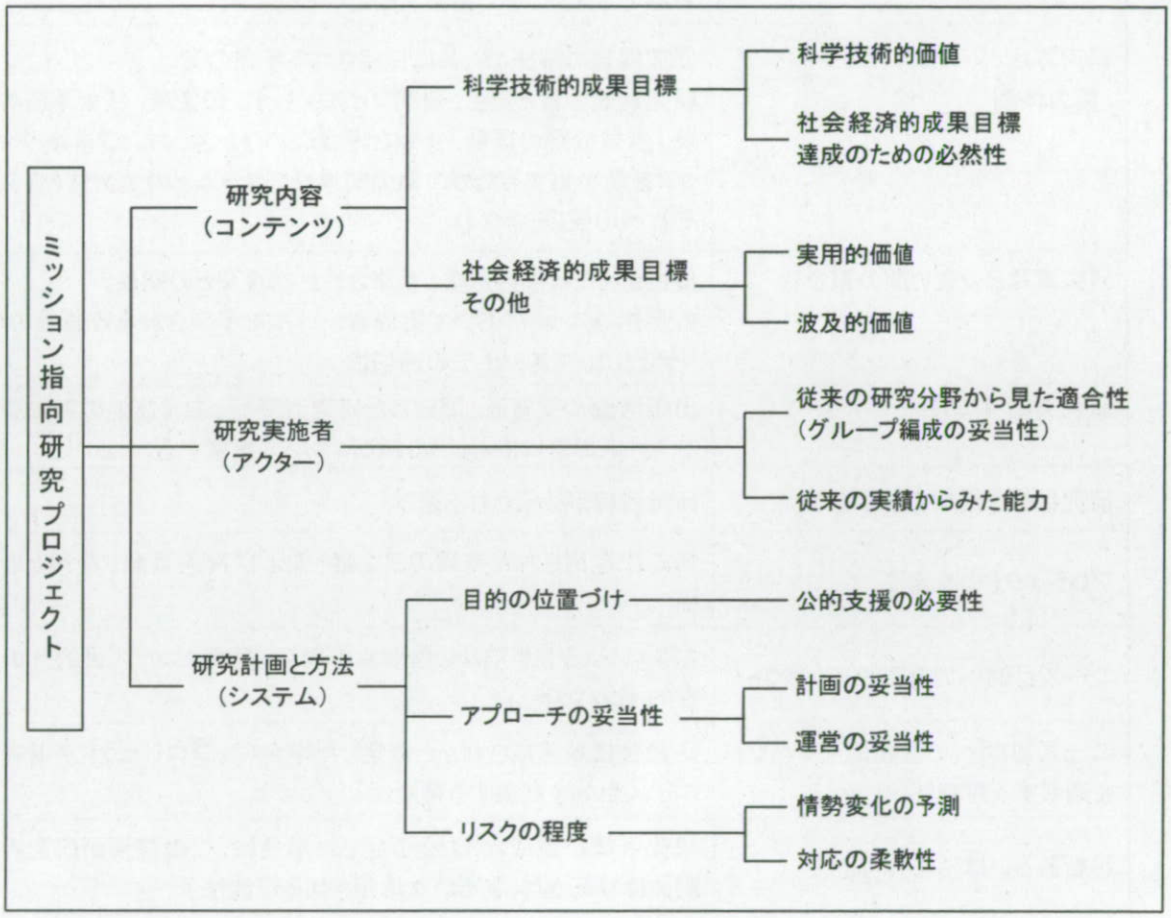

図4－2 ミッション指向研究プロジェクト（従属的プロジェクトの事前評価）の評価項目のMECEな関係

評価基準とは評定の判断を導く根拠のことであり、判断のための論理や指標あるいはその導出法等として設定される。評価項目によっては評価基準が単一で自明のこともあるが、通常は複数の基準が想定されるので、いずれの判断根拠を用いるのか、また複数の基準を採用するのか等を特定しておく必要がある。例えば、基礎科学の研究領域で、「研究の質」に関わる評価項目に対する基準として「独創性」とか「創造性」

表4－5　独立目的指向基礎研究プロジェクトの事前評価用評価項目と評価基準の例

（Ronald N. Kostoff, Research Program Peer Review より）

評価項目	評価基準
研究メリット	提案された課題領域の科学的進歩にとっての重要性。 技術的目標、最高技術水準の潜在的進歩可能性、技術に対する寄与のユニークさに考慮する。
研究方法/ターゲットポイント/協力体制	研究課題の解決のために採用される手順の質。 研究計画の質と焦点、研究マイルストーンの定義、技術革新の程度、当該分野の理解、実験と理論のバランス、および重複やギャップを最小化するための他の関連プログラムとの協力体制（又はそれへの配慮）を含む。
研究資源と目標の間の整合性	提案された科学的目標と要求された総資源との関係。 各実施ユニットにおいて実施者レベルで予定される資源量の「クリティカル・マス」としての適切性
研究実施者の質	出版活動や受賞歴、関連した研究活動歴、および他の無形要素でチームの質に寄与しうるものについて考慮する。
研究目標を達成する可能性	研究目標が達成される確率
プロジェクトの生産性	過去に産出された業績の質と量、および投入資源、費用との関係。
ニーズ目的への潜在的インパクト	本プロジェクトが成功した場合の研究/技術/ニーズ目的への潜在的インパクト
ニーズ目的への潜在的インパクトを達成する可能性	研究目標が達成されたと仮定した場合に、プロジェクトが潜在的ニーズ目的を実現する可能性
移転あるいは活用の可能性	研究目標が達成されたと仮定した場合に、この成果が研究者共同体に移転され、あるいは活用される可能性

が示されていることが多いが、このような概念では基準としての機能を期待できない。「独創性」や「創造性」の尺度があいまいであるからである。推奨されている評価基準は「知的フロンティアへの挑戦」や「知的フロンティアとしてのインパクトの大きさ」等である。

評定区分とは評定結果を表示する際に用いる区分のことであり、通常階層構造や平面分割構造をしている。評価結果を文章で表現するだけの場合では直接的には評定区分は必要でないが、その場合でも評価結果を表現するための論理構造を示すことは暗

黙的であるにしても必要であろう。評定区分はその論理構造を顕在化させたものとも言える。評定区分はその内容を単に記号で表すだけではなく、その記号の意味内容を特定して表示する必要がある。「1」,「2」,「3」等の数字で表現された階層区分であっても、区分構造としては必ずしも等価に区分されているとは限らない。評定区分と区分構造の例を表4－6に示す。

表4－6 評定区分の例

	評定区分の内容	特色
2段階 (1, 0)	・ 合、否 ・ 適切、不適切	採否等を明確に表現するための総合評価等に用いる。
3段階 (A, B, C) (3, 2, 1)	・ 極めて良好、概ね良好、良好でない 極めて適切、概ね適切、適切でない	中間評価や直後評価の対計画比等の表現に用いる。ほぼ計画通りであるなら「概ね良好」、「概ね適切」とする。
(A、B, C) (2, 1, 0)	・ 非常に期待される、期待できる、期待できない	事前評価に用いる場合、合格側を「非常に期待される」と「期待できる」に2区分し、不合格側を「期待できない」とする。従って合格ラインを50点とすると100－75, 74－50, 49－0の3区分となる。
4段階 (A, B, C, D) (4, 3, 2, 1)	・非常に優れている、優れている、劣る、非常に劣る	「平均段階」をあえて設定せず、優劣を積極的につける場合に用いる。
	・ 極めて優れている、かなり優れている、優れている、劣る	「優」「良」「可」「不可」のように、合格側(100－50)を 3 段階に区分し、不合格側(49以下)は1段階で表現する。
(AA, A, B, C)	・ 極めて良好、概ね良好、良好、良好でない	3 段階評価に特記部分(AA)を加える。「優れている」側のグレードに飛び抜けた開きがある場合。
5段階 (A, B, C, D, E) (5, 4, 3, 2, 1)	・ 非常に優れている、優れている、良好、劣る、非常に劣る ・ HIGH、GOOD、AVERAGE、FAIR、LOW	中間値付近を「良好」として等価に区分する。
(AA、A、B、C、D)	・ 極めて優れている、かなり優れている、優れている、劣る、かなり劣る	4段階評価＋特記部分(AA)の例。
6段階評価以上 (10段階評価の例)	(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	上記をさらに細分化する。 例えば10段階では、10－8.5(HIGH)、8.4－6.5(GOOD)、6.4－4.5(AVERAGE)、4.4－2.5(FAIR)、2.4—1(LOW)等の区分に相当するものと考える。

（4）評価結果の利用

評価はマネジメントの一角を担う「手段」であり、それ自体が「目的」ではない。評価結果は何らかの選択や処遇或いは見直し等に適切に反映され、研究開発や政策目的の推進に対して実効的に寄与すべきことが期待されている。評価を「評価のための評価」にしないためにも、また逆に評価を形骸化させないためにも、評価結果の有効な利用が図られなくてはならない。そのためには、評価結果がそれに続く意思決定に適切に活かされ、そして意思決定の結果が評価者を含む関係者に開示されることが重要である。

4．2　評価手法の体系

評価を実際に行うためには、評価対象に関する情報やデータをまず収集し、そしてそれらを分析して、評価手法に乗せることが出来るように処理・加工する必要がある。この意味で**調査法**、**分析法**、**評価法**の３者は互いに連関しているが、一方でそれぞれの機能を認識し、方法論としては３者を峻別する必要がある。このように「評価法」の用語を限定的に捉えることにより、評価法の理解に対する混乱を回避できる。評価の方法論に関する多くの解説書では、この３者を区分することなくこれら全てを「評価法」として紹介し、その結果として比較評価法やシステム評価法のような本来「評価法」として重要なカテゴリーを特定せずに、メトリックスの枠組みの中で紹介していたりするケースが多くみられる。この意味からも次に紹介する表４－７の構造を十分に理解していただきたい。

表４－７に、評価手法の体系とその前段として必要な調査法と分析法を一体的にまとめた。なお、評価手法としては、「単純評価」と「複合評価」の２種類に区分した。**単純評価**とはおよそ我が国で現在行われている程度の評価を意味し、より厳密に言えば研究開発に伴って得られる情報やデータのみに依拠するなど、評価を目的として構造化された調査や分析を行うことなく、限定された情報やデータの下で、特別な専門性なく運用できる手法を用いてなされる評価を意味する。これに対し**複合評価**とは、単純評価で用いられる調査法や分析法の他に、評価の専門性をもって評価のために特別に収集された情報やその分析に基づいて行われる評価を意味する。そのための手法は表４－７にあるように多岐にわたっていて、複合評価を適切に行うためには、評価エキスパートの支援が必要となる。彼らに要請される能力は逆の言い方をすれば、複合評価で使いこなすべき調査法や分析法に精通し、また評価手法を運用できることである。

表4－7　調査分析評価手法の体系

		調査法		分析法	手法区分	評価手法
		情報収集	データ収集			
調査分析評価手法	単純評価	インタビュー（関係者） ヒアリング（専門家）		ケース分析	定性的評価	ピアレビュー法 パネル法
			既存統計 実在データ		半定量的評価	評点法 比率評価法
	複合評価	アンケート調査 事例調査 インタビュー（関係者） ヒアリング（専門家） 顧客調査 社会調査		文献分析 ケース分析 社会分析 コンテント分析 計量文献分析（ビブリオメトリックス） 引用分析，共引用分析 共語分析，謝辞分析 計量技術分析（テクノメトリックス） 特許分析， サイエンスリンケージ分析 技術仕様分析 計量経済分析（エコノメトリックス） 費用指数分析 費用便益分析 生産関数分析 財務指標分析 消費者余剰分析 計量社会分析（ソシオメトリックス） CVM，汚染指数分析 健康・疾病指数分析 雇用指数分析 統計分析 相関分析，因子分析 判別分析，クラスター分析 構造化分析 マトリックス法，階層化法 関連樹木法，行列分析， マッピング システム分析 モデリング，プログラミング シミュレーション，ネットワーク分析	定性的評価	ピアレビュー法 パネル法
					半定量的評価	評点法 比率評価法
			アンケート調査 事例調査 既存統計 実在データ 申請・管理データ データマイニング その他のメジャメント		定量的評価	比較評価法 対比較年度比，対計画比 コントロールグループアプローチ （対照群・比較群アプローチ） 事前事後比較 非実行仮説 ランキング，ベンチマーク ポートフォリオ 指標法 単一指標法，複合指標法
					総合的評価	システム評価法 シナリオプランニング，ロードマップ ディシジョンマネジメント評価 ロジック評価法 ファクツ評価法，ケース評価法 レトロスコピック評価法 （歴史的遡及） レビュー法 デルファイ法，フォーサイト フォーカスグループ

（1）調査法と分析法

調査法は、対象の質的側面に関わる**情報収集**と、数量に関わる**データ収集**とに分けることが出来る。前者を定性的調査、後者を定量的調査ということもあるが、定性的調査の内でも特に質的側面に関わる調査に重点をおくべき意味合いを込めて、ここでは「質」と「量」に区分してある。なお、厳密には質を数量化して表現することも出来るし、また逆に強度因子のように質的側面を表す数量もある。数量による表現は明快ではあるが一面的でもあり、質に関わる定性的な表現に及ばないことが多いことに注意すべきである。また、「質的＝主観的」、「量的＝客観的」という誤った方程式はこの際完全に払拭しておくべきである。

分析法においても質的分析と量的分析があるが、構造化分析やシステム分析のように数量を媒介にした分析手法であってもその結果を図形やモデルで表現するものもあり、どの局面で定義するかによって分類が異なる。ここでも質的側面に迫る分析の重要性について強調しておきたい。それは多くの場合定性的分析であるが、一方で可能な限り定量的な分析を進め、評価対象の状況を明晰にする努力も欠かせない。

調査法と分析法において（また後で述べるように評価法においても）定性的アプローチと定量的アプローチの関係は、以上のように質的側面に迫るためには、定性的アプローチの方が重要であり定量的アプローチは状況の明晰性を高めるために補完的に用いられる。しかし、表４－７でみられるように、手法としては定量的手法の方がはるかに多く定式化されていて、また我が国においては専門性を必要とする定量的手法の理解が進んでいない。従って現在の我が国にとっては、定性的手法の質的向上と共に、圧倒的に遅れている定量的手法の理解を進めることの緊急性も極めて高いと言える。

○メジャメントとメトリックス

調査法と分析法における定量的アプローチを総称して「メジャメント」と「メトリックス」と言う。**メジャメント**は調査を目的としたデータ収集に係る方法の総称であるが、既に数量化されている**実在データ**や**既存統計**に関してのデータ収集法だけではなく、むしろ定性的側面に対して新たに測定基準を立てて、その基準に照らして測定を行うときに主としてこの用語を充てる。**メトリックス**は分析を目的とし、既に測定されている統計量や実在している数量等を加工して、対象とする評価項目の内容を表すように仕立てて分析を行ったり、逆に対象とする評価項目の内容を表現できる関係を特定しその関係を構成する変数を測定して分析データを得る場合等のことを言う。

測定対象に適用できる数量的尺度が無い場合、メジャメントの方法として次のよう

な工夫がある。適切な評定区分を設定して**評点付け**ないし**評点化**を行い、数量表示への転換を図る。また、一対比較法等により評価対象要素間の**順位付け**を行い、序列番号を用いて評価要素特性の数量化を図る方法もある。

何らかの分析を目的として工夫した測定を行う場合や、既存の統計量を分析のために再処理する場合等は、測定そのものを目的にしているのではないので、メトリックスと言うということを既に述べた。例えば論文数と引用数等の統計的な量を使って、それらの量の多寡だけではなく、分野毎の引用度（１論文当たりの引用数）を求め、その値でノーマライズした研究活動度のランキングや評点化を行う等である。

典型的なメトリックスの例としては、文献情報を対象にした**ビブリオメトリックス**、技術や特許情報を対象にした**テクノメトリックス**、経済情報を対象にした**エコノメトリックス**、そして社会情報を対象にした**ソシオメトリックス**がある。また対象に特化しない分析手法として**コンテントアナリシス**（内容分析）や**統計分析**の定型的な方法も有用である。

さらに、**構造化分析**や**システム分析**のように、数量による分析を媒介にするが、分析結果が数量以外の形式で表現される場合もある。このような手法は通常メトリックスに含めない。

○定性的調査分析法

定量的なメジャメントとメトリックスによる調査分析法の他に、定性的な手法を交えてオリジナルなデータを集め、その分析を深めるとこともよく行われている。オリジナルなデータは、当然のことながら**インタビュー**（関係者への訪問調査）や**ヒアリング**（専門家への訪問調査）によって得るとか、あるいは**質問紙**（**アンケート**）によってデータを集める（**サーベイ**）とか、あるいは**ケース分析**や**文献分析**により、全体像を掴むための様々な分析を行うとか、これらの方法を通して質的に深いオリジナルデータの集積とその実態に即した分析が行われている。

（２）評価手法

評価手法は**定性的評価**と**定量的評価**にまず分けることができるが、表４－７に示したように、その中間的な**半定量的評価**や、両者が複合化された**総合的評価**を加えて分類すべきであろう。ここで、評価の質的側面を担う評価法の重要性を再び強調しておきたい。

定量的評価＝客観的評価と考えることは、明らかに間違いであり、例えば主観的な価値観に基づいて数量化された定量的結果は、何ら客観的な評価には繋がらない。同

じことは定性的評価についても言える。両者の違いはむしろ次の点にある。数量による表現（定量的表現）は言語による表現（定性的表現）より、一般に明晰性が高く、情報伝達に際しての任意性が低いにすぎない。一方、定性的評価＝主観的評価と捉えることも間違いで、定性的な言語による表現の中で本質を見定めることは十分可能であり、実態を深く捉える際には数量ではなくむしろ論理化された概念による方が適切であることの方が多い。

欧米においても 70 年代前半までは、いわゆる定量的評価法が詳細に展開されていくプロセスにあったが、その限界を認識した後 80 年代をかけて定性的な質的評価が見直され、最近の GPRA にみられるように、定性的な枠組み、あるいは学習的な枠組みのような、非常に柔軟なアプローチを重視するように変わってきている。

研究開発評価において最も重要な評価手法は**ピアレビュー法**である。科学技術の質的に深い側面を捉えるためには、当該科学技術分野の専門家の知見による他に方法は無い。ただし、ピアレビューパネルの運営は極めてデリケートであり、適切なピアレビューアを得て細心の注意の下で運営されなくては、その効果を発揮出来ないばかりか深刻な弊害をもたらす。質の高いピアレビュー法の有り方に関しては、次章で詳しく述べる。

パネル法は、科学技術的ピア（同僚）のみで構成されるピアレビューパネルと異なり、科学技術の受け手の側（産業界や市民）やその他の関係者（資源配分機関、政策担当者、納税者等）を含む評価パネルで評価を行う方式である。通常別の仕組みで用意された調査分析データや情報に基づきパネルで検討が加えられる。ピアレビューが科学技術の質の評価に主要な任務が置かれ、ほぼ科学技術情報のみが提供されているのに対して、パネル法では社会経済的側面や科学技術の階層化された高次情報等も用意される。

評点法は、評価のための調査分析情報を評点化して評価する方法であり、また**比率評価法**は、評価対象の属性の比をとってその挙動や状況に基づいて評価する方法である。いずれの方法も半定量的評価法と言うべきで、評点法では質的な評価結果を評点化して階層的な番号や記号で表現するにすぎない。しかしこうすることにより評価結果を明晰に際立たせて表現できる。比率評価法は、インプット・アウトプット比のような評価対象の属性の比を取って評価する方法の一般名称である。比を取る 2 種類の変数が数量で表示されていれば定量的な意味を持つが、評点化されたデータのような場合では定量的とは言い難い。なお、単なる比率は意味のある指数にならないことが多いので、次に述べる比較評価法と区別する。

比較評価法は評価にとって意味のある比較局面を設定しその比較を行う方法であり、比較的簡便な評価法として多くの局面で利用されている。評価において、「絶対評価法」は評価値を定め難いのが一般で、多用することが困難であるのに対して、「比較評価法」は相対評価に相当し、意味のある比較対照を設定することにより容易に評価値を得ることができる。例えば、パフォーマンスそのものの絶対評価が困難であっても、対前年比のような**比較年度比**、あるいは**対計画比**や**対目的比**のように、パフォーマンスを相対化することにより評価の手がかりを容易に掴むことが出来る。このような方法は機関評価における基本的な評価法となっている。多少手の込んだ比較をするとすれば、例えば対チャンピオンデータ比のような**コントロール・グループ・アプローチ**（対照群・比較群アプローチ）や**事前事後比較**あるいは**非実行仮説**（計画を実施しなかったとした場合との比較）のような異なる状況との比較によりネットの効果を読み取り評価する場合がある。さらには、多数の比較対照を準備して評価を行う**ベンチマーク**、**ランキング**、**ポートフォリオ**等の評価方式があり、いずれの評価法も重要である。

メトリックスの結果を利用して直接的に評価を行うとすれば、通常**指標法**と呼ばれる典型的な定量的評価法になる。一つの指標によりある局面を評価する**単一指標法**や複数の指標を組み合わせて複合指標を構成し総合的な評価をめざす**総合指標法**等がある。単一指標法は限られた局面について評価しているにすぎなく、また総合指標法においてもデータの制約から全体像を描くことはほとんど不可能である。この意味で指標法によって表現されている多くの評価結果に対する解釈には注意を要する。

システム評価法、ロジック評価法、レビュー法は、いずれも定性的評価法や定量的評価法を組み合わせて評価を行う方式であるので、**総合評価法**として区分する。

システム評価法は、評価局面を表現するためのシステムモデルを構成しその要素となる変数やパラメータを測定したり推定したりして評価結果を得ようとする方式の総称である。**ロードマップ**や**シナリオプランニング**のような計画管理手法や、**ディシジョンマネジメント法**のように投資案件の意思決定支援手法のような局面で効果的に用いられている。かつてのOR的評価法とは基本的に異なり、システムズアプローチの手法により適切な評価局面の構成に主題を置いている。**ロジック評価法**は、**ファクツ評価**、**ケース評価**、**レトロスコピック評価**（歴史的遡及）のように、評価対象を実態的に深く分析しその過程で得られる知見を論理的に評価しようとする方式である。システム評価法は定量的色彩が濃く事前評価向きであるのに対して、ロジック評価法は定性的手法による分析を多用し質的な側面に焦点を絞り事後評価により質的

に深い教訓を得ようとする点に特色がある。また、**レビュー法**は、将来展望等を評価することを目指し主体参加型で行う総合的評価法である。**フォーサイト**、**フォーカスグループ**、**デルファイ法**等がその例である。

（3）評価結果の表現

評価結果は評価値で表現されるだけではなく、その評価値の位置がどこにありまたその位置がどのような意味を持っているか等の評価結果の表現方法が重要であり、この表現方法が高次の評価法そのものでもある。

ある多くの個別のプロジェクトが、ある一つのプログラムのもとで展開された場合には、そのプログラムの内部で展開されたプロジェクトの個々の評価値を、ある平面の中に位置付け、その分布の状況等を判断することによって、プログラムの様子が明確になってくる。このように、評価結果をどのように位置付けて表現するかは、非常に重要であるが、それをグラフや図形で表すことの他に、序列をつけるランキングや、ある平面の中に位置付けるポートフォリオ、あるいは方式とパフォーマンスとの定性的ないし半定量的な相互比較をするベンチマーク等、様々な評価結果の表現形式がある。また、マトリックス法や関連樹木法のような構造化技法の多くは評価法であると同時に評価結果の表現法でもある。

5章　研究開発評価の方法論

この章では、第4章で概観した評価法の中から、現在の我が国において研究開発評価の向上を図るうえで重要な方法論に焦点を絞り、具体的に説明する。個別方法論の解説に移る前に、各評価局面でそれぞれどのような手法が有効であるかについて概観しておく。

5．1　方法論の適用局面

方法論の適用局面を特徴づける切り口として3種類の軸を考える。第一は、評価対象の軸であり、プロジェクト、プログラム、施策や政策、あるいは機関等。2番目は評価フェーズの軸であり、事前、途上、直後、あるいは追跡評価である。そして3番目は、評価内容に関わる軸であり、科学技術性、経済性、そして社会性である。ここでは表4－7に示した典型的な方法論についてその適用区分を2種類の表にまとめる。

図5－1は、評価対象と評価フェーズを両軸に取り、マトリックスのセル毎に適用可能な方法論を示したものである。ピアレビューを含むパネル法はいずれの場合においても有効であり、比較評価法とロジック評価法は途上、直後、追跡評価において調査分析（メジャメントとメトリックス）を前提にした場合有効に適用できる。図5－1で方法論に付した下線は、その方法論の重要さの度合いを示したものであり、従属プロジェクトの事前・途上・直後評価を除き、調査と分析が極めて重要であることを意味している。また事前評価においては、本格的に行うとすればレビュー法やシステム評価法等を動員することになるが、従属プロジェクトのような比較的簡便な評価で良い場合にはピアレビューを含むパネル法で対処することで十分であろう。ただしその適用に当たっては「適切な方法」によるべきで、ピアレビューの質を高める努力を欠かすことは出来ない。

図5－2は、評価内容別に有効な方法論をまとめたもので、ピアレビュー法は科学技術性の評価には有効であるが、社会経済性に対しては有効でない。この場合は本格的な調査分析を踏まえた評価法を適用しその結果を参考にしてパネル法で詰めることになる。

	事前評価	途上・直後評価	追跡評価
従属プロジェクト評価	ピアレビュー法 パネル法	パネル法 ピアレビュー法 比較評価法 メジャメントとメトリックス	パネル法 比較評価法 指標法 ロジック評価法
プログラム評価 政策評価	パネル法 ピアレビュー法 レビュー法 システム評価法 メジャメントとメトリックス	パネル法 比率評価法 比較評価法 指標法 ロジック評価法 メジャメントとメトリックス	メジャメントとメトリックス
独立プロジェクト評価	パネル法 ピアレビュー法 システム評価法 レビュー法 メジャメントとメトリックス		パネル法 比較評価法 ロジック評価法 メジャメントとメトリックス
機関評価		パネル法 比較評価法 ロジック評価法 メジャメントとメトリックス	

図5－1　方法論の適用可能性

評価内容	有効な方法論	
	事前評価	途上・直後・追跡評価
科学技術性	ピアレビュー法	パネル法 比較評価法 ロジック評価法
社会経済性	パネル法 レビュー法 システム評価法 メジャメントとメトリックス	メジャメントとメトリックス

図5－2　評価内容からみた方法論の有効性

5．2　調査法

（1）基礎的な調査方法

研究プロジェクトや機関の評価を行うためには、まず第一に、それら対象の多様な側面についての定性的及び定量的な情報を集めることが必要である。そのうち、研究内容や研究結果そのもの、ならびに投入資源などの情報については研究実施報告書や自己評価報告書に研究者自身によって記されることが多い。一方で、研究成果のユーザー側からの評判や、他の評価対象との比較や当該研究分野全体の中でのポジション、あるいはプロジェクトに参加した研究者の意識や考えの全体としての傾向などは評価される研究者自身によっては収集されにくい情報であり、評価のマネジメントサイドやその支援機関のエキスパートによって収集分析すべき内容である。また、大きなプロジェクトの追跡評価などでは、プロジェクト終了後から追跡評価時点までの情報を新たに収集する必要がある。このような情報の収集方法のうち、どのような対象にも共通的な基礎的な方法をまず説明する。

○インタビューとヒアリング

インタビューとヒアリングはほとんど同じ意味に用いられているが、敢えて区別するとすれば、関係者に対する調査は**インタビュー**であるが、専門家からの意見聴取は**ヒアリング**の方が語感としては相応しい。しかし調査手法としては同一であるので、

以下では**インタビュー調査**なる用語を用い統合的に説明する。

インタビュー調査は、特定の人物に評価対象についての主観的な判断を聞くときや、その人物に関わる情報や経験した事項を提供してもらう時に用いる手法である。中でも調査すべき対象の人数が少なく、それぞれから深く情報を聞く必要がある場合に有効に用いられる。インタビュー調査が行われる場合の例としては次のようなものがある。

- 研究結果のユーザーにその判断を聞き、その社会経済的効果を把握する場合（ユーザーが少数に特定される場合）
- 機関評価において機関を構成する研究者にマネジメント上の問題点を聞く場合
- 特に追跡評価において、プロジェクトに参加した研究者それぞれに対してプロジェクトに参加した事によりいかなる変化が研究内容や研究方法・結果に生じたかを聞き、公的資金によるプロジェクトの効果を把握する場合

インタビューでは、質問項目をはじめから設定して統一（構造化）させる方法と、対象者に全く自由に話してもらう方法（非構造化）がある。インタビュー調査で気をつけなければならないことは、インタビュー対象が「自分が評価されている」と感じる場合には不利になる情報が得られなくなることである。そのため、被評価対象はインタビュー対象者自体ではなく、あくまでもプログラムや機関が対象であり、インタビュー対象者は関連情報を提供することを求められているだけというスタンスを伝えることが必要な場合もある。同時に、プロジェクトやプログラムに参加することが研究者自身に与えた効果を調査するインタビューにおいては、公的投資による効果を過度に強調する傾向があることが指摘されている。これは公的資金制度を維持させようとするためである。そのため、インタビュー対象者にはプロジェクト・マネージャーや各研究者、研究に関連しない部署の人間など様々な種類の人物を設定する必要がある。

○サーベイ

質問紙による調査、つまりアンケート調査のことを**サーベイ**と言う．サーベイはインタビュー調査と同様に主観的判断や個別の人物・組織に関する情報を収集する場合に用いるが、インタビュー調査とは異なり、比較的多数から情報を得る必要がある場合に用いられる。例えば次のような場合がある。

- プロジェクトに参加した企業だけではなく参加していない企業に対しても研究内容や成果を調査し、プロジェクト参加の効果を同定する場合
- プログラム内の全プロジェクトの参加研究者からの情報や、企業を対象として

プログラムの効果を同定する場合。

サーベイ調査でも、選択肢などが厳密に設定してあるものから、自由に主観的判断を記述してもらうものまである。また評価対象も評価目的に応じて、全数調査を行うのか、サンプリング調査を行うかなどを考慮する必要がある。

○事例調査

事例調査は、調査対象の状況や範囲に関して不透明な事項が多い場合に対して行う調査方法で、例えば幾つかの事例を取り上げて試験的な調査を行い調査の視点や調査項目をより明確にすることを目的にする場合や、事例があまり無い場合などでは試験的な調査ではなく本格的な調査分析対象として事例に取り組み詳細な情報までを収集することもある。この後者の場合では事例調査と言うより**事例研究（ケーススタディ）**の用語の方が相応しい。

例えば前者では、短期的には効果が見えにくい産学連携プログラムにおいて、実際にどのような種類の情報交流や連携がどのような相手とともに行われたかを明らかにする場合が挙げられる（例えば米国 Engineering Research Center の評価）。他方、後者では、古典的には米国 DOD（1969）による Hindsight や、NSF の資金提供によりバッテル記念研究所（1973）が行った TRACES（Technology in Retrospect and Critical Events in Science）のように、特定の成功したイノベーションに対して、公的資金提供によるプロジェクトやプログラムがどのような役割を果たしたかを歴史的に跡づける場合に行われた。このような方法によって、基礎研究の長期的なインパクトを評価することが可能となるが、当然、研究終了から評価を行うまでのタイムラグはかなり長くなる。また、このようにして得られる詳細な叙述もその解釈には主観が入るため、実際にはケーススタディにおいても定量的な指標を併用する事が必要となる。

（2）測定法

評価においては評価対象の特定の側面に関する情報を定量的な形で収集することも日常的に行われている。このような作業によって、順位付けを行うことが可能になったり、多様な側面の定量的データの収集を通じて総合的に判断することが可能になるなど、評価のための操作可能性が向上する。このようにある側面の定量的な情報を集めることが**測定**（measurement）である。ただし、測定の対象となるのは研究資金や研究者数のようにそれ自体が客観的な数値として存在しているものだけではない。評価者の主観的な判断も点数の形で表すことにより定量化（評点化）することが可能

であるし、「自然環境の保護」というような社会的価値も金銭という統一的な形で定量化する方法もある。そのため、測定で測られる内容は主観的なものから客観的なものまでスペクトル上に存在することになる。以下ではその概要を記す。

○評点付けと順位付け

調査者の主観的な判断を点数の形で数量化する方法が**評点化法**である。

測定値を得る第一段階としてのメジャメントを考えるとき、測定対象に適用できる絶対的な尺度が存在しない場合もある。その場合、プリミティブであっても定量化の手法について先ず考える必要がある。それは**評点化**ないし**評点付け**であり、測定基準を適切な評定区分に分割し、それとの比較を行うことから始める。評定区分の順番に数量を当てはめて評点化を行う。評定区分の構造が「程度」や「段階」を表す階層的なものであり、MECE な関係に区分されていることが望ましい。定性的に認識されている評価内容を数量的に表現する場合にこのような数量化手法が用いられる。

測定基準に相当する評価軸を敢えて区分せず、複数の評価対象を一対ずつ取り上げ、評価軸に照らして程度の大小関係を順次測り、評価対象全てを序列化することもできる。この方法は最も手堅い比較の手法であり、**一対比較法**と呼ばれる。つまり、評価しようとする要素の任意の 2 つを取り上げ、測定基準に照らして大小関係を比べる。これを全ての要素の組み合わせについて行うと、結局全体を順位付けることができる。この手法を**順位付け**と呼ぶ。そしてこの順位の番号を対象要素の数量として用いる。

○実証データ・統計的データ

上述のものは調査者による主観的判断を数量化するものであるが、評価対象そのものについての数量的な情報が実証的に得られる場合もある。例えばインプットとしては資金や研究者ないし研究支援者がどの程度投入されたか、プロセスとしてはプロジェクト参加企業と大学との間の交流頻度がどの程度であったか、アウトプットとしてはプロジェクトによる論文や特許等の数、育成された人材の数や、（例えば環境対策技術分野では）環境汚染物質の低減度合いなど様々なものがある。これらデータは、評価者が評価を行う際の参照情報として用いられるものであり、特に重要なものについては何らかの処理（例えばデータの値を 1 点から 5 点迄にノーマライズする）を行った後に総合点の算出に用いられる。実証データは対象のインプット・スループット（パフォーマンスとポテンシャル）・アウトプット・アウトカム・インパクトといった各側面について必要なものを選択し調査する必要がある。

同様に統計データは、上記のように対象について得られたデータを一般的な平均と比較する際などに用いる目的で既存のデータベース等から引用される。例えば同規模

で同一分野の機関との比較のための統計データなどである。

このような統計量を使って、それを指標として利用し定量的な評価に結びつけることは、一見合理的のように見えるが、一方で非常に危険性を伴っている。例えば、その統計量が実体の或る特定の側面しか表していない場合や、また良くあるケースではその統計量が実際に統計的な集計をし易い部分についてのみ行われていて、測れる指標で測ったに過ぎないようなメトリックスが横行することである。本来必要なことは、測るべき指標を独自に統計的に集計し、それを代表値として評価していくべきであるが、そのためには大変大きなコストがかかる。また、すべての側面を数量化して測定し尽くすことはほとんど不可能である。そこで、測れる指標でそれを代替し、（これだけでもかなり大変であったりするので）置き換えた部分が全体を表しているかのような錯覚に陥ったりすることになる。このような誤謬を避けるためには、指標そのものはあくまでも参考値であるとして、限定付きで利用するべきである。

実はこの背景には人文社会科学の研究における、数量的アプローチと質的アプローチの根深い対立がある。しかし、この両者は単に補完的関係にあると捉えるべきではなく、現実的な制約のなかで実態の本質に迫るためにはこの両者をどのように使い分けるべきかという視点から整理すべきであろう。

以下に数量的アプローチの中心にあるメトリックスの手法について順次述べる。

5．3　分析法

（1）ビブリオメトリックス

ビブリオメトリックス（計量文献分析手法）は、研究開発のアウトプットとしての論文や特許を用いた分析を行う方法である。

○論文分析

論文に関しては、単純な方法としては、次の値の計測による研究開発の生産性の分析が行われる。

- 論文数
- 引用数

例えば機関ごとのこれら生産性の分析や、プロジェクトに参加する前後での生産性の変化の分析がなされる。ただし、引用数の分析においては、主に用いられるデータベースであるISI社のScience Citation Index自体に様々な問題が存在する。すなわ

ち、分野によって収録雑誌数に偏りがあることや、言語の偏りがあるなどである。そのため、これらを十分注意して分析を行う必要がある。

また、論文数や引用数は分野によってその平均値が大きく異なるため、分野間で単純に数を比較しても全く意味はない。その際には、当該分野の全雑誌の全論文を基にした平均値で標準化した数値を用いるなどの加工を行う必要がある。図５－３はある大学に対して、各分野ごとの論文数がその大学全体の論文数に占める割合を日本の全大学の平均値と比較したものを横軸に、その大学の論文の被引用数を当該分野の全論文の平均被引用数と比較したものを縦軸にとったものであり、これにより当該大学においてはどの分野に強みがあるかを比較することができる。なお、図における円の大きさは論文数を表している。さらに詳細な分析を行うためには、引用数についても自己引用の除去などの処理を行う必要がある。

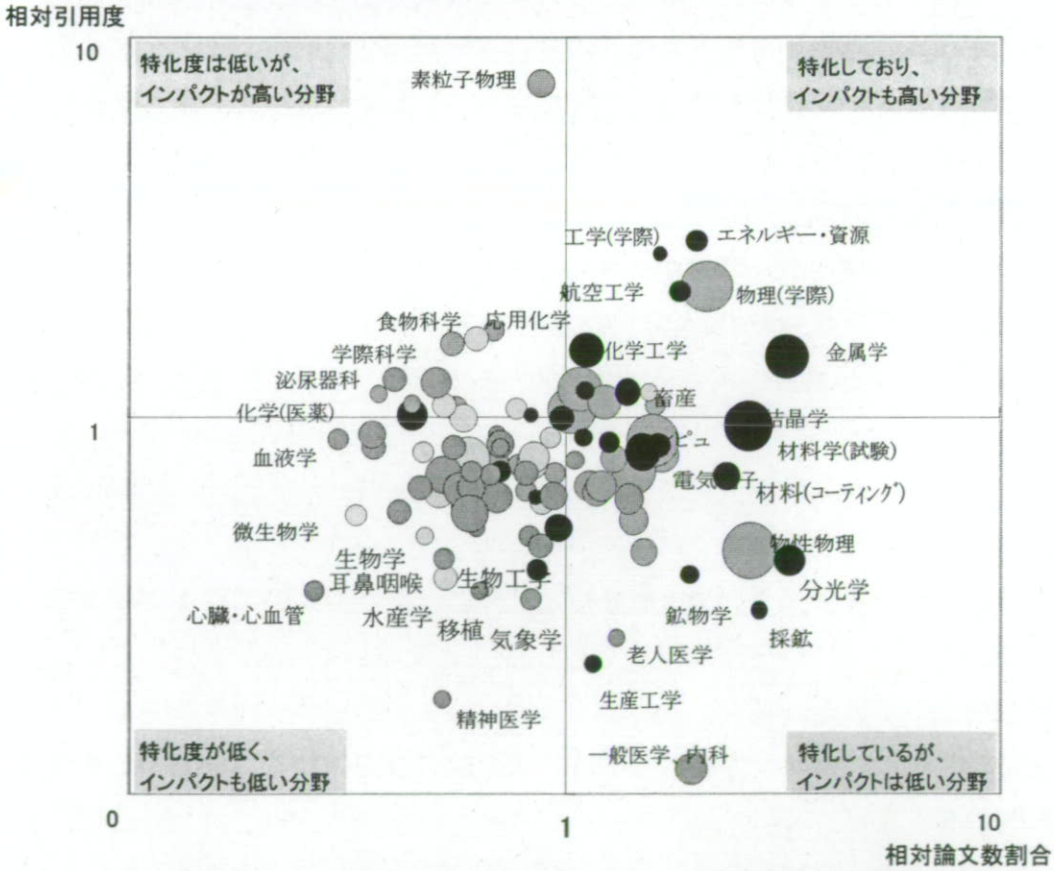

図5－3　論文数と引用度を用いた学科のポジショニング

一方、このような単純な生産性の分析以外にも様々な分析方法が存在する。

・ 共引用分析（共に引用されることの多い論文間の関係の強さを基にした分野の構造の分析）

・ 書誌結合（同じ論文を引用している論文間の関係を基にした分野構造の分析）
・ 共語分析（同じキーワードを用いている論文間の関係や、キーワード同士の間の関係を基にした分野構造の分析）
・ 共分類分析（同じ分野分類の論文間の関係や、分野分類同士の間の関係を基にした分野構造の分析）
・ 共著分析（異なる機関や国の研究者が共同著者になっていることによる連携関係の分析）
・ 謝辞分析（論文に記された研究資金への謝辞を基に、当該資金による各分野の発展への貢献度の分析）

前者 4 つでは、引用やキーワードや分野分類の共出現現象を用いて研究内容の構造を分析する。これにより、評価対象の機関やプロジェクトがそのような研究内容の構造の中でどのような分野に集中的にアウトプットを産出しているか、研究内容に特殊性があるか、研究内容の広がりと集中度はどの程度であるか等が分析できる。

また、プロジェクトやプログラムが当該分野の発展過程や分野間の融合や学際分野の発展にどのように影響したか、研究者の研究内容がプロジェクトの参加によってどのように変化したかの分析も可能である。

他方、共著分析では研究者の連携構造を明らかにするものであり、産学連携の推進などにプロジェクトやプログラムがいかに効果をもったかを分析することができる。

謝辞分析では、米国 NIH による資金提供効果の分析がなされており、また、英国 Wellcome Trust では謝辞情報も含む独自のデータベースを作成して自己の資金配分のバイオ分野の発展に果たした効果を分析している。

○特許指標

特許の書誌的側面を捉えて、論文分析と同様の分析を行うことが出来、特許の生産性の分析や構造の分析がなされる。特許分析はプロジェクトの事後に行うだけでなく、事前に行うことにより、プロジェクトで目標とする研究開発内容が既存の特許との関係においていかに優位性があるのか等を判断することができる。

また特許分析では、「サイエンスリンケージ」の分析がしばしばなされる。サイエンスリンケージは特許に引用された科学論文との関係を調べる方法である。米国特許では審査過程で（必要な場合には）先行論文を引用することが求められるため、このような分析が可能となっている（日本の特許ではそのような制度になっていない）。これにより、基礎的な研究がいかに産業研究開発へと結びついているかを評価することができる。

（2）テクノメトリックス

技術内容に関する数量的分析を**テクノメトリックス**（計量技術分析）という。例えば、製品の技術仕様（specification）が分析対象とされる。すなわち、製品にどのような機能が付与されているか、その機能の定量的特性（例えば容量や速度など）はどれほどであるか等が分析される。これにより、プロジェクトの目標値の設定やプロジェクト実施によるその達成、特定機能の他企業の製品への伝播などの分析を行うことが可能となる。

特許に記載された技術情報を手掛かりとして分析する場合には、特許分析ではあってもテクノメトリックスに分類する。例えば、ある組織や機関で展開された技術開発の様子を特許分析により跡付けることができる。

（3）エコノメトリックス

経済効果についての測定では、主にインプットと、アウトプット、アウトカムやインパクトの量を金銭的タームで測ることが行われる。古典的にはミクロレベルでの研究投資効率やマクロレベルでの生産関数を用いた分析がなされてきた。これらでは、研究開発プロジェクトへの投資額とそのプロジェクトによる成果に関連した製品の売上高等との関係が分析される。しかしながら、ここで注意しなければいけないことは、主に2点ある。

第一点はミクロなアプローチにおける問題であり、製品による売上高のうち、評価対象のプロジェクトやプログラムが直接的に関連している部分のみを抽出するという作業を行わなければ過大評価に陥るという点についてである。すなわち、特定の製品についても、評価対象のプロジェクトによる成果以外の様々な技術により製品は構成されており、さらに研究開発成果だけではなくマーケティングや流通や販売などの要素が売上や利益に大きく影響しているはずである。そこで、当該研究開発プロジェクトの売上や利益への寄与率を少なくとも2段階で同定する必要がある。第一は研究開発の枠組みの中での当該プロジェクトの寄与率であり、第二には全経営要素の中での研究開発の寄与率である。このような寄与率を客観的に決定することはほぼ不可能であり、関係者の主観的判断に基づき決定することになる。しかしながら、そのような場合においても、プロジェクトに直接かかわる範囲を同定するためには、プロジェクト参加企業からエンドユーザーまでに対してインタビューやアンケート調査を行うことが必要となり、膨大なコストがかかることになる。

また第2にはマクロなアプローチにおける問題点である。研究開発を含む経済活動

は一般に公的資金のみで行われているわけではなく、民間資金による活動との複合事象である。つまり、民間企業に対して公的資金を提供する研究開発プログラムやプロジェクトにおいては、単に参加企業の売上高の増加を計測したとしても、それは公的介入の効果そのものではない。従って公的投資による効果のみを取り出し、分離して把握する必要がある（図５－４）。この場合、公的資金提供によりそれがなければ生じない何らかの変化がもたらされていて、Georghiou や Luukkonen らはこの効果のことを**アディショナリティ**とよび、その分析手法の開発を意図してきている（例えば比較評価法を援用する）。

また一方では、公的資金を得た民間企業に対する効果だけではなく社会全体にいかなる効果をもたらしたかを分析する必要がある。この、プロジェクトに参加した企業以外の企業や産業全体あるいは消費者全体への効果のことを**スピルオーバー効果**と言う。経済学の用語としては厳密には対価を伴わない普及を**スピルオーバー**と呼び対

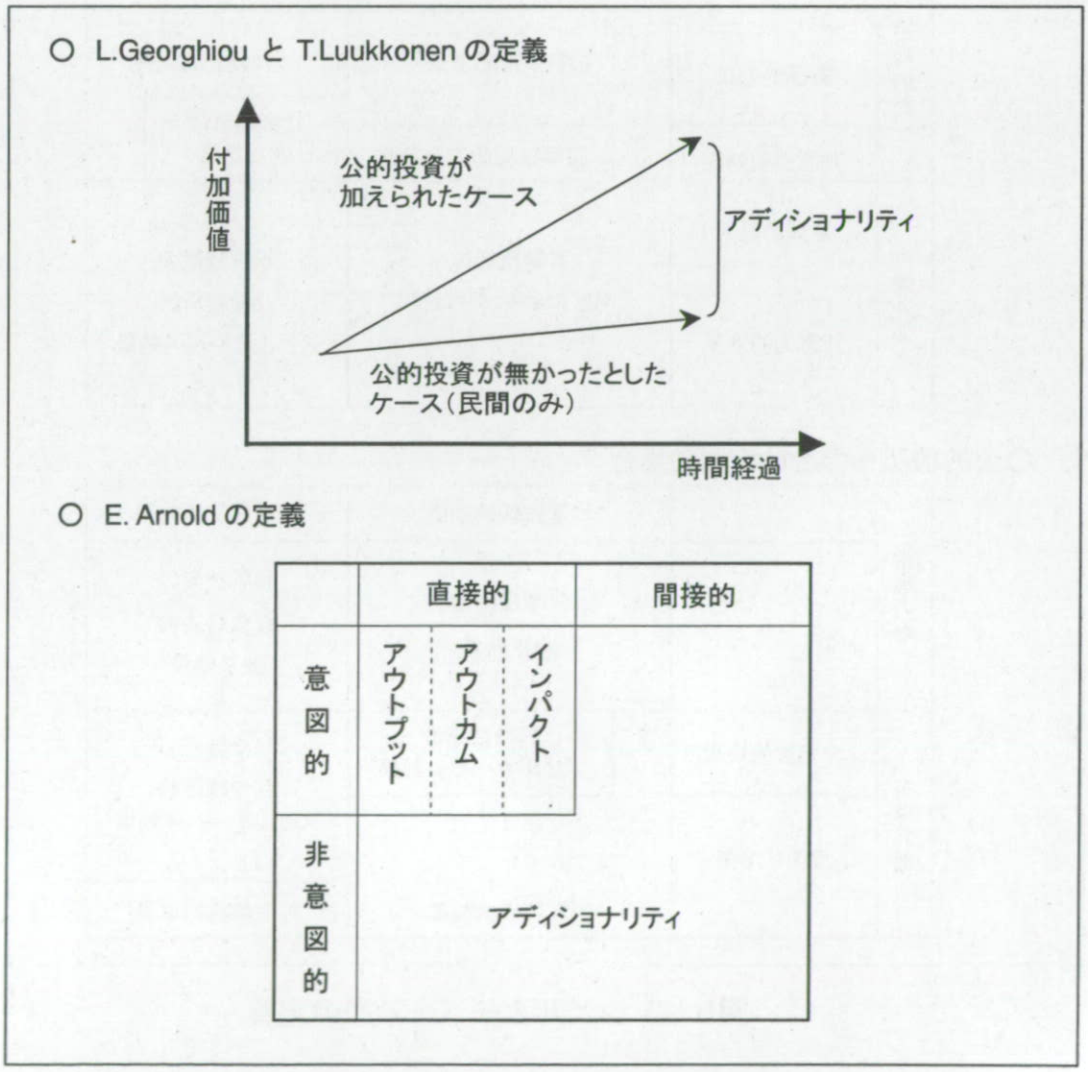

図5－4 アディショナリティの定義

価を伴う普及による効果は「波及効果」という。別の表現によれば市場内部で（従って対価を伴って）起こる普及と市場の外で（対価を伴わないで）起こる普及とを区別している。この意味での波及効果とスピルオーバー効果の測定を行う場合には、産業連関表等を援用し、直接プロジェクトに参加した企業だけでなく、顧客企業やサプライヤー企業の利益や仮想利益、ならびに競争相手企業の損失や仮想損失、さらにはエンドユーザーの利益や仮想利益も同様に計測されねばならない（図５－５）。エンドユーザーへの効果を分析する場合には、**消費者余剰**の観点からの分析がなされる。すなわち、エンドユーザーの需要曲線を想定し、研究開発の効果による価格変化などをもとに消費者が獲得する利益や仮想利益が計算される。

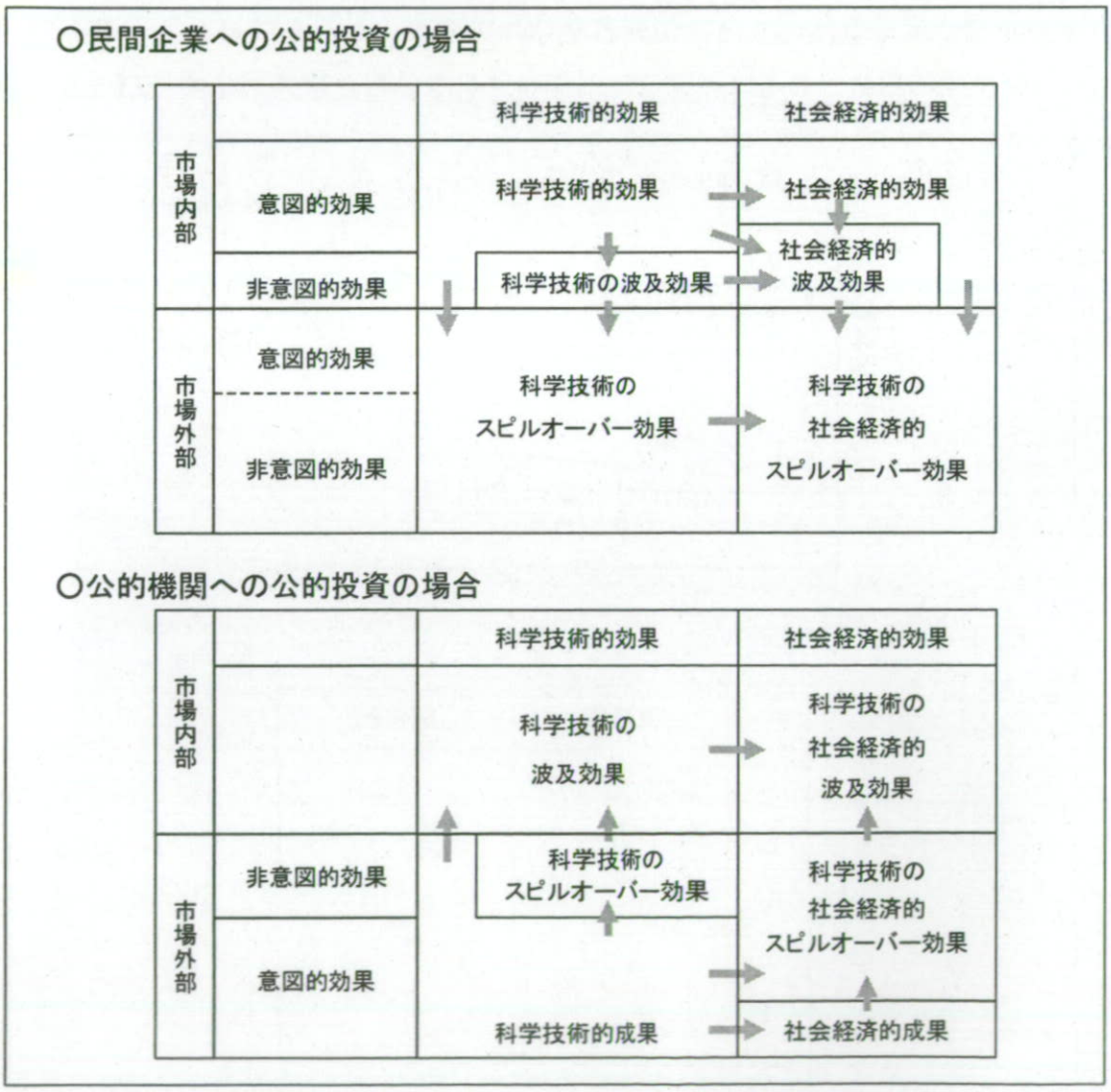

図5－5 スピルオーバー効果の定義

以上をまとめると、技術開発に伴う経済的効果はその技術が組み込まれて完成した製品の売上高や利益を単純に集計するだけではなく、以下の諸点を考慮する必要がある。

・寄与率：（技術開発内部における当該技術開発の寄与率）×（全経営要素における研究開発の寄与率）
・波及効果：市場内部で対価を伴って引き起こされる普及に基づく経済効果
・外部性：市場の外部で対価を伴わないで生起する普及に基づく仮想的経済効果（スピルオーバー効果）
・必然性：いずれの効果においても当該研究開発投資が無ければ生じない部分のみを抽出する（アディショナリティ）

いずれの場合においても、データ収集活動の困難さに加え、様々な仮定や推定を組み合わせる必要があり、信頼に足る分析を行うことは容易ではない。多くの事例分析を集積していく必要がある。

○BETA 法

BETA 法はミクロアプローチの例であり、フランスの Bureau d'Économie Théorique et Appliquée of the University of Strasbourg が開発した方法である。ESA（European Space Agency；欧州宇宙機関）プログラムの経済性分析に適用した事例に従ってその方法について紹介する。図５－６のように ESA プログラムに直接関連する行為（及びその他の行為）により企業が受ける影響・効果を、技術、商業、組織、労働の４つの要素に分類する。次にこれら要素が売上増加、コスト削減、企業の臨界質量（科学者や技術者の構造化されたチームの増加）に影響し、それらが付加価値の上昇や臨界質量の見積もり値に影響するというモデルを用いる。各影響関係において、どの程度の割合で影響があったか（寄与率）を参加企業（委託企業）のマネージャーを含む関係者に直接インタビューして調査し、各効果が金額として求められる。例えば、技術による経済的効果は次の式で得られる。なお、聴き取り調査で得る寄与率の回答には或る幅があるが、BETA 法ではその中の最小の寄与率を採用することにしている。

技術の経済的効果 ＝ 売上 × 付加価値率 ×（売上増加に技術要素が貢献した割合）×（ESA プログラムが技術要素に影響した割合）

この方法では、参加企業の多様な側面への影響を最終的には金銭的に算出できるという長所を持つ。その反面、単純にこの方法を用いるだけでは、参加していない企業

へのスピルオーバー効果について考慮していないし、また同様に経済社会全体に関するアディショナリティも算出したわけではない．一方、BETA 法による分析が適する対象は、開発側に近い内容のプロジェクトである場合や、中小企業支援のプログラムのように参加企業数も少数で独立性が高く、またそれら企業の研究開発のほとんどが当該公的資金により行われている場合などである。

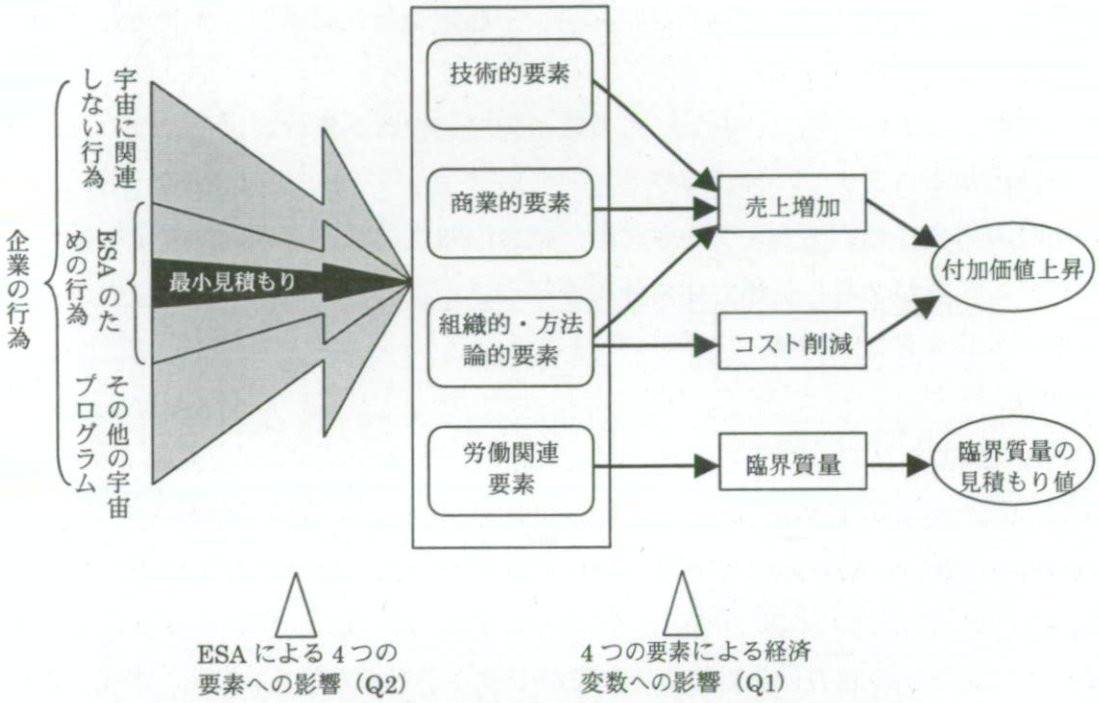

図5－6 BETA 法による間接効果の定量化（Cohendet 1998）

仮想的事例への BETA 法の適用例

PREST, AUEB, BETA, ISI, Joanneum Research, IE HAS, Wise Guys （2002), *Assessing the Socio-economic Impacts of the Framework Programme* より

《仮想的事例》

企業XXは精密機器・科学機器の分野の企業であり、ACPPプロジェクトという3年間のBrite-Euramプロジェクトの主委託先である。プロジェクトの目的は研磨作業の自動化の可能性の研究であり、最終的には自動研磨ロボットの仕様の作成であった。プロジェクトには他に自動制御の研究所、エレクトロニクスの研究所、光学分野の大学研究室、レンズ企業が参加した。これは新たなパートナーシップであった。XXはEUから€（ユーロ）300 000受け取ったが、プロジェクトにかかった全コストは€ 100万であった。

プロジェクトはほぼ100%成功し、翌年にロボットの試作が行われた。終了2年後には4台のロボットの販売、10台の追加注文を受けた。

プロジェクト中に、パートナーとともに新たなプロジェクトを計画し、ESPRITおよび国の資

金が得られる事が決まった。

プロジェクトによりXXは研磨加工技術についての知識基盤を向上させた。これにより、様々な材料向けの研磨加工の工具を最適化・品質向上でき、新たなデザインの研磨加工工具を装着した工作機器が販売されて競争優位性を得た。さらに、粉塵制御の技術は日曜大工工具を扱うXXの別部署に技術移転され、新たな装置の開発に活かされた。

また、XXはプロジェクトに参加し大学研究室と共同研究を行っている別の研究室に科学機器を販売した。

XXではプロジェクト管理者、光学研究者、メカトロニクス研究者の3人でプロジェクトを行い、管理者は自己の全業務の半分の時間、残りの2者は全時間を用いた。一人の研究者はプロジェクト終了後に退社している。プロジェクト管理者はEUプロジェクトの主委託先になったのは初めてであり、企業・個人レベル双方で管理能力（作業計画、資金配分、モニタリング、報告など）が向上した。また、企業XXは2年後に他のBrite-Euramプロジェクトを受託し、さらに1年後には新たなEurekaプロジェクトも受託している。

《BETA法による分析》

①直接効果：研磨ロボットの販売

販売台数：14

単価：€ 300 000

付加価値率：25 %（パートナーに支払われた10%のロイヤリティを含む）

直接効果 ＝ 14 × € 300 000 × 0.25 ＝ € 1 050 000

②間接効果１：研磨工具の新設計（技術的効果・製品）

販売台数（新たな研磨工具が装着された工作機器）： 45

単価：平均 € 160 000

付加価値率：35 %

係数Q1：この工作機器が売れている理由の3/4が技術の優位性であることがインタビューから得られた。

係数Q2：ACPPプロジェクトに関連した知識が工作機器の技術レベルに与えた影響は20～30%である。

間接効果１＝45 × € 160 000 × 0.35 × 0.75 × 0.2 ＝ € 378 000

③間接効果２：日曜大工工具（技術的効果・製品）

販売台数： 8 000

単価： € 380

付加価値率： 40 %

係数Q1：この工具が売れている理由は技術的優位性とマーケティングが50%ずつである。

係数Q2：ACPPプロジェクトから技術面への影響は10%である。

間接効果２＝ 8 000 × € 380 × 0.4 × 0.5 × 0.1＝ € 60 800

④間接効果３：科学機器の新たな販売（商業的効果・ネットワーク）

販売台数： 4

単価： €85 000
付加価値率： 40%
係数Q1とQ2：この顧客はプロジェクト開始以前には知られておらず、プロジェクトがなければ販売はなかった。よって寄与率は100%である。
間接効果３＝ 4 × €85 000 × 0.4 × 1 ＝ €136 000

⑤間接効果４と５：新たなナショナルプロジェクト・ESPRITプロジェクト（技術的効果・製品および商業的効果・ネットワーク）
二つの新たなプロジェクトで€200 000をXXは受け取った。
付加価値率： 100% （研究費)
係数Q1：新たな委託を得た理由は以下の2つ。XXの技術知識（50%）と優れたパートナーのグループに属していたこと（50%）。
係数Q2：新たな2つのプロジェクトの研究分野のうちの50%は当該プロジェクトを基に発展した。また、当該プロジェクトなくしてはパートナーと協力して新たなプロジェクトを申請することはなかった。
間接効果４（商業的・ネットワーク）＝ €200 000 × 1 × 0.5 × 1 ＝ €100 000
間接効果５ (技術的・製品) ＝ €200 000 × 1 × 0.5 × 0.5 ＝ €50 000

⑥間接効果６：プロジェクトマネジメント能力の獲得（組織的・方法的効果・プロジェクトマネジメント）
プロジェクトマネージャに獲得されたマネジメント能力はその他のチームにも伝播し、後の2つのプロジェクトについて2～4ヶ月の時間を削減した。最小見積もりでも、2人月に相当する。
プロジェクト管理者の給料＝€10 000 /月
間接効果６＝ 2 × 2 × €10 000 ＝ €40 000

⑦間接効果７：労働関連要素
ACPPプロジェクトは企業XXの知識基盤を拡大することに貢献した。しかし先端的でイノベーティブな研究開発に使われた時間は、プロジェクト管理者は10%、研究者は25～50%に過ぎず、残りはルーティンワーク（試料準備、データ分析など）であった。なお、研究者の一人は既に企業XXを退社している。
プロジェクト管理者の給料： €10 000 /月
研究者の給料： €7 200 /月
間接効果７＝36ヶ月 × €10 000 × 0.5 × 0.1 + 36 ヶ月 × €7 200 × 0.25 ＝ €82 800

⑧BETA法による分析結果
ECからXXへの研究投資＝ €300 000
直接効果／EU投資＝ €1 050 000 / €300 000 = 3.5
間接効果／EC投資＝€ (378 000 + 60 800 + 136 000 + 100 000 + 50 000 + 40 000 + 82 800) / €300 000= 2.83

５．４　単純評価法

本節以降は評価法そのものにかかわる解説であり，調査法（５．２)、分析法（５．３）に続き、評価法（表４－７参照）について順次説明する。なお、その際本報告書では「単純評価」で多用される方法論の質的向上を図ることに重点を置き、評価エキスパートが修得すべきプログラム、施策、政策等の評価で重要となる「複合評価法」には深くは触れていない。これらについては、企画を改めてまとめることにしたい。

（１）ピアレビュー法

ピアレビュー法とパネル法は多くの場合同時に適用される方法であり、概念を混同して用いられることもしばしば見うけられる。ここでは、以下のように両者を定義し本質的な違いを明確にしておく。

ピア peer とは「同僚」を意味し、**ピアレビュー法**は従って専門分野を同じくする者が当該専門分野の研究者ないし研究関連事項を対象として行う評価のことである。科学技術は、その内容が専門分化し、先端的な専門領域は、当該分野の専門家にしか適切に認識できないことが多い。従って評価に限らず、科学技術に関わる全てのマネジメント課題は、科学技術の専門家の知見を抜きにして適切に処理することはできない。この意味でピアレビュー法は科学技術や研究開発の評価法において最も重要な方法である。

しかしながら、ピアレビュー法は、専門家の知見に基づく評価であるが故に、科学技術の特定の専門領域の主として質的側面に関する評価にその効力が限定される。また評価対象が同一ないし近隣の研究者共同体内部に関わるため、評価者と被評価者が同一コミュニティ内部に所属することに由来する様々な攪乱要因の影響を必然的に受けることになり、その運用如何によっては大きな弊害をもたらすことにもなる。

ピアレビューの概念は上述のように単純であるが、専門家の知見を集約し、評価法として運用するためには弊害の混入を最小化するための定式化が必要である。ここでは特に我が国の現状で散見されるピアレビューの改善点を念頭に置いて、ピアレビュー法の要点を以下にまとめてみよう。

・ピアレビューアの役割をその専門領域に関わる状況認識と判断に限定する。

社会経済的側面やマネジメントに関わる判断はもとより、科学技術の非専門領域に関する評価者の一般的な知見等を安易に評価結果に取り入れるべきではない。あるいは、専門領域に関わる知見を尊重し、一般的知見と格差をつける。社

会経済的側面や運用的側面は、次項で述べるパネル法を援用し、また当該評価へのマネジメントそのものについては評価担当部署が責任を持つべきである。科学技術のピアレビューアが全てを仕切ってしまう方式は論外である。

・評価対象の専門領域の広さや特性に応じピアレビューアの構成やピアレビューの形式を柔軟に使い分ける。

科学技術のピアレビューアが最も能力を発揮できるのは基礎科学の領域であり、ピアレビュー法は基礎科学領域で有効なほぼ唯一の評価法である。逆に言えば、科学技術に関係してはいるが基礎科学以外の領域に関わる評価はピアレビュー法の他に、他の必要な評価法を組み合わせて実施すべきである。ピアレビュー法のみで基礎科学以外の領域の評価を間に合わせてしまうのは誤用である。

投稿論文や学位論文の審査のように評価対象が限定されている場合にはピアレビューアの構成は、論文が対象とする専門領域の関係者から成るように通常構成され運営されている。ピアレビュー法の原型は歴史的にもまた原理的にもここにある。公募型プログラムのプロジェクト採択に関わる評価の場合には、評価対象にある程度の広がりがあり、従って応募案件を専門性の特性に応じて区分し、サブカテゴリー毎にピアレビューグループを構成して評価することとなる。ここで新たに生じる問題はサブカテゴリーを相対化するための評価をどのように行うかという問題であり、「分野間の正規化ないし標準化問題」と言われている。通常この課題への対応もパネル法を援用することになり、分野間をカバーできる広いバックグラウンドと見識を有するパネリストで構成される第2段階評価パネルを設定し、そこでの判断に委ねられる。さらに学際的領域や社会経済的側面等を含む評価対象の場合では、第1段階としての部分的ピアレビュー結果ないし専門的分析結果を参考資料とするパネル法（第2段階）に力点があり、もはやピアレビュー法が中心ではなくなる。

・ピアレビュー法による場合の評価項目は、研究目的の重要性、研究内容の科学的意義、研究計画の妥当性、研究チームの能力、研究支援環境の適否、予算の妥当性などについてである。いわゆる研究メリット、研究手法、研究チームの質が主要3項目であるが、目的指向のファンドの場合ミッションに関連した第4の評価項目との整合性に配慮する必要がある。

社会経済的意義や社会経済的成果の可能性等の項目は、重要さの度合いにもよるが科学技術のピアレビュー法の範疇には入れるべきではない。それらを参考項目に近い形で加えるとしても、それはもはや純粋なピアレビュー法ではない。こ

のような項目の評価が必要になる課題の場合には、はやりパネル法を援用した2段階の評価法を設定するか、ピアレビューメンバーを部分的に含むパネルによる評価によることになる。社会経済的成果が問われる場合、当然のことながら科学技術的成果は最終成果ではないので、パネル法の運営の内部に科学技術ピアレビュー法が包摂されることになる。その際科学技術的成果の達成を自己目的化させてはならない。

・書面によるピアレビュー法の場合には、評価者と被評価者間及び評価者相互間の対面的接触をさけることが可能であるが、面接による評価やピアレビューパネルによる評価の場合、「グループダイナミックス」問題が必然的に発生する。

投稿論文の査読の場合には、"ブラインド・レビューイング"（被評価者の氏名等を査読者に対して秘匿する）が通常であるが、公募案件の事前評価のような場合では、書面による情報だけでは不十分なことが多い。また、対象領域を複数の専門家の組み合わせによりカバーし、その専門家グループによる合議が必要であることも多い。そこで多くの場合ピアレビューパネルによる評価を行うことになるが、その場合ブレーンストーミング類似の相補的ないしシナジェティックなプラス効果がある一方で、好ましくない効果も少なからず発生する。競って欠陥・誤謬・問題点探しに陥る「スタンドプレー効果」、少数意見を抑圧する「集団思考現象」、結論になりそうな意見に賛成する「寄らば大樹現象」、多数の中の一員であることによる「責任回避現象」等である。その結果として正統主義やリスク回避の傾向が強くなり、真に革新的、挑戦的な提案が低く評価されることになる。このような欠陥に陥らないためには、ピアレビューアの自覚や資質の向上に努めることも必要であるが、より基本的にはピアレビューパネルの司会者（責任者）のリーダーシップが重要であるといわれている。

・評価過程の公正さの維持とその過程の秘匿及び評価過程で知り得た情報の転用に関しては、最も厳格な倫理規程が守られるべきである。

しかしながらこのような倫理規程の整備や評価過程における倫理コードの確認等は十分には行われていない。「お手盛り」や「OBネットワーク」の存在が指摘されている。

以上述べてきたように、ピアレビュー法の改善問題は、マネジメント上の整理と共にピアレビューアの選任過程に課題が帰着する。我が国の場合、ピアレビューアの開拓はまだ進んでいなく、その選任はポジションや名声に依存することが多い。選択の枠組みを広げ、真に専門性と公正さを備えたピアレビューアの選任こそ最も重要な過

程であり、そのためには試行錯誤の中で次第に良質のピアレビューアを見出しストックとして集積していく以外に方法はない。

なお、以下にKostoffが指摘するピアレビュー法の質的向上のポイントと、Bozemanのピアレビュー法の使用上の指針についてまとめておく。

ピアレビュー法の質的向上のための8項目

Dr. Ronald N Kostoff "Research Program Peer Review: Principles, Practices, Protocols"より

最も重要な要素はレビューを行う機関のシニア・マネジメントの、高い品質のピア・レビューに対するコミットメントと、それに関連してピア・レビューを推奨する報償とインセンティブの配置である。

第2に重要な要素は技術的に信頼性のあるピア・レビューを行うレビュー・マネージャーのモティベーションである。レビュー・マネージャーはパネル評価において質疑応答を先導し、彼らのコメントを要約し、勧告を作成する。いくつかの機関においては、レビュー・マネージャーはレビューのプロセスおよび評価基準を選択する力量を持ち、すべての機関において、恣意的でない方法で評価者を選ぶ能力がある。もしもレビュー・マネージャーが、意識的にせよ無意識的にせよ、評価者を選ぶのに最高の基準に従わなければ、レビューのプロセスが始まる前に、評価結果は大きく左右されてしまう。

第3に重要な要素は、評価者の能力と客観性である。各評価者は担当分野について専門的に能力があるべきであり、研究対象の複合的側面（当該研究分野；隣接研究分野；研究により潜在的に影響を受ける技術、システム、ミッション）を評価者グループ全体の能力の範囲が覆い尽くしているべきである。更に、評価グループの専門分野は評価される研究領域に限定されるべきではなく（ジョブが正しくなされたかという問題）、当該研究の最高レベルの目的がカバーする研究分野にまで拡張されるべきである（正しいジョブがなされたかという問題）。この拡張された範囲によって、評価グループは正しいジョブがなされたかという広域の問題を扱い、革命的に新規なパラダイムにも公平な考察を行える可能性が増大する。

複数のプログラムやプロジェクトの比較に用いられるピア・レビューに関して、第4に重要な要素は、パネル間および分野間の正規化・標準化である。ある程度の類似性を持つ分野間に関しては、広いバックグラウンドを持つ（パネル間の）共通の評価者を用いることによりある程度の標準化は達成できる。極めて離れた分野の間では、当該機関にとっての各分野の相対的戦略価値を配慮する必要があり、採点の厳しさや偏向性の補正が必要となる。極めて離れた分野の間でも、広いバックグラウンドを持つ共通の評価者を多様なプログラムやプロジェクトの評価に用いることで、ある程度の標準化は可能となる。

第5に重要な要素は、評価基準の選択である。基礎研究提案の評価に際し、3つの主要基準は、研究メリット、研究手法、研究チームの質である。ミッション指向の機関に支援された研究の場合、ミッションに係わる第4の基準が便利である。ミッション指向の基準がより基礎研究指向の提案をふるい落とさないためには、ミッションとの関連性を極めて広く解釈することが必要である。基礎研究の場合には、長期的基準よりも、（転換や有用性といった）短期的基準の方が全体的な提案の質と点数との相関が高い。第5の評価基準として研究の総合的質を選ぶ

ことは重要であり、当該提案を評価するのに重要と評価者が考える評価リスト外の要素の影響を入れることを可能にする。例えば、評価者が当該提案は政府より産業界に支援されることがより適当であると感じたとしよう。この場合、この提案については、リストにある技術的基準の点数が高くても、総合点は低くなるかもしれない。

評価基準の第6の同様に重要な要素は秘密性である：評価者の匿名性および被評価者の非匿名性である。評価される研究の本質に関する正直で率直な見解が必要なら、評価者はレビュー・マネージャー以外には匿名でなければならない。提案（あるいは論文やプログラム）に対する否定的見解を表明する評価者にとって報酬はほとんどなく、評価者に対する懲罰や恨みは正直で率直な判断の科学に対する本質的利益をはるかに上回るだろう。

“ブラインド・レビューイング”つまり、被評価者の氏名と所属を評価者にわからないようにする事は、無名の研究者や評価の劣る研究所の研究者の仕事を公平に評価するという高貴な目的、あるいは性別といった個人的特性に基づいた偏向を除去するために用いられている。しかしながら、研究提案評価やモニタリング評価の研究によれば、プロジェクトの総合的質の決定に際し研究チームの質が最も重要な変数である事が示された。被評価者に関する情報を隠して研究評価を行うことは、方程式を主要な項を除いてから解くことに等しい。研究者に関する重要な変数を取り除くより、評価グループの視点を広めて、研究ターゲットの‘正しい仕事’の側面について言及できる追加的評価者を選ぶ事の方がより重要である。これにより、時流から取り残されたが論文生産が多く、被引用件数の多い研究が永続する事が無くなり、新しいパラダイムの新鮮な視点が公平に考慮されるであろう。

質の高いピアレビューにとって第7の重要な要素はコストである。ピアレビューの真の総コストは多額になりえる。しかしほとんどの報告事例では無視されるか過小評価される傾向にある。十分な数の専門家にパネルへの出席を求める質の高いピアレビューの場合には、実際の総コストは直接コストをはるかに上回るであろう。総コストの中で大きな割合を占めるものは、スタッフや評価者、発表者を含む、評価の実施に関するすべての参加者の費消時間である。質の高い実施者と評価者にとって時間コストは高く、評価コストの総計は総プログラムコストの中で無視できない割合を占めることとなる。多くの提案が拒否され、種々の支援者への重複応募が通例となっているような状況においては、支援される提案1件当たりのピアレビューのコストは大きく増加することとなる。提案（あるいは論文、プログラム）に関する質の高いピアレビューのプロセスを設計するにあたってコストは無視されるべきではない。

最後の重要な点であり、多分、質の高いピアレビューの基盤をなす要素であるものは高い倫理基準の維持である。真の専門家が評価者として望まれるとき、評価プロセス中には利害のバイアスや衝突が内在するので、科学的詐欺や科学的違法行為、秘密情報の流出、特権的情報へのアクセスにより不当な利益を取得するといった、倫理に関する多くの問題がある。

どのような文書に評価者がサインしようと、また最高の倫理基準を保つためにどのような望みを持つにしろ、彼らはアクセスする特権的情報に影響されざるを得ない。多くの経路を通じて知識移転は起こり、詳細な技術的説明を聞くことや詳細な提案書を読むことがおそらく最も効果的な知識の移転の方法であろう。通常の倫理的な評価者は秘密を漏らしたり、ピアレビューへの参加から不当な利益を得るために意識的で明白な行動をとることはしないであろう。彼等は大きな研究事業を手助けし、種々のアイデアに触れることにより自分の思考が拡張することの満足を参加の報酬として受け取る。高い倫理基準を保つために、専門家共同体内での自己規制と、はなはだしい倫理的逸脱に対し法的手段に訴える事が精力的に行われなくてはならない。

ピアレビューに基づく評価の改善のための指針

Barry Bozeman, "Evaluating R&D Impacts: Methods and Practice"より

指針１：ピアレビューを他の評価法と組み合わせて用いること。ピアレビューは主観的・解釈的・個人的なアプローチであり、専門家各人の持つ知識を最大限活用するが、同時に各人の持つ偏り、気質、死角を含むことも避けられない。ピアレビューの主観性は内在的なものであり、利益となる場合もあるのであって、それを減少させようとするよりは主観性の少ない他の指標で補完すべきである。この目的のためには客観的指標（論文数、引用数、特許件数などの生産性の尺度）の集合に基づくアプローチが最も適している。

指針２：ピアレビューを公開性の高い R&D 活動の評価に用いること。研究の大部分が基礎研究に近い場合でも、短期的な目標や固有技術との関連が明らかな場合にはピアレビューによる評価は適切でないことが多い。このような場合にピアレビューを行おうとすれば評価者すなわち競争者であるから、最良の場合でも感情的対立の可能性があり、悪くすれば訴訟に至る可能性もある。

指針３：評価者の選定が容易であること。評価者が得にくい理由には次のようなものがある。
(1) R&D 評価の実施責任者が当該科学技術分野との交流が乏しい。
(2) 評価される分野が小規模で特殊であるため適当な評価者を得ることが困難または不可能である（同分野の研究者が参加を拒むこともある）。
(3) 同分野の研究者が分散しており相互交流が少ない。産業界では特にこのようなことが多い。
(4) 分野が新しいため信用を保証するメカニズム（専門誌、学界など）が整備されていない。

評価者の選定の難易を問わず、評価者の適切性に何らかの疑問があればピアレビューによる評価自体が疑わしいものになる。したがって適切な評価者が得られる見込みがないか、参加に対する積極性が認められないならば、他の評価方法を採用すべきである。

指針４：内部の評価者を避けること。評価をインサイダーとアウトサイダーのどちらが行うべきかは評価や政策に関する文献で多くの論争を呼んだ主題であり、どちらにも明らかな利点が存在する。しかし R&D の評価においては、外部の評価者を起用する利益は他の場合にまさって大きい。

組織において、特に産業界の大組織において、自らに属する研究者に他部門の研究者の業績を評価させることはかなり普通に行われているが、この方式は正規の評価方法としては推奨できない。このような内部のピアレビューにもそれなりの用途はあろうが、評価結果が意思決定に影響するような場合は、内部ピアレビューは同系交配のようなものであり、視野が限られることになる。アウトサイダーを起用することは専門性を強化することに役立つだけでなく（組織内に十分な専門家が存在する場合もある）、視野や職業的・組織的文化の多様性を強める効果がある。内部ピアレビューは、廃止する必要はないが、正規の R&D 評価手段として用いてはならない。

指針５：脱機能的グループダイナミックスを防止すること。前述のようにピアレビューによる

R&D 評価におけるグループダイナミックスの研究例は僅かであるが（たとえば Porter & Rossini, 1985）、他の場合に極めて普通に見られる問題点をピアレビューのパネルのみが免れていると考える理由はない。グループ思考、「寄らば大樹」現象、声の大きい者が勝つ現象などのグループダイナミックスを防止する可能性のある方法として、ピアレビューをコンピュータ化し、匿名の評価者がリアルタイムで発言することが試みられているが（Bozeman & Bretschneider, 1990 など）、コミュニケーションの豊富さが減殺され、立ち上げが面倒であることが大きな欠点である。むしろ単に、このような現象が起こる可能性を意識し、それを防ぐために常識を働かせる方がよいであろう。たとえばパネルディスカッションを始める前に各評価者が個人的な評価を予め行っておくことは有用であり、これら個人的評価が匿名でパネルディスカッション以前に配布されればなお良い。

指針 6 ：尺度を用いるときはその妥当性と信頼性を検証すること。NSF や NIH の使用した評価尺度は極めて単純なもので、次元数も少なく、少数の段階を含む段階評価を行うにすぎない。このような単純さはある場合には望ましいものである。しかし政府の研究助成機関で自らの使用する尺度の妥当性や信頼性を検証しているところはほとんどない。ピアレビューで評価尺度を用いるときは、次のような問題のいくつかを検討しておくべきである。まず、評点は信頼できるか（同じ案件を繰り返し評価したとき、各評価者は前回と同じ評点を与えるか）。評価のずれは本質的に偶然誤差である。また、評価者間の信頼性はどの程度か。評価に大きな差があるならば、ピアレビューを利用するに充分なコンセンサスが達成されていない可能性がある。

定量的尺度は適切か。100 点満点の尺度を用いるのは、もし評価者がそのレベルの詳細さで質を識別できるのでなければ無意味である。反対に少数段階の評点では識別力が不十分な場合もある。尺度の適切さは経験的な問題であり、検証が必要である。評価の各次元の間にはどのような関係があるか。通常は各次元に等しい重みが与えられるか、または評価者にとっての重要性に応じた重みが与えられている。しかし次元間の多重共線性すなわち重複は統計的に検定すべき問題である。もしある次元での変動が本質的にすべて他次元で説明できるなら、その両次元に同じ重みを与えるのは適切でないであろう。

そもそも定量的尺度を使用する必要があるか。この問題には明確な解答は存在しない。まず必要なことは、評価者が定量的な評点をつけることを便利と考えるかどうか、またこれに関連して、用いられる評点システムを適切と考えるかどうかを知ることである。定量的評点の使用は評価の目的と関連していなければならない。多数のプロジェクトに関して助成継続の可否や助成のレベルに関連する決定を下すためには、定量的尺度は有用である。しかし単に進行中のプロジェクトの改善が目的であるならば、定量的評価はほとんど役に立たない。いずれにせよ、定量的評点には文章による説明をつけることが望ましい。これによって数値の意味や評価者の判断基準が明らかになり、また改善すべき点が示唆される可能性もある。

指針 7 ：評価者に偏りに関する申告書を出させること。National Academy of Science のパネルでは、まずパネリストが個人的な偏りや利害衝突の可能性に関する広範囲にわたる申告書を提出する。利害衝突に関しては明確な規則が大部分の政府助成機関に存在する。このような申告は、偏りの危険をすべて避けるには至らないとしても、部分的には有効な手段である。

（2）パネル法

パネル法は、専門分野の異なるパネリストから成る評価ボードを設定し、合議により評価結果を得る方式のことである。前項で述べたように、評価対象が広く、科学技術の専門性のみでは対処できない評価案件に対しては、知見を集約する最終的な場として、評価パネルを組みパネル法を実施することになる。その意味で、パネル法も極めて広く用いられる評価方法である。

パネル法のパネルメンバーは、従ってピアレビューアとは本質的に異なる資質を有する人物であるべきであり、それは狭い分野の深い専門性ではなく、広い相対的な知見と見識を備えるべきことである。また評価を中心とした経営の枠組みに関する知見や経験あるいは思考的枠組みを備えていることが望ましい。

科学技術のピアレビューパネルは、科学技術者の専門的知見を集約する場として設定されるのに対して、社会経済的側面に関しては、そのような当該課題に見合った専門的知見を備えた専門家は、学問の成立状況から考えてもほとんど存在しない。そこで、エキスパートによる調査分析を前段の活動として用意し、その知見をパネルに注入することになる。実務的な技術的課題に関する評価にしても、必要な知見は技術的学問領域とは異なる次元にあり、またそのような知見を有する実務的専門家が存在しているとしても当該人物はまさに競争関係にある人物であり、専門家として評価パネルに加えるわけにはいかない。従って評価に必要な知見は、何らかの調査分析的手法により生み出し、評価パネルにもたらされる必要がある。

パネル法の運営は対象案件の質的違いにより、異なるアプローチを適用する必要があり、科学技術の純粋ピアレビューパネルの場合は、当該分野の研究者による主導が妥当であるのに対して、社会経済的側面が含まれる混合パネルの場合には、案件のニーズ側の知見や政策的課題に詳しい専門家が主導するべきである。その中には政策担当者も含まれる。

社会経済的側面を含む課題に対する事前評価のための「メタ理論」

いわゆる実用化を目指した研究開発課題の事前評価の場合、実務的な視点からその内容を評価するために有効な「メタ理論」を以下にまとめる。「メタ理論」とは大きな枠組みや傾向としては妥当であるが、個別には例外を生じる類の、「理論」になる前段の理論のことである。組織現象は、本質的にこの種の「メタ理論」に依存することになる。

・大型ないし超大型の開発課題の場合

フェーズ管理の手法が推奨される。重要な科学的課題が克服されるまでは、技術的課題に本格的には着手するべきではない。大型装置やシステムの建設にとりかかる前に、重要な要素技

術（材料開発等）の開発やシミュレーションによる技術アセスメントを十分に行い、また、当該課題の経済性や社会性についての十分なフィージビリティ・スタディ（FS）により、その有効性や受容性（市場や社会への）が確認されていること。特に FS は重要で代替機能や、競合製品やシステムを含めたコストパフォーマンスの分析の他に、信頼性、安定性、効率性（空間利用効率、エネルギー効率等）、供給体制の可能性、協同事業領域の生態学的な成熟性等について十分に検討されるべきである。

・通常の開発課題の場合

当該研究開発案件が終了した段階で、実用化の見通しがついているためには、

既存市場の場合：コストパフォーマンスがよほど優れていること。

新市場の場合：材料開発の場合では用途が決まっている、部品開発では組み込む装置が決まっていることが必要で、つまり完全なニーズ型の開発体制になっていること。装置開発の場合は代替機能との競争力が格段にあること、システム開発の場合は、システム全体に関わる他の事業が成立し成熟した市場環境になっていることが必要である。

これらの「メタ理論」は過去の事例分析から得られたものである。

（3）評点法

評点法は評価者の主観的な判断を数量化して評価する方法であり、本格的な数学的方法を用いないという意味で、ここでは半定量的方法に分類した。次項で説明する**比率評価法**も単に数量の比率を取るだけであるので、同じ理由から半定量的方法とした。評点法も比率評価法も多くの場合他の評価法の一部に組み込まれ、要素的な評価法として利用されている。その意味において利用価値が高いといえるが、逆にこれらの方法の適用のみで評価が完結することは稀で、もしそのようにしたとすれば、評価作業が単純すぎないか反省してみる必要があろう。

評点法においては評点付けの手法を用いる（64 ページ参照）。

○評定段階の設定

評点法で評価対象プロジェクトに点数を付ける際には、評価項目ごとにどのように評定段階を付けるかについて先ず設計する必要がある。しばしば使われるものは、1 点～4 点、1 点～5 点、1 点～7 点、1 点～10 点などの段階づけである。

次ページの図５－７の例 1 に示すように、1 点～4 点のような偶数個の段階付けの場合は中間点が存在しないために、良いか悪いかの 2 分的な判断をつけたい時に用いられる。逆に例 2 に示すような 1 点～5 点の奇数個の段階付けでは「平均」を意味する中間点（3 点）が設定されており、多くの評価対象がこの中間点に評点付けされる。そのため、平均的なもの以外の、特に優れているもの（5 点や 4 点）や特に悪いもの（1 点や 2 点）に注目したい時に用いられる。この他に 7 段階や 10 段階といったよ

り細かい段階付けは、評価項目ごとの評点の総合点（合計点など）を最終的に算出する場合に用いられることが多い。

また、プロジェクト申請において多数の応募から上位のものだけを選択する時などは、平均以下のものに細かい段階をつけてもあまり意味がない。そのような場合には例3のように異なる幅をもたせた段階付けが行われることもある。

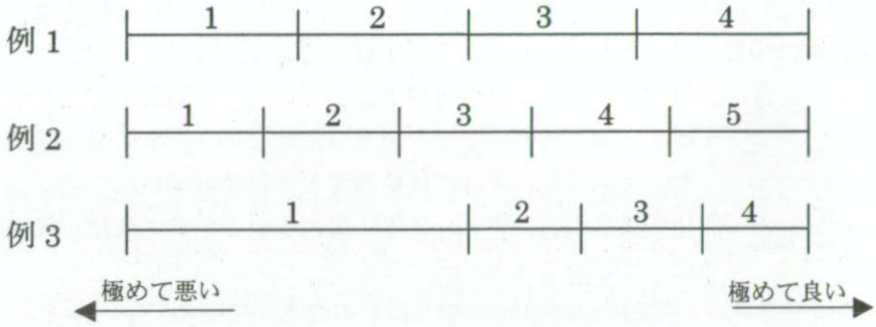

図5－7 スケール設定の例

ただし、実際に評点付けを行ってみると、例1の場合であっても「あまり悪い点をつけたくない」という評価者の心理から「3」が平均点として扱われてしまう傾向もある。また、英国の大学の研究評価であるResearch Assessment Exercise（RAE）では7段階設定であっても「1～7」ではなく「1、2、3a、3b、4、5、5*」といったラベルを用いている。これにより、数の多い「平均」（3点）をaとbで2つに分けることや、非常に優れているものだけを5*という特別な表現で切り出すといった工夫が行われている。このように、段階付けは、評価の対象や目的にあわせて設定し、いくつかの試行を経て適切なものに変更していく必要がある。

設定した段階付けは、それらの評定区分がどのようなことを意味するかを文章で明確に記述しておく必要がある。特に例3のように段階の幅が異なる時は、例えば「4点」は「極めて優れている」という抽象的なものではなく、具体的・解説的にどのような場合であるかを明文化しなければ、評価者の間で評点の付け方にばらつきが生じる。

○評点付けと総合化

上記のような段階づけのもとに各評価者は評点付けを行う。評点法においては、まず考えられる評価項目についてのリストを作り、項目にもれがないか、項目に重複的な内容がないかを評価者を含めて検討することが必要である。次に、評価者はヒアリングや報告書や各種データを基にして各項目ごとに評点を付ける。

必要な場合にはこのようにして得られた評点を合計して総合点を算出する。その方法にも様々なものがある。

・加算方式（各評点の和を計算する）
・連乗方式（各評点の積を計算する）
・加乗方式（下位項目については和をとり、上位項目については積をとる）
・ウェート補正方式（プロジェクトのフェーズ等にあわせて評価項目ごとに重み付けを行って和をとる）
・確率方式（評価項目ごとにリスクを推定してかけあわせ、和をとる）
・実測値との併用

以上のような方法から、評価対象にあわせて適切な方法を選択しなければならない。例えば、単純な加算方式は複数の評価項目の間で一つが低くても他がそれを補うほど高ければ良いという判断のもとで用いる。これに対して、連乗方式では、１つでも極端に評点の低い項目があると計算結果は大きく低くなるので、鎖状的に進行する連続現象の総合評価に適している。また、オペレーションズ・リサーチの分野ではファジー評価といった評点のバランスを重視する方法や、試行を繰り返すことで評価者の全体的直感と総合点を一致させるように各評価項目ごとの重み付けを決定していく方法などが提案されている。

また、単一の総合点を算出してしまうことによって逆に各評価項目ごとの数値が見えにくくなるという問題がある。そのためには、複数項目の評点の間に偏りや特異なものがないかを視覚的に把握するための下記のような方法がある。

・チャート方式（折れ線グラフ方式）
・ブロック方式（碁盤目グラフ方式）
・スケール方式（棒グラフ方式）
・ラジアル方式（円形グラフ方式）

ラジアル方式の例は図５－８のようなものであり、どの種類の項目群で評点が低いかを視覚的に見てとらえることができる。

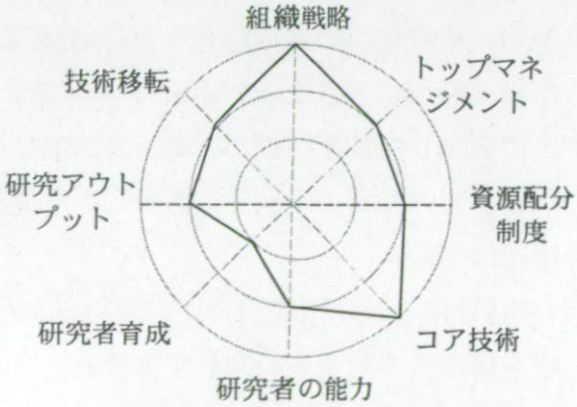

図5－8 ラジアル方式の例

（４）比率評価法

比率評価法は「効率」や「達成度」のような質的な指標（**示強性の量**）を算出して評価

する方法である。例えば「インプット」と「アウトプット」はいずれも量的な指標（**示量性の量**）であるが、その比である（アウトプット）/（インプット）はこのプロセスの「効率」を表し、プロセスの質的な側面を示す量となっている。評価においては基本的に示量性の量を把握しただけでは判断できないことが多い。それはその量に関する絶対尺度を我々が持ち合わせていないからである。そこで何らかの対比をとりその程度から判断をするのが一般的である。

対比の取り方としては以下のような比率評価指標が通常用いられている。

・対計画比（達成度）
・対目標比（到達度）
・対基準年比（成長率、成長度）
・対投入産出比（コストパフォーマンス）
・費用対効果（効率）
・対類似他者比（ベンチマーク）
・対チャンピオンデータ比（到達度）

また、この他に或る量の内部構成比のような比率の取り方もある。例えば、示量性の量である全コストの内の管理費の割合のような比率である。これはいわばコストの質に関係した量である。

このような比率指標を評価に用いる際に注意すべき点が二つある。第一は後節で述べるシステム評価法とも関連するが、評価対象を構成する変数の関係を評価システムモデルとして表してみると、一般に単純な線形モデルでは表すことが出来ないことが多い。このように変数の関係が非線形である場合、2つの変数間の比が何を意味するかは一義的に言うことは出来ない。つまり関連する他の変数如何によるからである。上に例示した比率指標が評価上意味ありげに思えるのは、評価対象の一部のみを取り上げ、他の変数との関係を捨象した単純線形モデルを想定しているからに過ぎないのではないか。

第2には、比率指標自身に定性的要素が含まれていることが多い点である。比率指標を構成する数量を吟味してみると、主観的な判断や推定に基づく数量が含まれていることがしばしばある。このような指標を**定性的比率指標**といい、実測値のような確定的な数量の比から成る**定量的比率指標**と区別し、評価指標としての信頼性の差に注意してそれを扱う必要がある。

このような問題点が有るとはいえ、比率評価法を抜きにしてはほとんどの評価が成り立たないこともまた事実である。

5．5　複合評価法

複合評価法はプログラム、施策、政策等のようなより高度な評価対象に対して主として適用すべき方法論である。ここではその概要を紹介するにとどめる。

（1）比較評価法

複合化された指標あるいは、生の指標そのものについて、複数対象間での比較を行うことによって評価対象の位置を把握することが一般的に行われている。また、同一対象であっても、時間的な前後や、「もし公的資金によるプロジェクトがなかった場合」などの非事実仮説との比較を行うこともある。

○ランキング

ランキングは多数の評価対象を特定の指標や総合指標によって順位付けするものである。ランキングはその結果として評価対象が一元的に並ぶので評価対象のポジションを理解しやすく、評価者への影響力が大きい。しかしその反面、適正な処理を行ってから比較するのでないと解釈の間違いを犯すことになる。例えば、研究成果の量の多い順にランキングする場合には、規模が大きい対象が有利で、上位に位置することになる。また、分野を明確に区分しないと、特に論文や特許では医学系などの平均産出頻度の高い分野の対象の方が上位となる。さらには、対象によって研究目的が異なるものを無理に一元的にランキングする場合もみられるが、このことは評価対象自体の目的と整合しないこともある。

また、ランキングにおいては1位から個別に順につけていくのではなく、第一グループ、第二グループというようにグルーピングのみを行うことも行われる。これは、順位の若干の違いが統計的にも意味が無い場合に、ランキングによる弊害をさけるための措置である。

プロジェクトの選定を行う際に、評価対象（プロジェクト）ごとに独立に点数を付けるのではなく、対象ごとの比較（どちらが良いか）を繰り返して全体の順位付け（ランキング）を行う方法がある。その典型的な方法が「一対比較法」である。一対比較法では、ある評価項目（例えば「技術発展のインパクトの大きさ」）について n 個のプロジェクトの間で主観的な二者比較を繰り返し行っていく。最終的には、比較によって優先順位配列を入れ替えていくことで決定するソート方式と、相対的な点数をつけて合計をとる方法がある。

米国 DOE の化石エネルギー局の先端研究プログラムにおけるランキングは後者の

方式を利用した例であり、探査、採掘、処理、利用の４つのプロジェクトを各評価項目に関して比較し、どちらがより重要であるかを４×４のマトリクスに主観的に点数で書き込む。このようにして付けられた得点から、各プロジェクトの順位付けを行うことができる。表５－１では各 a_{ij} は「比較して特に優れている」場合には5、「比較して特に劣る」場合には１の点数が書き込まれる。最終的には各プロジェクトごとに点数の合計がとられる。

欠点としては、プロジェクトの数が多いと比較回数が非常に多くなってしまうことや、人間の判断の一貫性のなさのために推移律が成立しない状態が生じることが挙げられる。

技術発展のインパクトの大きさ	探査 p_1	採掘 p_2	処理 p_3	利用 p_4
探査 p_1	3	a_{12}	a_{13}	a_{14}
採掘 p_2	a_{21}	3	a_{23}	a_{24}
処理 p_3	a_{31}	a_{32}	3	a_{34}
利用 p_4	a_{41}	a_{42}	a_{43}	3

表５－１　一対比較法の例

○ベンチマーク

ベンチマーキングは、焦点となっている項目について最良の結果（ベストプラクティス）を出している対象との比較分析を行い、自己の欠点を把握して改善を行う方法である。そのため、ベストプラクティスの対象が明らかでない場合は前項のようなランキングなどの作業が前提的に必要とされる。例えば機関評価において、マネジメントの面で優れていると言われる他機関との比較を行い、パフォーマンスの差と採用されたマネジメントの違いとの関係を分析しマネジメントのあり方について評価する。

○ポートフォリオ

ポートフォリオ法は個々のあるいは総合化された指標を軸として、評価対象の研究プロジェクト群をマッピングすることによりそれぞれの特性を可視的に示し、限られた資源の中で最適な配分バランスを見出すための方法である（図５－９）。軸の設定の仕方は様々であり、プログラム目的からブレークダウンして設定されるものや、市場の大きさ、リスクの高さなどを設定できる。

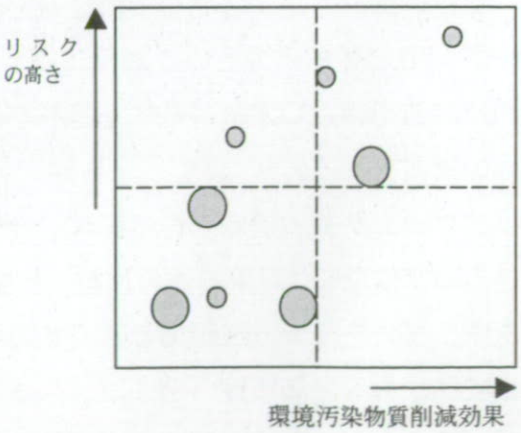

図５－９　ポートフォリオ

○コントロール・グループ・アプローチ

コントロール・グループ・アプローチは、規模や研究分野や事業分野がほぼ同一の

企業や研究組織について、公的資金を得たものと得なかったものを比較する場合等に用いられる。これにより、資金提供を受けた企業や組織が異なった質や量のアウトプットを産出しているかなどを分析する。ドイツのISIではドイツ経済省における中小企業支援プログラムの評価を行っている。しかしながら、分野によってはそのようなコントロール・グループを見つけることは不可能な場合もあり、たとえ可能であったとしても、別の要因が入る場合も多く、実際に分析を行いえるかを十分注意する必要がある。

○前後比較

コントロール・グループ・アプローチには上記のような問題があるため、同一企業や組織について、プロジェクト参加の前と後といった特定時間の間に行為やアウトプットがどう変化したかを比較する方法が前後比較である。問題点としては、プログラム外の影響を切り離すことが困難なこと、インタビューやアンケートでは公的支援を過大評価するような傾向が経験的に存在することにある。これまでISIにより「マイクロエレクトロニクスの応用」プログラムによるマイクロエレクトロニクス技術の製品への応用の振興効果の分析や、CAD/CAM導入の加速効果についての分析が行われている。

○非事実仮説

公的資金によるアディショナリティを判断するための論理的な方法は非事実仮説（公的プロジェクトがなかったと想定した場合）との比較である。この方法では、分析の妥当性を高めるためには詳細な仮説シナリオを設定して分析を行う必要がある。

（2）指標法

○指標間の比の算出

測定して得られたアウトプットやアウトカムやインパクト（さらにはスループット）の値に対して、それをインプットである研究開発費や研究者数などと比較することによって、研究投資の効率性を評価するものである。これは事後評価において行われるだけでなく、事前評価においても期待される効果を基にしてプロジェクト選択を行うのに用いる。

$$\text{研究投資の効率性}=\frac{\text{アウトプット}+\text{アウトカム}+\text{インパクト}}{\text{インプット}}$$

企業会計の分野においては、単純な売上高／研究開発費の比率だけでなく、様々な公式が提案されている。特に、事前評価において成功確率を組み込むものや新製品に

より利益が見込める年数を合計計算に組み込むものである。例えば下記はオルセンの利益指数法である。

$$\text{研究開発収益見積額} = \sum \left(\begin{array}{l} \text{新製品売上高の 3\%} \times 5 \text{ 年間} + \text{改良製品売上高の 2 \%} \times 2 \text{ 年間} \\ + \text{工程合理化節約額} \times 1 \text{ 年間} \end{array} \right)$$

$$\text{研究開発収益指標} = \frac{\text{研究開発収益見積額} \times \text{成功確率}}{\text{研究開発費見積額}}$$

また、オペレーションズ・リサーチの分野では、評価対象ごとに評価対象に有利になるように重み付けを変更する包絡分析法（DEA）なども提案されている。これは複数の対象を一律に評価するのではなく、各対象の特徴を考慮して評価するときに適している。

しかしながら、このように研究開発を対象に比率評価を行うには、十分な注意を要する。一つには、研究開発のフェーズが基礎寄りであればあるほど、アウトプットやインパクトが得られるまでには多くの年月が必要となり、短期的に測定を行うことは困難である。この場合にはプロジェクトのプロセス（スループット）のインプット比や目的との論理整合性により焦点がおかれることになる。また、効果が現れる長い年月の中ではより多くの要因が商業的成功にも関係するのであり、プロジェクトに起因する部分を抽出することは一層困難になることにも注意が必要である。さらに、当該評価対象に関わる範囲（特にスピルオーバーの範囲）をいかに設定するかによってこの比は大きく変化するものである。そのため、事例に即していかなるスピルオーバーが存在するかを定性的に調査した上で、比較を行うためには共通した境界基準を設定する必要がある。

○指標の総合化

複数の指標を総合化させて一つないし少数の指標にまとめ上げることがしばしば求められる。そのためには、まず第一に、指標が対象の各側面についてバランスよく収集されているかを再度検討する必要がある。研究の質などの重要な側面ほど単純には測定しにくいものであるため、再度、対象の設定する目的・目標からのブレークダウン及び対象の概念化によりその過不足を検討することが求められる。次に、それら複数の指標の総合化を行う。そのための方法としては、一つには評点法の項で示したような加算法や加重法を用いることができる。そのためには、各指標について基準を設定して評点に変換するなどの標準化を行うことが必要である。また、加算や加重を行う際には同時にそれぞれの項目の重み付けを行うことが必要であり、評価対象であ

るプログラムやプロジェクトや機関の目的・目標に照らして、評価者及び被評価者がともに検討してその重み付けを設定する必要がある。また、総合化の別の方法としては統計的手法を用いることができる。例えば多変量解析手法である主成分分析やクラスター分析を用いることにより、複数の評価項目や指標の間の関係を求めたりその総合化を行うことで、可視的な形で評価を行うことができる。

他方、評点法の項でも述べたように、指標の総合化をおこなわずに、それぞれの指標をレーダーチャートに表示するプロファイリングを行うことにより、対象の性質を把握する方法も用いられる。図５－１０はある大学について、科学的知識、教育・人材育成、イノベーション、公共政策、共同研究の５分野について指標をプロファイリングとして示したものである。

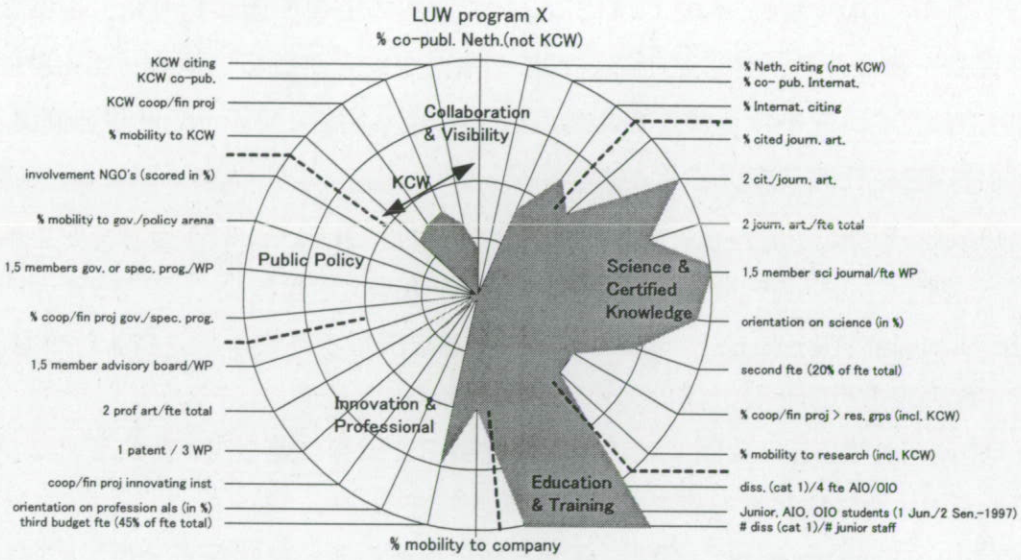

図5－10 プロファイリング

（3）システム評価法

システム評価では、上記のような指標間の比較だけでなく、より複合的に手法を用いるものである。一つの方向としては、ニーズや問題事項の把握や意思決定事項の選択といったプロセスをも評価のプロセスの中に明示的に組み込むことで意思決定システムとの統合をはかるものである。また一つには、複数の将来シナリオの設定やリスク設定によってシミュレーションを行う方法などがある。

対象を整理する方法として、マトリクス構造とツリー構造が用いられる。マトリクス構造では２次元などのマトリクスに整理することにより、複数の軸によって対象の構造化を行う。ツリー構造では、階層的な構造化を行う方法であり、リスク評価など

によく用いられる。また、将来の技術や社会変化の予測を行い、どの分野の技術を強く推し進めることが必要であるかを判断し優先順位付けを行うためには、「技術予測」手法が用いられ、シナリオ法、クロス・インパクト分析、デルファイ法などが用いられる。

以下では、システム評価の事例の一つであるDecision Management法を説明する。

○Decision Management法

Decision Management法（DM法）はスタンフォード大学R. Howard教授が確立した個別研究開発テーマを評価する手法であり、住友電気工業（株）などの企業で用いられている。この手法は、評価対象の構造的特徴を把握するための定性的分析と、期待利益などを計算するための定量的分析がシステマティックに連結していることに特徴がある。DM法は、大きくは構造分析、主不確実性要因決定分析、確率論的分析の3段階で構成され、さらにその内部は何段階かに分かれる（図５－１１参照）。

① 構造分析　研究を行う背景や問題点を明確化、不確実要因と正味現在価値との関係図を定性的に作成する。

a) Force-Field Diagram　プロジェクトの実行に賛成・反対する意見を集約して整理し、対立点・問題点・不確実要因を明確化させる。

b) Decision Hierarchy　既決事項、今回意思決定すべき事項、将来行う意思決定事項を明確化し、議論のポイントを絞る。

c) Strategy Table　今回すべき意思決定事項を横軸に並べて、それぞれについてどのような代替案や戦略があるかを整理する。

d) Influence Diagram　不確実要因と期待損益の関係を関係者全員が把握し、定量モデルの骨格を作成。

② 主不確実要因決定分析　不確実要因についてその最大値、基本値、最小値を数量的に把握する。不確実要因から正味現在価値を求める計算式を作成し、重要な不確実要因を把握する（定量分析）。

a) Spreadsheet Model　定量モデルを作成する。不確実な要因については楽観的な最大値、基本値、悲観的な最小値の3通りの予想案を考える。

b) Tornado Chart　不確実要因が損益に与える影響の大きさを把握する。全ての変数について計算をまず行い、次にその1つをとって変数を最大値から最小値まで変化させて、どの程度値が変わるかを求める。その変化量の大きい要因を上から並べる。経験的には30から50程度の不確実要因のうち、6程度の要因が大きく影響する。

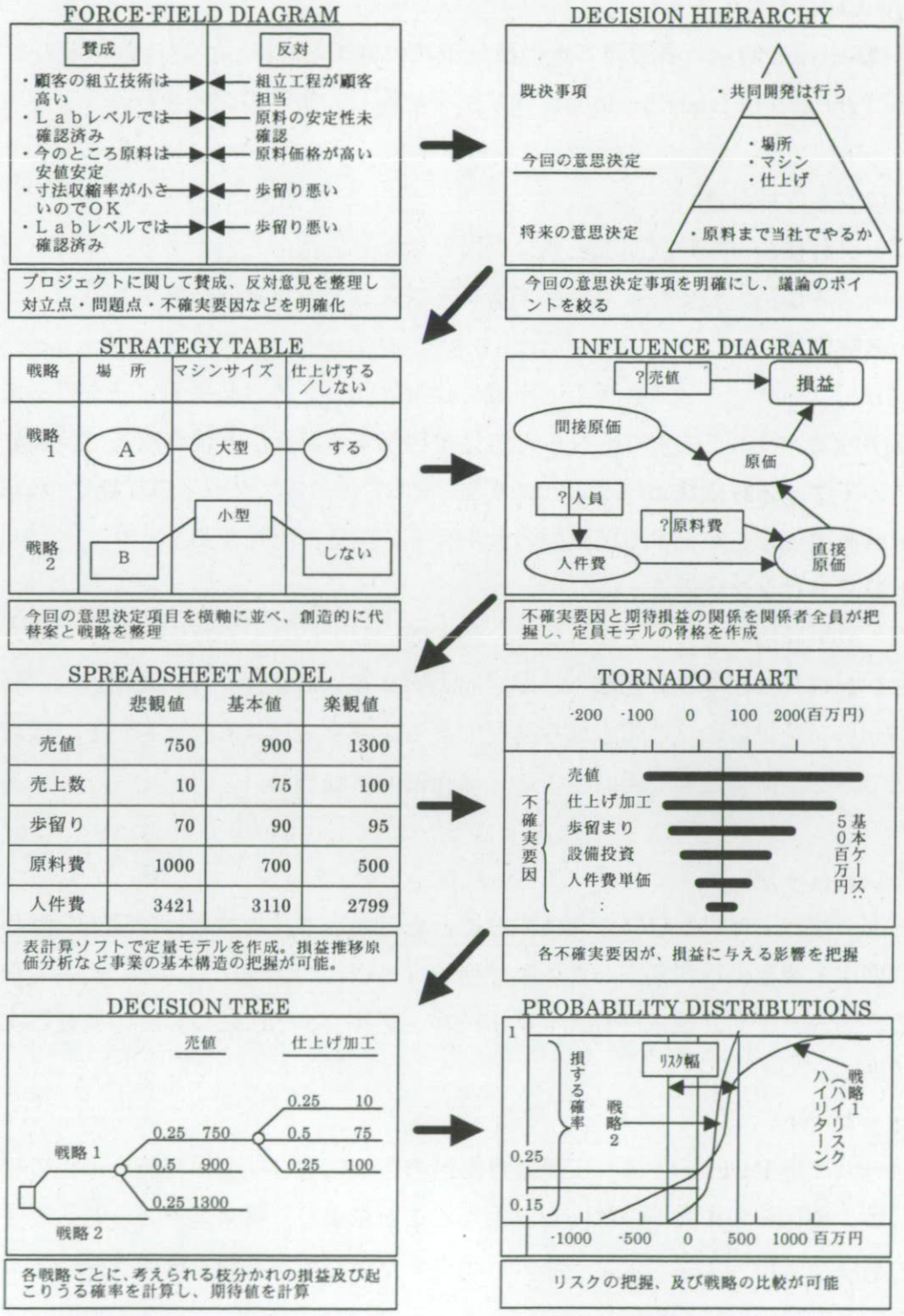

図5－11　Decision Management 法のツールと手順

③ 確率論的分析　不確実性要因のそれぞれの値が起こる確率を階層構造で計算し、期待利益を算出する。

a) Decision Tree　各戦略ごとの枝分かれの確率を計算し、期待値を求める。

b) Probability Distributions　リスクを把握し、戦略の比較を行う。

（4）ロジック評価法

ロジック評価法は論理展開を主要な手段とする評価法のことであり、事例を対象とするケース評価法や歴史性を対象とするレトロスコピック評価法等がある。

○ケース評価法

事例分析ないしケーススタディは施策や政策評価のための基礎的知見を得る最も有用な手段であり、ここではそれらの知見を利用する評価法と言う意味で調査分析作業を含めて**ケース評価法**と呼ぶことにする。なお、ケーススタディの詳細については、施策・政策評価を主題とする別企画のテキストに譲ることにする。

○レトロスコピック評価法

施策・政策評価の有力な手段として、歴史遡及分析に基づく評価があり、ここではそれを**レトロスコピック評価法**と呼ぶ。通史的な知見に基づく評価に相当し、歴史のキーパーソンを対象にした聞き取り調査であるオーラルヒストリー等がその代表的な手法である。前項と同じ理由でここでは詳細には触れない。

（5）レビュー法

レビュー法は、将来展望等を評価することを目指して、当該の科学技術分野やイシューに関する複数の専門家等の主体参加型で行う総合的評価法である。代表的な手法であるフォーサイト、フォーカス・グループ、デルファイ法についての概要は次のとおりである。

○フォーサイト

フォーサイト Foresight は、広範な専門家の参加により社会・経済・市場の相互に関係し合う要因等の中長期の動向を同定することにより、将来ニーズに対処できるように、科学・技術・工学・インフラストラクチャーに要求される展開を探索する方法である。

社会や産業界のためにビジョンを創りあげ共通する課題等を抽出し、社会の各主体に、将来に関する筋の通った統合された見方を提供し、変化に関する情報を伝えて、将来に備え対処することを助ける狙いがある。すなわち、技術予測とは異なり、科学

技術動向のみならず、産業界や需要側の主体も参加させて市場や社会の動向も勘案し、潜在機会を洞察して技術の商業化時期などを想定することに特徴がある。将来を見通すことを試みるために、広く関連主体をパネルに参加させてシステマティックなやり方を導入する。

英国で初めて 1994 年に着手されたプログラムは、1 万人に及ぶ社会に開かれた幅広いコンサルテーション過程を通じて、科学技術・イノベーション政策目標の策定や研究予算の配分にまで適用する指針を見出そうとするものであった。当初は政策過程の基礎に置かれ参照されたが、現在は影響力が低下している。地域などのバランスをとりつつ、経済界、科学基盤（高等教育機関や研究機関等）、研究公益団体、政府から専門家が選ばれ、テーマや産業セクターやクロス・テーマ群別にパネルが作られる。各パネルは詳細情報を提供するタスクフォースにより支援され、また、特定トピックの将来動向は専門機関や研究・技術機関により調査される。また、個人や企業が将来ビジョンを描けるように、プロセスを通じた情報を知識プール化して利用できるようにしている。

○フォーカス・グループ

フォーカス・グループは、ピアレビューやワークショップなどと同様に、グループディスカッションにおける意思決定や問題解決などの協調作業による総合的評価方法である。なかでもフォーカス・グループは、あるひとつのテーマと参加メンバー人選において焦点が絞り込まれた、非常に組織化された集団討議である。政策、プログラムの企画立案や評価に有効とされる。テーマやトピックの専門家により選ばれた主題について集中的に討議が行われるので、非常に迅速に、効率的に多くの具体的な情報や豊かな知識を得ることが期待される。米国大統領サイドが政策評価に利用したり、国際標準の設定などに利用されている。また、フォーカス・グループは、プロジェクト評価などで、インタビューよりも効果的で未来指向の成功要因にフォーカスすることができるともいわれる。

○デルファイ法

デルファイ法は米国 RAND により開発された方法で、多数の専門家に対する同一内容の繰り返しアンケートにより、専門家集団の直観的な判断を集約する方法である。2 回目以降のアンケートでは、前回のアンケート結果から中位点、上下四分位点、四分位点外の回答者のコメント（無記名）などの情報をフィードバックさせることにより、回答分布が収斂する傾向を利用している。比較的長期の技術予測に利用され、わが国では旧科学技術庁が 1971 年から 5 年毎に未来技術の開発・実用化時期の予測に

採用し、70％程度の精度での予測結果を得ている。

ブレーンストーミング法のように会議を採用せず、討論の場合に起こりがちな心理的錯乱要因、例えば権威従属性や多数意見に対する付和雷同性を避けることができるとされるが、質問内容や専門家の選定に留意が必要である。

第６章　評価推進体制の策定

６．１　共通事項

評価を実際に行う場合に、まず必要となるのは、評価推進体制の設計である。評価推進体制は評価目的、評価対象、評価フェーズ等の違いにより大幅に異なる。しかし個別の評価体制の説明に移る前に、各評価体制に共通した基盤的事項についてまとめておく。

評価推進体制の設計にとってまず必要なことは、評価の枠組みを明確にする事である。評価の枠組みとは、評価システムと評価体制の枠組み、つまり広い意味での評価方法とそれを実施運営する際の人の問題である。

（１）評価システムの枠組み

評価システムの枠組みは事前評価とそれ以外の場合では全く異なる。

事前評価の場合、「目的の妥当性」、「目的の重要性」、「目的の実現可能性」が主要な評価項目であるのに対して、途上評価、直後評価、追跡評価等の事前評価以外の場合では、「実績の状況」、「費用対効果の様子」、「実績（成果とコスト）の比較」、「マネジメントの妥当性」が主要な評価項目となる。

事前評価では評価結果は案件の採否等の意思決定のための参考資料に供されるが、それ以外の場合では案件の継続の可否の他に「目標や計画の修正」、及び「目的自体の見直し」のための材料に用いられる。

いずれの場合においても、評価システムとしては評価項目、評価基準、評定区分を具体的に設計する必要がある。上記項目を具体的に展開した評価項目と評価基準の標準的あり方の例を表６－１にまとめる。

事前評価の場合、「目的の妥当性」は上位目標やその状況変化との関連において評価されるので、案件がそのように整理された階層構造の下で設定されていない場合にはそのための調査分析が前段の知識として必要になる。この作業は簡単ではない。「目的の重要性」は案件の内容的妥当性に係る評価項目で、具体的には「期待される成果の意義やメリット」が十分であるかどうかが問われることとなる。科学的ないし技術的な意義やメリットの場合は、対象領域を専門とする科学者や技術者の専門的知識を動員

表6－1 評価項目と評価基準の総括表

評価フェーズの区分	評価項目	評価基準
事前評価	1. 目的の妥当性 a. 上位の目的に対する適合性 b. 上位の状況の変化に対する適合性	・上位の目的や計画に包摂されているか（予想される大状況の枠内に目的が設定されているか） ・上位の状況が変化している（変化が予想される）場合、それに対応した目的の設定になっているか
	2. 目的の重要性 a. 内容の妥当性	・期待される成果の意義やメリットは十分か －科学的意義やメリット －技術的意義やメリット －経済的意義やメリット －社会的意義やメリット
	3. 目的の実現可能性 a. 計画の妥当性 b. 手法の妥当性 c. 実施者の編成 d. 実施者の能力 e. 実施環境の適否	・成果を挙げるための手順は適切か ・成果の実現時期は適切か ・適用される方法論やアプローチは適切か ・チームメンバーの構成は適切か ・リーダーの指導力は十分か ・チームメンバーの能力は十分か ・実施環境は十分支援的か
	4. 実施の可否	・実施に移すべきか
途上評価・直後評価・追跡評価	1. 実績の状況 a. 成果の内容 b. 成果に対する寄与 c. コストの内容	・個別課題毎に、どのような成果が挙ったか。量的側面と質的側面について －アウトプット成果 －アウトカム成果 －インパクト成果 －副次的成果 を挙げる ・個別成果毎の当該案件による寄与率はどの程度か ・個別成果毎に要した費用はどの程度か
	2. 費用対効果 a. 費用対効果の様子	・主要成果の費用対効果はどの程度か ・当該案件全体としての費用対効果はどの程度か ・その他の比較（対計画比、対基準年比、対類似他者比等）における費用対効果の実績はどの程度か
	3. 成果の比較 a. 対計画（目標）比実績 b. 対基準年比実績 c. 対類似他者比 d. 対チャンピオンデータ比	・成果の達成度は計画（ないし目標）と比較した場合どの程度か ・基準年の成果に対する成果の成長率はどの程度か（対前年比等） ・類似した他の案件と比較し、成果はどの程度か（ベンチマーク、ランキング） ・チャンピオンデータに対する成果の到達度はどの程度か
	4. コストの比較	3. の成果に対する比較と同種の比較
	5. マネジメントの妥当性 a. マネジメントの枠組みの妥当性 b. マネジメント手法の妥当性 c. リーダー等の指導力 d. メンバーの編成 e. メンバーの能力	・マネジメントの枠組みやシステムは適切であったか ・取り組み方は適切であったか ・リーダー等の指導性は十分発揮されたか ・メンバーの構成は適切であったか ・メンバーの能力は十分発揮されたか
	6. 継続の可否（途上・中間）	・継続すべきか
	7. 目標や計画の修正（途上・中間） a. 目標の修正 b. 計画の修正	・修正すべき目標は何か ・目標をどのように修正すべきか ・修正すべき計画は何か ・計画をどのように修正すべきか
	8. 目的の見直し（直後・追跡） a. 目的自体の見直し	・当初の目的を見直すべきか ・目的をどのように見直すべきか

することにより、その価値判断は比較的容易につけることが出来る。しかし、経済的ないし社会的意義やメリットについては、そのこと自体を対象とした学問が体系化されている訳ではないので、当該案件に詳しい専門家を見出す必要があり、利益相反のないその種の専門家を見出すことは極めて困難である。従ってこの場合は、通常やはり調査分析を事前に行っておく必要がある。「目的の実現可能性」についても同様の状況があり、上記と類似した作業が必要である。

途上評価、直後評価、追跡評価の場合では、「実績の状況」、「費用対効果の様子」、「実績の比較」、「マネジメントの妥当性」の各項目について把握する必要があり、そのためには、単に実施者からの報告書や聴き取り調査だけでは不十分であることが多い。特に、アウトカム、インパクト、副次的効果の把握については実施者から離れた独自の調査と分析が欠かせない。また、「目標や計画の修正」、「目的の見直し」については、評価者の幅広い見識が必要である。

なお、機関評価は基本的に途上評価のカテゴリーに属する。

（２）評価実施体制の枠組み

評価システムの枠組みが以上のようなものであるとした場合、その実施体制としては、評価システムの設計者である評価担当部署の担当者の他に、評価パネルの責任者とメンバー、そして評価対象の調査分析を担当するエキスパートが欠かせない。

評価パネルに関しては、パネルの設計上以下の諸点に注意する必要がある。

まず、ピアパネルとして組織すべき案件であるかエキスパートパネルであるべきかの区別である。この点に関しては５．４節で詳しく述べた。

２番目は、機関評価の場合に関してであるが、評価パネルをどのような位置付けに置くかという問題がある。例えば、内部の者達で行うのか、あるいは外部の専門家を招聘するのか、あるいはそのパネルの位置付けとして、外部から評価者を招く場合であっても、内在的な立場でパネルを運用するのか、あるいは第三者として外在的位置から評価を行うのか、こういうパネルの位置付けをどのようにするのかということは、如何なるパネルメンバーが相応しいかを判断する際だけではなく、評価システムとしても重要なポイントになる。

このような共通事項は何を評価対象にするか等によって、その評価要素の重要さが異なり、またさらに評価要素の詳細なポイントを詰めておく必要がある場合もある。

次に、いくつかの典型的でかつ重要な評価局面を取り上げて、その各論を述べたい。

6．2　従属プロジェクトの評価

従属プロジェクトの評価は、最も典型的であり、また現在我が国においてこの種の評価の質的向上が喫緊の課題となっている。その具体的な評価システムについて海外の設計事例と共に詳しく見てみたい。

（1）従属プロジェクトの事前評価

当該プロジェクトの目的が科学技術の価値の発展にある場合、つまり科学技術的な知識の増進に目的がある場合と、社会経済的な何らかの効果を意図している場合とでは、その推進体制をかなり変える必要がある。科学技術内部の価値形成に目的を限定したケースは、より単純であり、この場合には評価パネルのメンバーは、いわゆる科学技術のピアレビューアで良く、設定領域の広さにより文書ないしメール方式を用いるか対話を必要とするピアパネル方式にするかを選択する。推進体制の設計のポイントは、如何に科学技術のピアレビューの質を高めるかにあり、そのポイントは、評価項目、評価基準、評定区分から成る評価体系をどのように設定するか、さらには評価パネリストと評価パネルをリードするパネルディレクターの質如何にかかっている。

NASA の宇宙科学関係のプログラムの評価方法についての事例を参考資料 3.1.1 に挙げておく。

社会経済的な効果を目的に含んでいる場合には、科学技術のピアレビューアの他に、社会経済的な評価が担当できるエキスパートを評価パネルの中に入れる必要がある。比較的小規模の場合には、そのエキスパートパネルの判断を尊重するだけで、社会経済的な効果についてはある程度明確に判定できるが、より広い範囲の複雑な社会経済的効果を目的としている場合には、評価パネル用の資料が必要になり、定性ないし定量的な調査分析が必須となる。また、その調査分析の一環として提案者に対するヒアリングを行う事も必要になるであろう。このケースの事例として、テクノポリスが担当したイギリスの DETR の事例（参考資料 3.1.2)、米国メイン州の科学技術振興プログラムにおけるプロジェクト評価事例（参考資料 3.1.3)、及び米国商務省の ATP の事例（参考資料 3.1.4）等を挙げておく。

（2）従属プロジェクトの中間評価と事後評価

我が国では、資金総額が 1 億円を超える程度の比較的大型の従属プロジェクトの場合、中間ないし事後評価は、実質的に 2 段階で行われている。第 1 段階の評価はプロ

ジェクトの推進ないし助言を兼ねた評価委員会であることが多く、いわゆる内在型で評価が行われる。第２段階では実施者と第１段階評価者からの報告を基にして、第２段階用に設定された評価パネルで評価される。第２段階の評価パネルは対象領域毎に設定され、従って複数の案件が対象になる。この場合、パネル法による評価が行われることになり、その目的に相応しい広い識見を有するパネリストを用意する必要がある。

また、ある場合にはこの第２段階評価の結果がさらに上位の評価パネルに年度ごとに報告され、プログラム全体のプロジェクト案件の年度ごとの評価が行われることがある。これは、実はプログラムの途上評価に相当するが、我が国では案件全体に関する事前の調査分析が通常ほとんど行われていないため、本格的なプログラム評価とはなっていない。この関係を図６－１に示す。

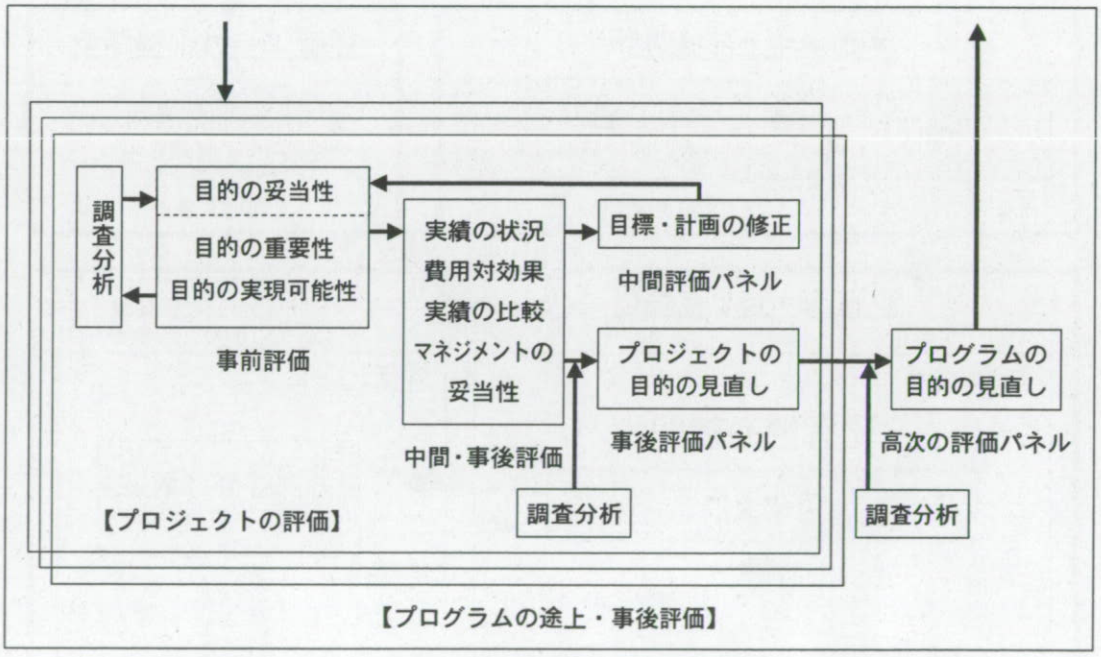

図6－1　プロジェクトの評価とプログラムの途上・事後評価

６．３　機関の途上評価

機関評価は通常、機関の途上評価 monitoring evaluation に相当し、年度ごとに行われる自己評価と、数年に１回行われる外部評価がある。機関評価は、従ってプロジェクト評価の実績評価と類似している。

機関の途上評価の場合には、大きな評価の枠組として実績評価とマネジメント評価

がある。実績評価は、設定された目標や計画に照らし合わせて、それが実現されているかどうか、どの程度実現されたか、あるいは、予期しない効果が発生しているかどうか、といったことについての評価であり、そのための独自の調査や分析に基づきデータを整理することにより状況を明確に認識することができる。このような実績評価自体は、それほど困難なことではない。

機関評価では、評価パネルの位置付けが重要である。また、評価パネリストの選任をどのように行うかについても、評価パネルの位置付けと関連して整合性のある方式とすべきである。評価パネルの位置付けとパネリストの選任方式については、図６－２に示す４通りがある。

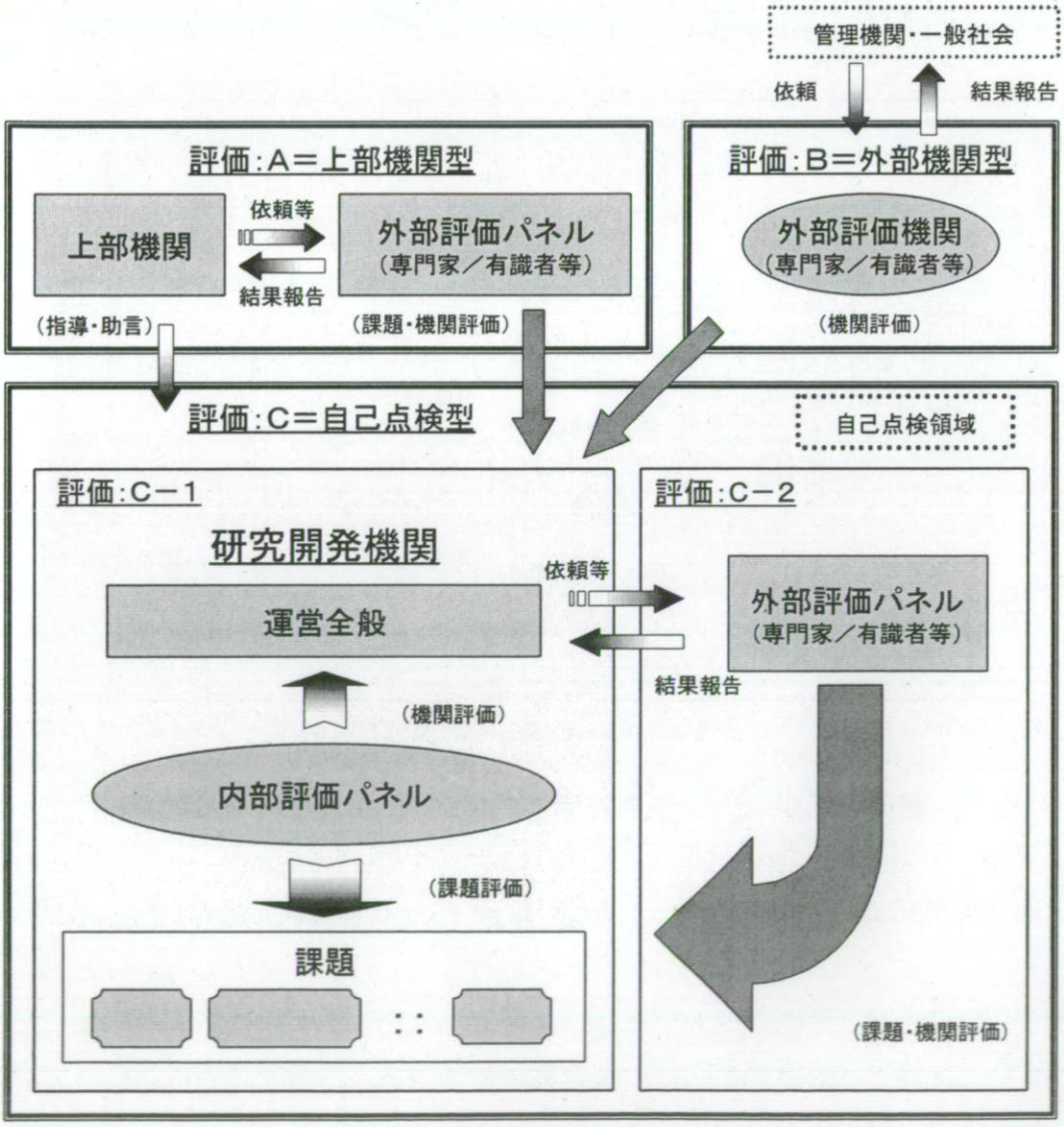

図6－2 研究開発機関における評価の枠組み

我が国における例では、独立行政法人の評価委員会は上部機関型であり、大学評価機構は外部機関型、また国立研究所の大部分は研究所が依頼した「外部評価委員」による自己点検型（C-2）である。いずれの場合においても機関評価を行う前に自己評価（C-1）を実施するのが通例となっている。地方自治体の試験研究機関の場合は、上部機関型が多いが、機関評価の出来栄えを左右する大きな要因は自己評価が緻密に行われるかどうかにかかっている。なお、我が国ではC-2のタイプを「外部評価」と呼んでいるが、C-2タイプは本来自己点検型であり、内部だけでは見過ごす欠陥や弱点についての助言を「外部パネリスト」から受けることが出来るメリットはあるが、そのパネリストは被評価者が依頼した評価者であるため、A、B 型の「外部評価」がもつ公正さや透明性に欠け、紛らわしい呼称である。

また、自己点検型はもとより上部機関型や外部機関型（第三者評価）であっても、公的資金による機関である以上、評価経過や評価結果は公開されるべきである。また、社会からのコメントを受け取る窓口を同時に設定し、オープンアドバイスを受ける体制を整えることも重要である。

機関を対象とした評価システムと作業の流れについて、図６－３にその概要を記し

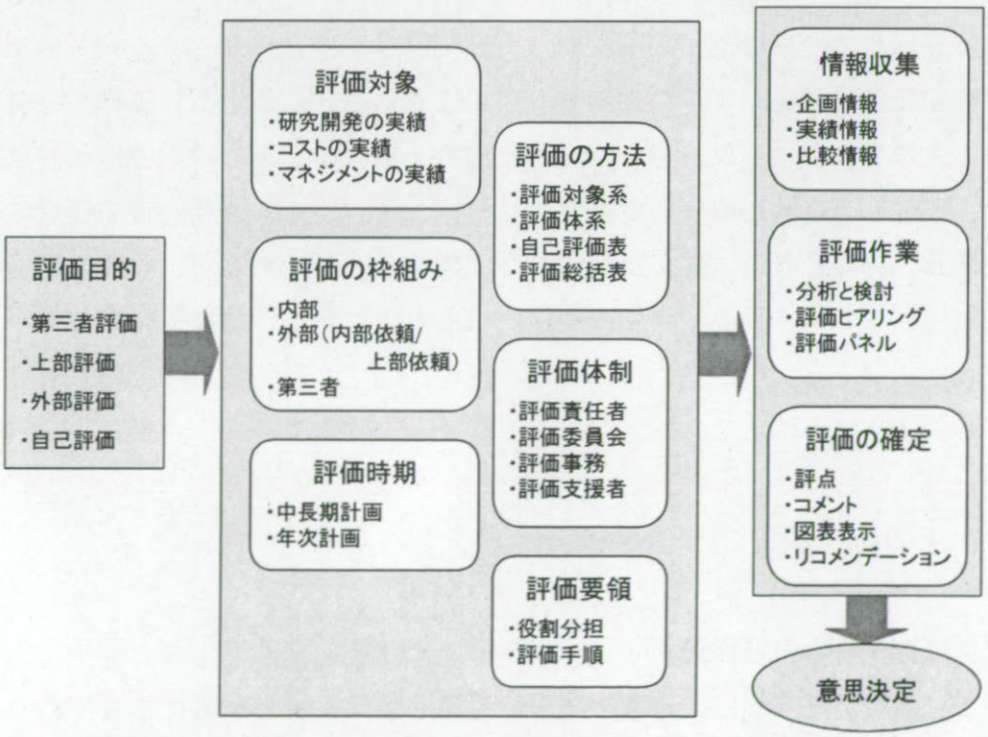

図6－3　機関を対象とした評価システム

た。評価対象は、途上評価で重要となる「内容の実績（成果とコスト）」と「方法の実績（マネジメントのあり方）」についてであり、評価の方法としては比較評価法（対計画比等）を評価要素毎に適用し、そのような個別評価データをパネル法で総合するのが常道である。

機関評価の質的向上を目指す場合、例えば特定領域に関しインタビュー調査等を深めケース評価法やレトロスコピック評価法等のロジック評価法を動員したり、各種メトリックスによるデータ分析を深め、類似機関とのベンチマークやランキング等の比較評価法により隠れたファクツを見出したり、あるいは機関全体のパフォーマンスをポートフォリオで表現し全体状況と各部署の状況とを明示するなどの多様な手法を適用し、多面的に認識を深める努力が必要である。

このような作業は機関内の評価事務局のみでは過重であり、外部の評価支援機関の援助を仰ぐべきである。ただし、我が国においてはそのような支援機関の充実が極めて遅れていて、信頼できる機関は皆無に等しい。その強化が喫緊の課題である。

最後に、上記のように認識を深めた後で、機関評価のまとめを行うこととなる。評価のまとめはおよそ2種類の内容から成っている。第一は評価そのもののまとめであり、ファクツと評価システムに則った比較的客観的な結論である。評価結果は通常、評定区分にしたがった評点とその背景的な説明に相当するコメントの文章で構成される。評点は単に項目ごとに記載されるだけではなく、レーダーチャートやポートフォリオ等の全体状況を示す図で表現されることもある。しかし、評価結果の骨子はあくまでも文章でまとめられるべきである。また、第二のまとめの内容は、評価者のリコメンデーションに関わる事項である。リコメンデーションは評価者の見識に基づくものであり、評価者の価値観が反映された主観的な主張であるが、前段の深い認識と思慮を前提としている以上、評価依頼機関はもとより評価対象機関の側でも十分に検討すべきものである。そして、その実施状況については評価者だけではなく、社会に対しても公表されるべきである。

6．4　その他の評価

（1）プログラムの追跡評価

プログラムの追跡評価は、政策や施策を見直す手段として非常に重要である。政策や施策の事前評価は次項で述べるように極めて困難な課題であるため、通常の評価のための分析の多くが、実効性の高いこのような評価局面に注がれているのが世界の実

情である。

プログラムの寿命は通常それほど短くはないので、プログラムレベルの評価は追跡評価より途上評価のケースの方が多い。途上評価の場合には、プログラムそのものを見直すことが主要な目的であるのに対して、追跡評価の場合では、そのプログラムを設定した政策それ自体を見直すことが主な目的となる。しかしいずれの場合であっても、新たにプログラムや政策を設定する際の教訓を得る事ができ、その種の知見の集積が新政策の形成にとって極めて有効である（図６－４）。

通常、プログラムの下で多くのプロジェクトが何年にも亘って展開され、既に終了しているプロジェクトが多く存在しているという状況が想定される。科学のフェーズで閉じているプログラムの場合は、そのような既存プロジェクトに対する計量文献分析法等のメトリックスを主要な方法論とする分析法が有効であり、技術以降のフェーズに係るプログラムの場合は、いずれにしても社会経済的な効果を把握する必要があり、計量経済分析法を中心としたメトリックスを分析の基盤的方法とすることになる。結局、プログラムの社会経済的なインパクトまでを分析するということは、その下で展開された個々のプロジェクトの社会経済的なインパクトまでを個別に分析し、そして、それらを足し合わせることに相当する。この作業は膨大であり、評価パネルを設

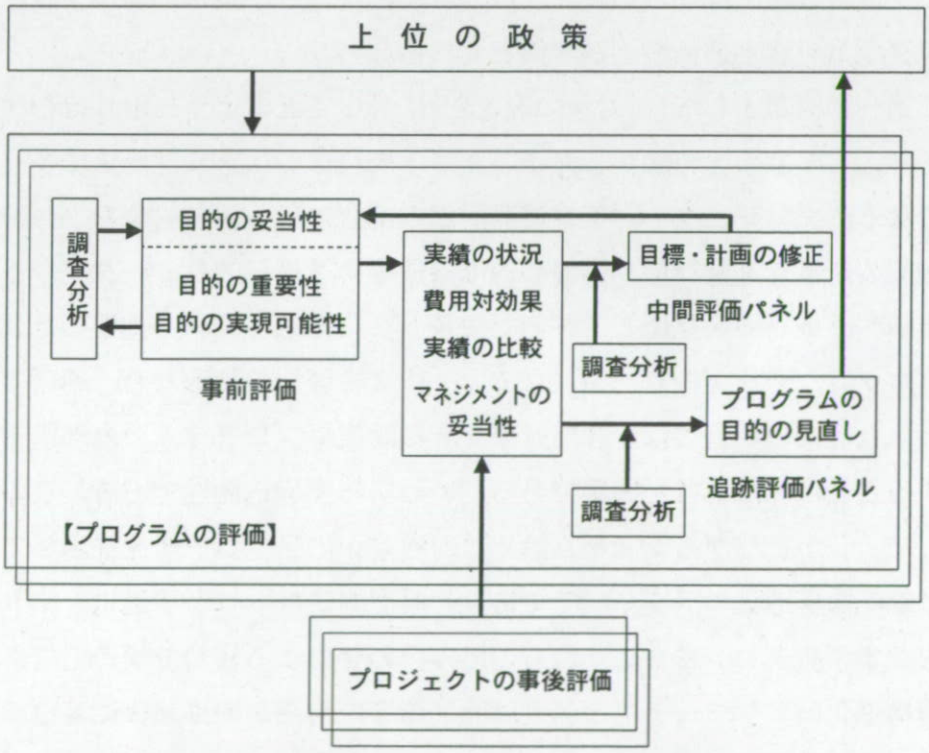

図6－4 プログラムの追跡評価と政策の見直し

定して直ちに評価結果を出せるものではなく、調査分析が作業の中心になる。逆に、プロジェクト結果の単純集計程度のデータで気づく教訓は表面的な現象論に過ぎない可能性が高い。実務的には、その調査分析を担う外部の支援機関や調査分析機関に作業を委託し、その分析結果を踏まえてパネルが設定されることとなる。図６－４にこの間の事情を示した。この種の調査分析は、評価対象とするプログラムの内容に大きく依存するため、定型的な方法論は存在しない。むしろ、事例を通して、多様な方法論の組み合わせ方や使いこなし方を修得すべきである。

プログラム評価の事例として、SRI インターナショナルが担当した NSF のエンジニアリング関係のプログラム評価について参考資料 3.2.1 に紹介する。また、DOE の大型のプログラム群に関する途上評価の事例を参考資料 3.2.2 に示す。また、テクノポリスが担当した事例を参考資料 3.2.3 に紹介する。

（２）政策の評価

政策の評価についてもプログラム評価と同種の階層構造を想定することが出来る。ただしこの場合は通常の政策の上位概念としては、基本政策や基本計画ないし総合戦略、あるいは省庁等の設置目的に規定された使命などが考えられる。この種の階層構造の中で、政策の途上評価や追跡評価は、対象の複雑さの飛躍的増加を除けば、基本的にプログラムの途上評価や追跡評価と変わらない。

そこで第一の問題として、「対象の複雑さ」に対してどのように取り組むべきかに関してまず考えてみよう。結論から先に言えば、それはシステム評価法に多くを依存することになる。プロジェクト評価や機関評価の場合には、評価対象をプロジェクトの構造や機関の組織構造及び職務内容の内部構造等に準拠してブレークダウンし、評価対象の要素や評価局面を見出して評価対象系としてそれら総体を把握した。しかしながら、プログラムや政策のレベルになると、特に論理化の薄弱な直感的動機で打ち出されたような施策の場合には、評価対象である政策やプログラムの内部構造を明確に構造化することは、はなはだ困難である。いわゆる悪構造問題である。このような対象に対して、しかしながらシステム論では何通りかのアプローチを提示している。第１には形態的な構造化のアプローチであり、典型的にはトリー構造化とマトリックス構造化を基本手法とし、さらにそれらの組み合わせによる複合的構造化である。第２には論理構造化のアプローチで、例えば因果律等の何らかの推移律に基づく論理構造化である。欧州の評価研修プログラムで推奨される「ロジックモデル」はこの原初的な例である。第３には機能構造化のアプローチで、要素としては実体そのものではなく

機能を担うエージェントを想定しその構造化モデルを運用する場合である。第４にはオートポイエティックなアプローチで、構造を既定のものとせず「行為」による「産出」の結果として構造化を想定する立場である。このようにして認識された「構造」に対して、構成する要素も、確定的な属性を示すだけではなく、不確定な幅や分布、あるいは確率的な振る舞いをするものとして想定される場合もある。システム評価法はこのように多様なモデルによる総合化を手法とする評価法である。実際には、インタビューの調査や各種メトリックスによる分析結果等を踏まえてシステム評価法が適用されるので、得意な方法論を有する異なる数名がチームを形成して調査、分析、評価に当たることになる。

第二の特徴的問題は、評価対象の特性値の大小がパフォーマンスの良否として検討されるだけではなく、システム自体の適否を検討すべき点である。このことは第一の悪構造問題であることとも関係するが、より本質的には意思的人間によって構成されている「人間活動システム」に係る知の特性に由来した特徴である。その原理的な話にはここでは深入りしないが、政策評価では、政策を体したシステムのシステム上の欠陥や弱点を指摘することの方が、その特性値のパフォーマンスの大小を検知することより本質的な評価であることを主張している。

第三に新たな問題として、さらに、最上位にある政策の評価をどのように取り扱うべきかについて考える必要がある。特に政策の「目的の妥当性」の根拠となる上位概念が原理的に存在しない場合、それに代わるいかなる根拠を求めればよいかについてである。例えば、科学技術共同体全体を対象とするような基本政策、端的には科学技術基本計画であるが、この種の場合には、関係者は社会全体に広がっているので、社会全体を巻き込んだ評価システムが構想され、政策の「内容的な妥当性」の他に「決定過程の正当性」が問われるべきである。そのような体制とは、フランスのギャランターシステムやイギリスのフォーサイトのようなものであり、関係者に対する全体性と将来に関する先見性が担保されなくてはならない。最近の例では、オーストラリアのイノベーションサミットをてこにした政策の見直し過程がその好例であり、この場合は最終的には国政選挙を通じて信任を得た。

政策の事前評価は、上記の途上評価や追跡評価より一段と困難な課題である。その理由は、「政策目的の妥当性」を担保するために必要となる先見性確保の困難さにある。将来を見通すことの困難さと過去を分析することの困難さとの比較から明らかなように、データの存在しない将来をいかに信頼性をもって把握できるかの困難さの程度にかかっている。特に、政策のような広範な適用範囲をもつ対象を先見的に把握す

るためには二重の困難さを克服しなくてはならない。そのための方法論は、既に記した通り本格的には企画を改めてまとめることとしたいが、その基盤となる考え方は参考資料１にまとめた。

しかしながら、実務的にはいくつかの試みがなされており、世界の潮流として、第一には最上位の基本政策を社会に開かれた政策形成システムによって形成し、その枠内で下位の政策を展開するか、あるいは第二のアプローチとして政策を施行しつつ見直しを繰り返し循環型の評価体制のなかで政策をブラッシュアップしていく方法等がとられている。前者の先駆的事例はイギリスのフォーサイトであり、また方法論として興味深いのはドイツで最近行われたフトゥア Futur である。後者の典型的事例はアメリカの GPRA である。

（３）独立プロジェクトの事前評価

独立プロジェクトの事前評価は実務的に重要な課題であるが、上位の政策的枠組みが存在しないため、やはり事前評価が困難な評価課題である。独立プロジェクトは通常大型の研究費を長期にわたって投入することになるプロジェクトである。このような大型の独立プロジェクトにも、何種類かのタイプの違うものがある。例えば、ITER のような統合された目標に向かって集中的に資源を投下していくケース、あるいは大きな予算を投下するが、研究自身は小さいユニットに別れて分散的に行うような分散型の大型研究もある。

分散型の場合には、まずその巨大な資源配分を行う枠組自体の設定に関する妥当性の評価が必要であるが、上位の枠組みのない状況で妥当性を確認するためには多くの他の可能性との比較等、膨大な作業を必要とする。またさらには、特に社会経済的な効果を含む目標を設定する場合には、その事前予測は極めて困難である。その期待される効果を予測するためには具体的な個別プロジェクトレベルにまで降りて評価する必要があるであろうが、対象の目標が分散しているので、個別ケースを積み上げることの困難さがある。

これに対して集中的な資源配分を行い大きな実験装置等を開発して研究を行う種類の大型研究の場合には、その対象とする装置の妥当性を含む、様々な角度からの評価が必要になる。この場合も、当該「目的の妥当性」を確認することの困難さは上記と同様に残るが、期待される成果の側の評価は、比較的定型的な手法を適用できる。つまりこの場合の評価手法は、科学的、技術的、経済的、社会的の４フェーズに分け、それぞれのフェーズについてフェーズ固有の方法論を具体的に適用し、分析を深めて

いけばよく、他のケースと比較すると評価の枠組みは単純である。また、技術ロードマップのような全体計画を事前に策定する必要があるが、そのような枠組みを整備した場合、実施段階に入ってから途上評価を繰り返し、計画を管理していくことも容易に行える。

参 考 資 料

参考資料１．方法論の基盤

参考資料２．ＭＯＴの枠組みと評価

参考資料３．事例

参考資料４．主要文献

参考資料５．欧米の評価研究者ネットワーク

参考資料１．　方法論の基盤

研究開発評価は、評価対象となる科学技術の他に、研究開発に携わる研究者や研究開発組織に関わる側面と、その評価を展開する思考世界の状況とを併せて扱う必要がある。このような組織現象や思考世界を扱う方法論は、実在する自然科学的な世界を対象とした方法論とは基本的に異なるものであり、そのため、自然科学や工学で自然や人工的な物理システムを対象にして思考を展開している通常の科学技術者が、同じ思考の枠組の中でしか、その研究開発評価という課題に取り組まない傾向があるとすれば、このことは非常に大きな欠陥を生むことになる。組織的な対象、組織現象や思考世界を扱うための方法論は、自然科学的な方法論とは基本的に異なる原理に根ざしており、そのことを理解する必要がある。ここでは、そのような通常の科学技術者が対象としている世界と異なる対象領域を扱うための方法論の基盤として重要なポイントを幾つか選択して解説しておきたい。このような対象を妥当な形で理解するためには、高等教育過程において勉学を積むことが必要になるが、通常我が国では、このようなカリキュラムを備えている高等教育機関はほとんど無く、従って多くの評価に携わる人たちは、本文で述べた評価の概要を理解しただけでは不十分であり、以下で述べるような思考の枠組を新たに備えることが本来望まれる。

1.1　認識論の転換

まず、認識論の転換についてである。意思的な人間、あるいは意思的な人間によって構成された組織や機関には、普遍的な内在原理が存在しないと考えるべきである。自然や人工的物理システムであるならば、それらの対象を貫く内在原理、いわゆる自然の摂理や科学的な法則等が存在していて、そのような内在原理を見つけ出し、あるいは、その内在原理を導き手として対象をより深く理解していくことができる。しかしながら、意思的な組織現象にはそのようなものが存在しているとはほとんど期待できないであろう。

例えば、質問紙によるアンケート調査を行ったとする。その調査結果は当然のことながら、調査対象に依存する。また、同じ調査対象を扱ったとしても、その回答率によって、当然調査結果は異なる。また、調査時点がずれると、それによってまた異なる結果になるということが通常経験される。質問紙が同じであるならば同じ回答が得られると考えるとすればむしろそれは荒唐無稽だと通常思われるであろう。しかしながら、その調査対象の意思的人間組織に普遍的な内在原理がもしあるとすれば、そして、その調査質問紙がその内在原理を掘り起こすような仕組みになっているとすれば、これは、いつどのような対象を取り上げても同じ結果が返ってくることになるはずである。経営学の研究者ですら、アンケート調査の結果、アンケート調査にもり込まれたモデルを是とする回答者数が多いというような現象を捉えて、そのモデルが正しいということが実証されたといったような論文を書いていたりするわけであるが、このようなことは論理的に間違えていると言わざる

を得ない。

しかしながら、組織論、あるいは人間学においても、何らかの普遍性が存在すると考える考え方もあり、それは西欧の学問体系の一角で常に唱えられている。それは神の概念と関係し、完全なる神が作った人間であるならば、その神の完全性を反映した何らかの普遍性が人間の中にも存在していると理解する。もちろん人間であっても、その人間を医学の対象、あるいは生理学の対象として捉えるような、つまり、自然システムとして捉える捉え方もあり、このような側面を捉えた場合には、なんらかの普遍性が存在している。また、自然システムとしての人間ではなく、何らかの論理性に基づいて思考を行う人間という側面を考えた場合には、その思考の中味にも関係するが、その限定的な思考の枠組の中で成立する部分的な普遍性、つまりある種の思考の公理系の中で成り立っている論理的に無矛盾な範囲で普遍性を想定することが出来る。しかしながら、意思的な存在、情緒等に支配されたり、あるいは個人的な価値観に左右されたりする場合の人間を想定してみるとそこにはもはや普遍性を期待することはほとんど不可能である。このような通常の人間、あるいはその人間の集団である組織や機関をどのように取り扱うべきか、つまり普遍的な内在原理が存在しない対象として人間や組織を認識したときに、それらをどのように取扱うべきかが新たな課題になる。このような課題に関して、例えば Peter Checkland は Soft System Methodology というアプローチを推奨しているが、さらに一般的に言うならば、システム論の展開の歴史の中で、「対象」に内在した妥当性を求めるのではなく、「行動」の合理性を求めることによって活動システムとして人間や組織を合理的に把握しようという立場が確立されてくる。このような認識論にたつ場合に、それでは何を手がかりとして論を進めることが出来るのであろうか。

1.2 メタ理論の構築

完全なる論理が存在しないとするならば、不完全ながら、しかし大まかには信頼できるような種類の、理論になる手前にある理論を我々は集積しておくことが妥当だと考えるべきであろう。そのような理論のことをメタ理論と呼ぶ。メタ理論は、大きな傾向や大枠の原理に関わる理論であって、個別には多くの例外が存在するが、大枠については信頼できるという種類の、理論になる前の理論である。

例えば、人間組織を変えていこうという場合に、多様なアプローチがあるが、構成者の意識変革を求めたり制度変革に取り組むことを考えたりする。しかし最も有効なのは、人間の構成を変えることである。それは一部であってもよいが、例えばリーダーであるとか、あるいはその組織のある部分に別の組織から人を移してくる等人間の構成を変える。この組織変革を起点とすることにより、その新たに導入され、また新たな構成になった組織は、その構成者の変化に合わせた意識の転換が起こり、その意識の転換により、それに見合った制度が出来、そして組織が変わっていくという道筋が容易に出来る。このように組織変革において、人の構成、それに続く意識の変化、そして制度や方法等の変化という一連の

流れが存在するというのがひとつのメタ理論の事例である。自然科学的なアナロジーで言うならば、これは自然科学的なアナロジーであって必ずしも適切ではないが、例えば砂場の砂山に上から水をかけたとすれば、その水は砂山の上から下に向かって流れていくけれども、どういう流路を通って流れるかということまでは言い当てることは不可能である。つまり、大きな傾向として、上から下には流れていくが、ミクロにどの流路を通るかまではとても議論できないということになる。このようなメタ理論は、対象に応じて様々なレベルで数多く存在していることが理解できる。

例えば、現在我が国において行政改革が行われたが、行政改革が行われると次にはその改革された組織形態に合わせた新しい運営方式、つまりマネジメントの課題が発生するであろう。次にはそのマネジメントの課題を担える人、人材の問題が顕在化するようになるであろうことを容易に想定できる。このような機構改革、マネジメント、それから人材、という課題の連鎖の存在を想定することがひとつのメタ理論になる。

1.3 論理化手法

メタ理論の背景には何らかの論理に支えられた思考が存在している。組織現象はまず現象論的な対象であるが、その現象論的な対象を我々は何らかの形で論理に置き換える必要がある。論理化されると、その論理的な演繹性を利用して、理論を展開することができるようになる。現象を完全に論理化することは難しいが、ある側面を論理化することは可能である。論理化の方法論も何通りかあるが、その代表的なものは“ロジカルシンキング”で紹介された MECE（Mutually Exclusive and Collectively Exhaustive：ミッシーと発音する）な組み合わせに依存するやり方である。MECE は、要素が互いに独立でその要素を足し合わせると全体になるような要素の組み合わせのことである。互いに独立で要素を合わせると全体になる、そのような要素の組み合わせを要素の「完全な組み」ともいう。例えば、組織現象を考える場合、その組織がどのようなやり方、制度、つまりシステムで運用されているかというシステムの側面、それからどういう人たちがそれを担っているのかというアクターの側面、それからまた、どのような課題を取扱っているかというコンテンツの側面、こういうシステム、アクター、コンテンツというのが第一次的な要素になると考えられる。この 3 つの要素の組み合わせ、システム・アクター・コンテンツは、互いに独立であり、また、その 3 つを合わせるとその対象である組織現象の静的な側面を全体として捉えることが出来ることになる。残された別の側面は、ダイナミックな側面である。このようなシステム・アクター・コンテンツという切り口では、ある組織現象のある時間的断面が捉えられたことになるが、その組織現象としての時系列の繋がりをあからさまな形で捉えたことにはならない。このように、スタティックな側面とダイナミックな側面、この両方を合わせるとさらに高次の全体像を把握することができるようになるわけである。

このような要素の組み合わせを多数作成し、その中の任意の二組を取り上げてその二組を軸とする平面を想定する。そうすると、その要素が互いに分割する平面上のセル（ハコ）

ができ上がることになる。その平面上に現象論として観察された事象をプロットしていく。次にプロットされたセルの位置付けを使って観察された現象を、今度はセルの特性に置き換えて議論をするということができるようになる。これが論理化のひとつの道筋ということになる。

2つの軸はこのようにマトリックスを構成する場合と、ツリー構造に展開する使い方も出来るわけである。つまりひとつの軸は階層を分ける軸と考え、もう一つの軸は各階層の平面を分割する軸と考える。マトリックス構造として利用するか、ツリー構造として利用するかは、その論理化したい対象の目的に合わせて選択すればよい。

あるいはまた、組織現象を何らかの論理化されたモデルに置き換えて考えるということも通常よく行われる。この場合も組織現象全体を映し出すことは出来ないが、ある重要な側面を論理化されたモデルとして描くことは有効である。組織現象の場合に、もしもその組織を構成している人そのものを要素とした、例えばAさんとかBさんとかという実体的な人を要素としたモデルを作ったとしても、それはその現象論の域からは抜け出ることができない。論理化するということは、例えばそのAさんやBさんが担っている機能、これは当然複数の機能が担われているであろうが、その機能を担っているエージェントをモデルの要素と考えるならば、そのエージェントを要素として機能モデルを作ることが出来る。

通常機能に置き換えることは非常に重要な意味を持っていて、それは、その機能を組み合わせることによって、物事を実現していく筋道を明確にすることができるからである。ニーズとかディマンドをターゲットとしてそれを実現していくことが、このような課題を扱うときに必要になるが、このニーズやディマンドを実現できる機能の組み合わせということを利用すれば、もしその機能によるモデルが出来ていれば、その機能の組み合わせによってアプローチをしていく道筋が見えてくることになる。もしもそれを機能エージェントに置き換えないで、Aさんや Bさんという実態モデルで表現している限り、これはいつまで経っても現象的な思考からは抜け出せないことになる。

1.4 全体性の確保

我々が思考世界を展開する時にしばしば陥る過ちは、自分の当面思考できる範囲に思考を限定して進めてしまうことである。これは全体ではなく部分を扱っていることに相当する。全体を見た上で判断しなくてはいけない事柄に対して、部分しか見ていないとすれば、当然誤った思考に陥ることになる。よくある過ちとして、部分最適化、つまり部分だけを認識してその限りにおける最適なものがあたかも全体を支配していると錯覚してしまうことである。

部分部分で最適化した解を足し合わせて全体の最適解になる現象は非常に例外的なケースであり、これはいわゆる要素がリニアな関係になっている時、つまり、全体と要素が単に足し算や引き算の形で表現できる時にのみ成り立つ。しかしながら、組織現象のような話では、複雑な絡みがあり、例えば要素同士がかけ算になっていたりする。そういう項が

ある場合ノン・リニアな関係になっているという。このような現象の場合、部分部分を認識していくとしても、ノン・リニアな構造がどのようになっているかを認識しておくことが先決であり、そのような構造が分かった後、部分ごとの認識がその構造に合わせて集計されるなら間違いない表現形式になる。しかし、ノン・リニアな構造を同定することは非常に困難な作業を伴う。

このような関係を認識したうえで、部分ではない全体を捉える簡単な捉え方についてまずみてみよう。これは、先ほどのMECEな組み合わせと全く同じであり、互いに独立で足し合わせると全体になるという要素の組み（要素の完全な組み）を導き、そのような要素からなる軸を多数用意して、そのうちの任意の2本を取り出して平面を作成する。少なくともこのような軸で区切られている平面である限り、限定された局面ではあるが全体を捉えていることになる。このような平面をたくさん用意して、色々な切り口で全体性を把握していく中で、全体の構造等が見えてくることになる。MECEな組み合わせの軸というのが、全体性を確保するための第一のアプローチということになる。

MECEな軸であっても互いに直交した2本の軸を見出せない場合，第二のアプローチに移ることになる。これはシステム思考によるアプローチであり、要素を完全な組みを成すようにもれなく集め，考察すべき場において互いに関係づけを行う。さらに何らかの推移律を用い、要素間の関係を構造化する。このようにして得た構造モデルを模擬し，実体を表現できるように修正を加える。要素の抽出、関係づけ、構造化、シミュレーション、そして前段のどこかに立ち戻って修正を行い、全体性を備えた実態モデルを形成していく。その際、要素は実体そのものの概念からはじめるにしても、モデルとして有効なのは機能概念による場合であることを銘記すべきであろう。

1.5 先見性の確保

全体を捉えることの困難さよりさらに困難なことは、先見性を確保することである。通常、対象の分析を行う場合、それは過去にその対象が演じた現象を把握して、その過去のデータを分析することになる。しかしながら、先見性は将来に対する見通しであるので、将来を分析しなくてはならない。残念ながら我々は将来のデータを手にすることはできない。しかしながら、工夫をすることによって将来を見通せる可能性もある。それは将来起こるであろうことの予兆を発見することであり、予兆を発掘しそれを予兆として捉えることである。

単純には過去の分析をして、過去の分析のトレンドの先に将来の姿があるとして予兆を見ることがよく行われている。しかし、これだけでは信頼性は高くない。信頼性のある予兆はどのようなものか。過去の現象であっても、それが不可逆的な現象であるならば、これは逆行しないわけであるから、その不可逆的な現象の将来に伸ばしたトレンドは信頼性がある。例えば、加齢は不可逆現象であるから人口統計は少なくとも10年くらいの単位で考えるならば、今までの人口統計、つまり過去の統計を使って10年程度先を予測すること

はそれ程困難ではない。つまり、死亡率とか出生率の変化はそれほど激しくは変化しないし、また、何よりもそこで今存在しているその人間集団が、一年ごとに加齢、年齢を重ねていくという、そういう不可逆的な現象であるからである。

不可逆的な現象の他に 2 番目には、論理性に依存することも十分信頼性があるやり方である。論理が備えている演繹的な展開力が予兆を発見する契機を与えてくれる。思考世界における先見性は論理性に依存することになるが，人・組織等に関わる実態的な論理性は先のメタ理論によることになる。先ほどの事例で言うと、例えばある組織の人の構成が変わったとすると、その組織はやがては意識が変わり、その組織の構成の内容に合わせたような方向に変化していくだろうということが予兆として捉えられるわけである。また、ある組織において構造改革が行われて組織構造が変わったとした場合に、その先には新たな課題として、その変わった組織に合わせたマネジメントの課題が生まれるに違いないし、また、その課題が困難な難しい問題であるとすると、それをどのような人材が担えるかという、その担える人材を養成しなくてはいけないといったような課題が新たに生まれてくるに違いないと予測できるわけである。このようなメタ理論による論理性は、いわゆる発展段階論に相当するものとして各所に存在しているが、間違えてはいけないのはその発展段階論のとおりに必ず変化するわけではなくて、あくまでもメタ理論が指し示している比較的信頼できる予兆であるに過ぎない点である。

3 番目にシステム論では、例えば先行指標が使われたりもする。同じ原理に基づいている中で先行的に変わっていく指標を捉えて後続する指標がどのようになるかということを予測し、先回りして考える。例えば軍事用の技術というのはやがて民間で利用されるようになるであろうから、軍事用で到達している技術のトレンドにある種のタイムラグを見越すと、民間で使えるようになるのにどの程度時間がかかり、どのような技術がいつ民間で使えるようになるかということが読めてくる。

1.6 形成的思考（システム思考・システム論）

通常、自然科学では分析的な思考が卓越している。対象を掘り下げていくことによって対象を深く認識する思考のトレンドがあるわけであるが、それに対して今我々が扱おうとしているような思考世界では、考えついている要素を汲み上げて、それで全体像を作っていくような、形成的な思考が必要とされるわけである。分析的な思考に慣れている者にとって 180 度思考の向きを変えるのは容易なことではなくて、その形成的な思考をやはり訓練する必要がある。その形成的な思考はシステム思考、あるいはそのシステム論によって裏打ちされている思考ということもできる。その詳細な内容をお話しするには、高等教育機関でのいくつかの講義と演習が必要である。1.4 節で示したアプローチは，その一例である。

1.7 組織知の形成

評価を組織的に取り組むことが要請されている。評価は決して評価システムを運用する担当者が独自に行えるものではなく、研究者やその研究者が提案するプロジェクトの内容が理解できる評価者や、あるいは社会経済的な側面が理解できるエキスパート、このような人たちの協力なしには評価は行えない。また、評価結果は組織的にまとめられた後、タックスペイヤーに対してそれが公開されるということになるであろう。このような評価を巡る組織的な対応が必要である。その場合に評価の方法や評価によって得られる知見が関係者によって共有できる形であることが必要となる。これは、とりもなおさず、研究者とタックスペイヤー双方から、あるいはまた評価に携わる評価者自身を含めて、お互いに信頼される評価システムであり、また評価結果であることが必要であることを意味している。

そのように、それぞれの立場から信頼され共有されるシステムや知を得るためには、組織的な対応、組織知を形成する仕方が必要になる。その組織知を形成する方法は、例えば、Soft System Methodology のような関係者がある思考プロセスを共有しつつ、行動において同じ方向を目指すような形でその組織がまとめられることが必要で、その評価システムを設計する段階から各関係者が組み込まれていて、当事者や様々な立場を代表している人たちによって、inclusive な形で評価方法等が形成されていくことが必須であるわけである。このような組織知を形成する方法論なしに評価システムを組織に定着させることは殆ど不可能であるし、また導入されたにしてもそれが有効に機能するとは思えない。

Soft System Methodology をはじめとした参加型の方法論に関しては、これもまたこの理解だけでもかなりの時間を要するであろうが、本格的な参考書に詳細な解説はゆずることとする。

参考資料２．　MOTの枠組みと評価

研究開発の評価は技術経営の重要な部分を占めている。評価は技術経営（MOT）の枠組みの中で位置付け、その一部として運用することが必要であり、またさらに大きな枠組みとしては、経営全体、つまり行政経営全体の中で技術経営を位置付け、そしてまた必要な評価をそれに組み込んでいくことが必要である。このような、評価の外側にある重要な概念を持たないで、単に評価だけを取り上げたとすれば、全体として有効な評価にはとてもならないであろうし、また評価結果を有効に活かすことも困難である。技術経営の概念も、また多岐に亘っているが、ここでは、その中で評価と直接係わる部分に関して、参考のためにまとめておきたい。

2.1　評価を巡るフェーズ概念

技術経営の枠組みの中で、研究開発の流れを位置付けるとすれば、それは研究開発のパイプラインと称する一続きのフェーズにおいてである。まず、最初は戦略のフェーズであって、研究開発を始めるにあたってその戦略を明確にする段階である。2番目は、その戦略に基づいて具体的な展開計画をつくること、つまりロードマップ等を整備する段階。3番目に、そのような戦略と計画の枠組みの中で、具体的にどのようなプロジェクト等を選ぶかという選択の段階。そして実際の研究開発を実施する段階がそれに続き、その結果をしかるべき受容者、成果を受け取る側に移転し、また受け取る側からは、受容のフェーズの後に利用結果の情報が研究実施側にフィードバックされる。

このそれぞれのフェーズの区切りのところで何らかの評価が必要となる。例えば最初の段階でいうならば、戦略としての妥当性に関する評価があり、2番目では計画の妥当性に関する評価があり、3番目の段階では、まさにどのプロジェクトが妥当であるかといった選択の評価がある。研究を実施した後にも、研究を続けている途上での評価があり、研究を実施した直後の評価がそれに続き、また受容者の受けとめ方等を考慮した追跡的な評価がさらにある。このように、各フェーズに関係して様々なタイプの評価が実施されなければならない。決して選択のフェーズだけの評価ではないことを認識すべきである。

このような研究開発のパイプラインを想定してみると、通常、研究開発評価の事前評価は、プロジェクトを選択するフェーズに相当し、また、途上評価や直後評価をより明確に行おうとすると、事前にどのような戦略や計画あるいは目的が策定されているかが明確になっていなくてはならないことが理解できる。従って、評価を明確に行おうとすれば、まず目標を明確にし、目標を明確にするためには計画が明確につくられていなくてはならないし、また良い計画をつくるためには妥当な戦略が策定されていなくてはいけない、というような一連のマネジメント作業に遡及的に焦点が移ってくる。

我が国では、技術経営の中で「評価」から取り組むことになったが、このようなパイプ

ラインを想定すると、やがては戦略形成をどのようにすべきかという、別種の新たなマネジメント課題と我々は取り組まなくてはならなくなると想定される。そのような戦略形成、あるいは計画へのブレークダウン等が妥当に行われて初めて、その枠組みの中で適切な評価が実施できることになる。

2.2 R&D のフェーズモデルとメカニズムモデル

研究開発を実施する段階においてもまたフェーズモデルを想定することができる。最も原初的にはリニアモデルと称される、基礎、応用、開発、つまり研究から開発へという流れの中で研究の部分を基礎と応用に分ける、つまり基礎研究、応用研究、それと開発というフェーズが単行的にリニアに移行していく、このようなモデルがある。しかしながら、このリニアモデルに従って研究開発を行うことの非効率性ないし不経済性は各所で認識されていて、もはやそのようなモデルに従うべきでないことは明確である。にもかかわらず、基礎研究のプログラムとか応用研究のプログラムというフェーズ毎のプログラムを設定することが通常行われている。その場合如何に評価を適切に行ったとしても、非常に効率の悪い運営になることは明確である。技術経営の研究開発のフェーズモデルをどのように採ればよいかの知識が、評価を実施する前の枠組みとして必要になることが理解できるであろう。

研究開発に限らず、多段階の意思決定問題を効率的に解くためには、最終段階から遡及的にそのプロセスを定めていくことが正解であることが、ベルマンの原理として定式化されているが、研究開発においても最終的なターゲットから、どのような研究開発が必要か、どのような技術や科学的知識が必要かということをブレークダウンし、それらに直接取り組むのが効率的なアプローチである。このようなアプローチは例えば基礎、応用、開発という 3 つのフェーズを想定したにしても、そのフェーズの間をリニアに移して行くわけではなくて、ノンリニアにそれらを組み合わせるということに相当する。総称するならばノンリニアモデルで考えるべきである。

我が国の場合には開発に相当する部分を「開発研究」と呼ぶ習慣がある。これを英語に直すときに、その概念をうまく一語で表現できる単語が無い。Development Research とか Developmental Research と言うにしても、これはリニアなモデルに慣れている欧米の研究者にはなかなか理解できない概念である。そもそも Research と Development は対極にある概念である。実は「開発研究」と言う概念は、日本企業がキャッチアップ体制の中でうまく機能させた概念である。キャッチアップ体制の下では、目標が明確である。そのため必要な技術者と研究者が同じ組織の中で一緒に研究開発に取り組むのが合理的であり、この方式を開発研究と呼んだ。言葉を換えて言えば研究段階と開発段階、あるいは基礎研究、応用研究、そして開発が垂直的に統合されたやり方で取り組むことである。ターゲットが明確であればこのようなノンリニアなアプローチが効率的である。

キャッチアップでない場合であっても研究開発のターゲットをまず定めた上で、それに

必要な組織を形成したり、あるいは必要な知識を部分的に外部から移転してくる等、ノンリニアなメカニズムで研究を進めることが効率的なアプローチであることは論をまたない。またそのターゲット自身が社会経済的なものではなく、科学技術的なものである場合もある。そのような研究の場合であっても、科学技術の体系の中で、どのようなフロンティアに挑戦するかが明確に位置付けられるならば、その知的フロンティアを明らかにしていくための道具立てが、科学技術の枠の中で工夫されることになるであろう。これも又ノンリニアなメカニズムである。

しかしながらまた、科学技術の枠の中の研究において、そのような目標を定めないで研究することが、思いがけない発見に繋がるいい方法であるということを唱える研究者もいる。これは豊富なリソースと十分な時間があるような環境条件の下で、そのようなゆったりとした研究があってもおかしくはないが、全体的な効率から言うと、それは遙かに非効率的なアプローチである。もちろん予期しない成果は、特に基礎科学の研究においては重要であるが、それは漫然と研究をしていて予期しないものが出てくるのではなくて、明確なターゲットに挑みながら、しかし予期しないものが出てくるのでその知的な価値がそれとして認識されるのである。このように研究開発のアプローチをどのように採ろうとしているかは、計画段階の評価、あるいは実績評価をする際のマネジメント側の評価として重要な項目になる。

2.3 機能概念による研究の設計

科学技術活動のような知的活動の成果は、社会に対して2通りの効果をもたらす。

第1は社会を知的に先導する機能であり、2番目は社会の中にある問題を解決するために寄与する機能である。公的資金による研究を考える場合、納税者は、主には自分の身の回りの課題を科学技術が解決してくれることを期待している。例えば経済的な繁栄、雇用の確保、競争力の確保、疾病の治療や健康の維持、環境問題の解決であるとか環境との調和の維持等、様々な課題がある。しかしながら研究者の側から考えると、社会に対する知的な先導的役割への寄与が自身の任務として大きいと、通常考えている。このような側面に関しても、社会は十分認知する必要がある。

社会を知的に先導するための成果は、科学技術の側から発想される必要がある。つまり、社会はどのような先導が起こるかを認識していないわけであるから、科学技術の側から発想することが必要であって、これを科学技術を種とするシーズ型のアプローチという。これに対して、社会の中の問題を解決することに対しては、問題の所在とか問題の内容等を深く捉えた上で科学技術のポテンシャルやあるいは新たな成果を適用して、それに対処するアプローチ、つまりこれは社会の中にあるニーズとかディマンドから発想していくアプローチが必要になる。

むずかしいことはディマンドが長期的であり、ずっと先のニーズとかディマンドを対象にするが、現下から基礎的な科学技術の研究を続けて行かないととてもそのような問題を

解決することには至れないというような、シーズまでのプロセスが非常に離れているようなケースであるとか、ディマンドが明確には確定できないがそのためには科学技術の基礎的なある特定領域の研究を集積させる必要がありそうな場合等である。このような長期的な課題に関しては、最終的なニーズとかディマンドの側から発想して、それを実現するための「機能」を中間段階で想定し、その機能を実現する研究開発に取り組むというような計画の立て方が妥当であると考えられている。機能概念で中間段階を設定することが、シーズとニーズの両面に研究を結びつけるために有効である。

2.4 ナショナル・イノベーション・システム

研究開発の上記のような特性を活かして国全体の研究開発が進められるように、研究開発システムが適切に設計されていることが必要である。またそのシステムの部分部分を担うプログラムが適切に他と連携を取って展開されていることが必要である。

先導的な知の創出を目的とするプログラムの場合には、そのクライテリア、つまり評価基準は「独創性」といった曖昧なものではなく、「知的フロンティアへの挑戦」というような、測り易いクライテリアを設定すべきである。また、社会の問題解決型のプログラムである場合には、真にターゲットとして重要であるとか、あるいは社会経済に対するインパクトが大きいことを評価基準にして選ぶべきである。多くの国の研究はその中間にあって、長期型とでも言うべきものであり、20年30年の歳月をかけて、最終的には社会経済的に大きなインパクトをもたらすであろうが、その原初的な基礎研究の段階を国が現在担っているといったような研究開発プログラムが多い。このようなプログラムの場合でも、そのプログラムを設定する際に、例えば社会経済ニーズをブレークダウンして認識される、「機能領域」をターゲットとするプログラムにするべきであって、科学技術領域をターゲットにもししたとすれば、これはシーズ型で研究開発を進めてしまうことになる。結局発散的な研究になり、成果が集中的に展開されないであろう。

さらに組織構成として、フェーズ毎の組織区分にするのではなく、基礎、応用、開発を担う人たちが一体になってプロジェクトを担えるような、垂直統合型の組織で研究開発が進められるような組織構成や、あるいはアライアンスが組まれなくてはならない。こういう仕組みがナショナル・イノベーション・システムの枠組みとして存在しているならば、それに見合った評価は非常に明確にでき、またその評価によって効率よい研究開発が展開できると考えられる。

参考資料３．　事例

3.1　評価システム

3.1.1　NASA における評価の事例

NASA には宇宙科学等を主題とした公募型の科学研究プログラムが多数設定されている。NASA の公募型科学研究のプログラムは、宇宙科学や地球科学分野をはじめとして、微小重力のもとでの生命科学や物質科学、あるいはその応用等の分野等にわたっている。行政組織としては、NASA 本部の「宇宙科学局」、「地球科学局」、「生命と微小重力に関する科学と応用局」、「機会均等プログラム局」、「人的資源と教育局」に該当プログラムが分布している。

これらのプログラムは、従来からピアレビュー方式によって評価すべきことが決められている。研究開発プロジェクトの評価に関して、OSTP と OMB はピアレビュー方式（メリットレビュー方式、パネル方式とも呼ばれる）の採用を推奨している。その経緯や論拠に関しては、（参考）コラム「ピアレビュー方式の検討」にまとめた。

（参考）ピアレビュー方式の検討

専門家相互による評価であるピアレビューは、歴史的に公募研究の選抜に広く使われ、また、それほど多くはないが R&D プログラムの評価やそれらの効果の評価にも使われてきた。

米国の省庁が実施する公募型研究プログラムでは、提案された研究テーマを外部有識者に評価してもらう専門家評価方式（ピアレビュー方式、メリットレビュー方式、パネル方式などと呼ばれる）が一般的であり、各政府機関のミッションに沿う必要が指摘されている（GAO: Federal Research"Peer Review Practices at Federal Science Agencies"1999）。これらの手続きを貫いている原則として、米国連邦法である連邦主題別法令集 United States Code に記載されている「調達契約制度における最大限の競争、グラント及び協力協定実施における競争の奨励」（Title31）の哲学が浸透している。政府資金利用の研究制度の根本精神は共有し、実務レベルではミッションに即して各省庁の独自のマネジメント・システムが工夫されている。

NRC 科学･工学･公共政策委員会の最近の報告では、連邦予算による研究プログラムの評価の最も効果的な方法は、専門家評価であると結論している（Evaluating Federal Research Programs: Research and the Government Performance and Results Act, 1999）。専門家評価の最もよく知られているものは、ピアレビューである。これには 2 つの側面があり、一つは研究がプログラムや機関のミッションに適合しているかどうかを評価する関連性評価と、もう一つは研究が科学的、技術的知識の先端にあるかどうかを評価する国際的ベンチマーキング評価である。ここでは、さらにフロンティアへのチャレンジの程度や科学的なインパクトの大きさが問われる。

ピアレビューに関する米国連邦政府の統一的な定義はない。OSTP ではピアレビューをメリット・ベースド・プロセスあるいは独立のメリット評価と定義している。各機関でピアレビューはやや異なった定義になっている。だが、すべての機関のピアレビューの定義や説明には、基本的に十分な科学技術的能力を保持し、また応募者と利害の関係のない個人による、科学的・技術的なメリットの評価がうたわれている。

評価者は一般的に応募研究者と同等の能力をもった専門的科学者あるいは工学者と考えられている。加えて、評価者には研究のメリットと目標との関連性について固有の評価ができなければならない。OSTP のマニュアルによれば評価者は解決できないほど大きな利害関係の対立が想定されない限り、資金助成機関の外部の人である必要はないと述べている。

OSTP によれば、政府機関がピアレビューを実施する必要があるかどうか、あるいはピアピアレビューをどうやって実施するか等を記述した、広く政府レベルの機関全体に適用されるべき規則や政策はない。だが、OSTP はピアレビューの利用と研究の質を評価するための「ピアレビュー利用の手引」を提供し、そのなかで OSTP は予算管理局（OMB）とともに、ピアレビューを推奨している。1996 年度の初めから、2 つの局は共同で、ピアレビューされていない研究よりもピアレビューされた研究への研究資金の増額を強調することで方向性を示した。その結果、OSTP は各機関がピアレビューされた研究により多くの資金を提供するようになったと述べている。また、「ピアレビュー利用の手引」に誘発されて研究評価の問題をどのように解決するかについて、多くの研究者から提案が寄せられるようになったと述べている。

OSTP は各機関のピアレビューの手順が柔軟で各機関のミッションや研究のタイプに合わせたものであるべきだと述べている。ある特定のやり方がすべての機関やすべての連邦によって助成される研究に適用されるべきでないと述べている。各種のピアレビューの方法が各機関で適切かつ本質的に検討されるべきだし、研究の性質や目的、研究のタイムラインの違い、R&D 遂行者の幅の広さ、そしてグラント、契約、共同協約といった研究資金提供のメカニズムの違いを反映すべきである。例えば、各機関はあるピアレビューを機関の評価、あるいはプログラムの成果に関する論文発表以前の研究結果の評価のような全く異なった目的に用いることができる。

ピアレビューは何人かの有識者によれば、質の評価基準が社会的、経済的、政治的なものより優位になる確固とした確信があると見なされている一方、他の有識者はピアレビューが経済的なことを考慮しなくなる科学のエリーティズムの本質であると見なしている。

NSF はある条件の下では、研究資金決定においてピアレビューの強調しすぎは、革新的研究への研究資金提供を阻害する可能性があると述べている。なぜなら、評価者は普通、やや保守的に研究を評価する傾向があるからと述べている。OSTP はこの現象は他の機関でも気にとめるべきであり、どのようにしたらピアレビューの手順がイノベーションをよりよく促進できるかについて、各機関を超えて努力するべきであると述べている。

1996 年 7 月、NSTC はピアレビューの導入における柔軟性の必要性について強調した報告書を作成した。そこでは、ピア評価に伴う各機関のメリット評価が、基礎的な科学プログラムの成果の評価のための一つの戦略と見なされると述べている。そのレポートで述べられた根本的な原則を、次に示す。

「科学機関は科学的プロセスの性質に適合し、国家目標のもとで基礎的な研究の役割を可能にする評価戦略を工夫しなければならない。戦略は驚きに応え、廻り道を許容するよ

うに設計されなければならない。また、国民の未来の福祉の向上にむかって新しい科学技術的な情報の創出機会を組み立て、そのプログラム・アジェンダを柔軟に変更できるようにしなくてはならない」

さらに、ピアレビューアの選抜に際し、分散化を図り、またNASAの関連5局のそれぞれやや異なる実施内容と要求内容を理解しそれに対応することも極めて重要なこととされている。5局のプログラム内容の詳細について、(参考) コラム「NASA 5局の関連プログラムとその位置付け」にまとめた。

(参考)「NASA 5局の関連プログラムとその位置付け」

(ⅰ) 機会均等プログラム局

「マイノリティのための大学における研究と教育プログラム (MUREPs)」は、議会と大統領の指示に基づき NASA では機会均等プログラム局に設定された。これは「歴史的黒人のためのカレッジと大学 (HBCUs)」、「ヒスパニックとヒスパニック・アメリカン支援協会 (HSIs)」、「少数民族カレッジ」への入学機会を増加させることを目的としている。

NASA の MUREP は科学技術の研究と、教育における少数民族協会の能力を強化するよう努めている。また、MUREP は主要な研究者を経常的あるいは公募型研究や、教育過程に関与させるようにしている。これらは NASA の仕事の多様性に寄与し、数学や科学・技術の支援による国家教育目標の実現をうながし、大学や大学院、あるいは高校レベルの先進的教育活動に寄与している。

(ⅱ) 人的資源と教育局

NASA の教育指導の方向性は、国家科学技術目標への政府機関の5つの寄与機関の一つとして、「NASA の戦略計画」に位置付けられている。そこでは、教育における優越性の確保と、教育界への包摂を優先することとしている。

こうした寄与は、教育界での活動や関連する種々の活動や努力に寄与する NASA の広範囲の教育プログラムの実施を通じて達成される。それは、参加者だけではなく NASA のミッションへの利益ともなる。

NASA の教育プログラムは、教育での優越性を支援する役割を果たすために、教師や学生を参加者やパートナーとしてミッションに組み込み、研究に参加させることを目的としている。なにをおいてもまず、教師や学生が、NASA の科学者や技術者、施設、研究開発活動に関与する機会を提供する。そうすることで、教師や学生は研究開発活動に参加することにより利益を得るし、かつ NASA の活動の広がりを理解させることができる。そして、教育界で分かち合うべき先進的な知識やスキルを教室に伝えることができる。NASA のミッションに組み込むことは、教育での優越性を向上させることに寄与する。教育プログラムの本質は、数学・科学・技術の領域で、顕在化した能力を発揮していない多様な人々をそこに導き参加させることにある。

(ⅲ) 宇宙科学局 (OSS)

NASA の宇宙科学プログラムはこれまでの間に十分進展し、また新たな展開が始まっている。科学的な知識との対話、物理・化学の基礎的な理解、あるいは太陽の星進化学的過程、地球の宇宙環境、惑星と、太陽系、星、銀河系、そして全体としての深宇宙に関わる

その他の現象の理解等。主要な科学分野は宇宙論、太陽物理、惑星間物理、宇宙物理、惑星の地質学と化学、他の惑星システムの研究、宇宙生物学が含まれる。OSS は、このプログラムの科学的能力を向上させ、アメリカの科学と技術のリーダーシップを確保するために新技術の開発と科学的発見や技術的進歩を社会に還元することを目指している。また、OSS は「機会公告」、「研究情報」に従ってプロポーザルの公募による先進教育／公衆プログラムを育成し、資金を提供している。

具体的に言うと、OSS は地上研究、コンピュータモデリング、基礎理論、航空機や気球や測鉛線ロケットによる軌道下の測定装置、シャトルによる低軌道上あるいは自動化された地球軌道や惑星間宇宙船による宇宙実験を含む活動を統合したものである。これらの活動は JPL を含む NASA のフィールドセンター、他の政府機関や施設、国内の宇宙企業を通じて、大学や他の研究組織の参加や支援により実施されている。加えて、多くの OSS のプログラムは、ヨーロッパや日本との国際的共同のもとで行われている。そして、ロシアやラテンアメリカの参加が増えつつある。これらの科学調査のすべてが、公式のプログラム広報に応じて提出されたプロポーザルの競争的なピアレビューによって選抜されている。

（iv）生命と微小重力の科学と応用局と宇宙飛行局

有人宇宙探査開発事業（HEDS 事業）は生命や微小重力科学を進展させるための宇宙あるいは関連する地上の実験装置を開発するプログラムである。そして、米国の産業や学界の適切な場所に、実際的な応用のためにこれらの成果である知識を移転することを目指している。

HEDS はヒトの健康や生産性向上のための生理学、NASA 職員の健康の理解に応用される。展開されているプログラムは、重力生物学と生態学プログラム、先進的生命維持技術プログラム、生物医学研究と対応策プログラム、微小重力研究プログラム、宇宙生産開発プログラム、航空宇宙医学プログラム、労働医学プログラムである。

加えて、HEDS は連邦政府の他省庁との連携プログラムであり、また世界の宇宙開発を実施している国々との連携プログラムでもある。

（v）地球科学局（OES）

地球科学プログラムは地球システムの全体を理解することと自然と人間によって引き起こされるグローバルな環境変化を理解することとを目指している。将来の世代のため、かつできる限りの科学的理解を基礎とした世界の貧困政策のために、地球環境を守り、向上させることである。宇宙からの観測のユニークな卓越性は、他の方法では観測できない地球表面、大気、氷、海等の情報を提供できることである。グローバル変化の研究と共に、OES はアメリカ企業の戦略的優位性を確実にするため、国レベル、国際レベルの複雑な政策決定への支援のために必要な理解を開発している。

OES は現在 5 つの科学研究テーマを実施している。エコシステムの生物学と生物地球化学とグローバルな炭素サイクル、グローバルな水とエネルギーサイクル、気候変化と予測、大気化学、物質地球化学、それに加えて地球科学の学際的研究も実施している。NASA のプログラムの役割は、環境研究のためのグローバルな観測システムの開発と運用、モデリングとその過程の理解である。OES の科学と技術のプログラムは、研究部門、応用と先進部門、プログラムプランニングと開発部門の 3 つの部門によって管理、遂行されている。

地球科学事業は宇宙船、他のリモートセンシングプラットフォーム、現場でデータを収

集する能力、収集されたデータや情報の管理、グローバルデータの処理、保存、配布、新しい地球システムの理解のためのデータ変換の研究や解析といった統合されたものから構成されている。

競争を基礎とした選抜、ピアレビューされたプロポーザルは、地球科学事業の科学プログラムを構成する基礎となるものである。

NASA の資金は組織の内外を問わず第一義的に公募型のプロポーザルを介して提供されている。その方式は、グラント、共同協約、委託契約、他機関との協定等多様であるが、公募情報は、「NASA 研究情報（NRAs）」、「機会公告（AOs）」、「共同協約公示（CANs）」、のメカニズムで流されている。応募された申請書は、ピアレビュー方式により審査される。

NASA ではピアレビューが科学的評価方式として位置付けられ、NASA 内外の専門家の協力によって行われている。ピアレビュー方式は、非公募型の提案や申請に対しても適用されている。ピアレビューの評価項目は、NASA の目的そのものと関係づけられて設定されている。また科学技術的なメリットを含む本質的な効用、研究者の能力と資格、そしてコスト等の項目を含む。

ピアレビュー制度は、NASA そのものの質と NASA と研究との関連性を向上させる唯一の手段と考えられている。この観点から NASA における全研究は実質的にピアレビューにかけられ、個々の申請内容のメリットが評定されている。NASA の「連邦調達規程補」（Federal Acquisition Regulation Supplement）1835.016 と 1872.403 に評価と研究の資金提供のための選定にピアレビュー方式を用いるべきことが定められている。またピアレビューに関するポリシーが、公募手順等を定めた公式文書の中にも述べられている。

NASA の個別のプログラム担当部局や適用領域の間で、実施されるピアレビュー方式はそれぞれ多少異なっている。例えば、コスト、適用される資源、あるいは受理されるプロポーザルの数やタイプ。評価委員会の形式も常置委員会、アドホック委員会、その混合等。しかし、当該領域の専門家による評価と、プロポーザルからの所見を記載することは NASA 全体のピアレビューを通して共通している。

ピアレビューとしては、当該領域の専門家である限り、その所属や国籍を問わない。ピアレビューア選抜のクライテリアは、その研究内容、発表論文等、知識と経験、そして公平なレビューを行う能力である。NASA とピアレビュー支援契約機関は、妥当なレビューアを特定するために、専門領域毎に専門家のデータベースを作成し常にその更新に努めている。領域の定評のある専門家や、かつての申請者からも質の高い他の専門家の推薦を受けている。このような外部のレビューアの協力により、新しい観点、代替的な視野、そして最高水準の理解を実現し得ている。

評価作業は資料郵送によるかパネル会議で行われている。両手段を使うこともある。資料郵送による場合は、非常に特化した科学技術上の専門性を持ったレビューアによる評価が期待できる。パネル会議では資料郵送された評価者間の評価の違いを調整したり、大規模な多面的内容を含む計画を評価するときなどに有効である。パネル・メンバーは 1 パネ

ル当たり通常 7～10 人程度であるが、15～20 人からなる大規模なものや、サブ・パネルを備える場合もある。非公募型の場合は、通常 NASA 内部のパネリストに資料郵送型で評価してもらう。書簡や電話等で評価を集約する場合は、当然評価会議への出席旅費は返還する。プロポーザルの申請者は当然評価には関与しないし、また直接競合関係にある申請の評価にも加わらない。全てのレビューアはその評価対象と利害関係がないことを確認して選抜される。

評価基準はプログラムの対象領域によって具体的には異なるが、常に基本的な 4 種の評価項目を踏まえて設定されている。NASA のミッションや目的との関連性、科学技術的な本質的効用、研究担当者の能力、そしてコスト。評価者は、通常 5 段階評価で評点をつけるが、言語による所見を常に付帯させる。この他に、NASA の当該プログラムの目的や期待される成果との関係や、提案されている予算の費用便益性も評価項目に加えられる。しかし、これらの評価項目の枠組みの中で、プログラム毎には様々なバリエーションがある。例えば、あるパネルでは、プログラムの目的との整合性に焦点を当てて、どの提案が最も適しているか、その序列を付けることに力点を置くのに対して、別のケースでは、その作業はパネル会議では行わないで NASA 本部の担当官に委ねる。NASA の担当官は、パネル会議からの評点にウェイト付けをして、プログラムの目的やコストと比較したり、科学的なリスクを考慮して採択提案の組み合わせがプログラムの中でうまくバランスがとれるように選択する。

3.1.2 英国 DETR（環境･交通･地域省）における評価の事例

英国環境･交通･地域省 DETR は 1 万 5,000 人のスタッフ（うち 3,900 人が中央の省に勤務）をかかえる政府機関である。昨年度の予算規模は 130 億ポンドであり、今年度は 155 億ポンドになる。うち 70 億ポンドが住宅と建設関連、45 億ポンドが交通運輸関連、35 億ポンドが環境ならびにエネルギー関連、1 億 3,000 万ポンドが地方整備施策関連、1 億 8,000 万ポンドが公衆衛生関連である。

そのうちの 1 部局に「建設局」がある。建築の質と投資効果を高めるために、職能訓練や雇用の改善、建設に関する技術開発や行政手続きの改良などにもあたっている。建設産業のイノベーションを推進し、「持続可能な建設」を実現するためにさまざまな研究プログラムを支援している。

DETR の 2000 年度年次報告書によると、建設産業の活動の質と活力をたかめることを目的にした研究開発に 2,600 万ポンドを助成した。さらに Partners in Technology（同省の助成プログラム）と LINK IDAC（同省と工学・物理科学研究会議が共同で行う助成プログラムのうちの 1 つ [注1] によって建設分野の研究開発を支援するために 1,200 万ポンドが追

注1 DETR と EPSRC（工学・物理科学研究会議）は LINK と名づけた 2 種類の研究助成プログラムを持っている。一つは IDAC（Integrating Design and Construction）で、もう一つが MCNS(Meeting Client's Needs Through Standardisation) である。建設分野での産学協同を促進するための助成である。

加された。

同省は建設事業計画 Construction Business Plans を作成し、政府とアカデミズムと建設業界の新たなパートナーシップのもとに企業の活動と収益性を高めるようプロジェクトを計画し、実行している[注2]。このプログラムに沿って同省は Partners in Innovation (PII) という研究公募のスキームを実施している。その公募研究プログラムは Construction Research and Innovation Business Programme と呼ばれている。公募された提案（プロポーザル）は専門評価者チームによって選定される。この公募プログラムは、国内のみならず国外の建設会社、建設関連業界、研究開発機関、大学に対しても開かれている。採用されれば、そのプログラムの費用の半分までを政府が負担することになる。

またこの助成プログラムの枠内に Construction Best Practice Programme というプログラムがあるが、これは企業に対して経営活動の最適化の意義を知らせ、その実現を支援し、そうした仕事にあたるリーダーを養成することを目的にしている。

1998 年には DETR の行政官と RMC（Research Management Contractor: 契約プロジェクトの運営担当官）だけが事前評価者となり応募されたすべての提案（プロポーザル）を判定していた。しかし 1999 年からは外部から評価者を入れ、第 1 段階の評価（アウトラインの提案でアイデアの重要性を判別）と第 2 段階の評価（全文の提案で中身を判定）にあたらせた。

その理由は、

1) PII のスキームがより広範囲からの応募が可能になるものに変わったため、応募された提案を評価するには幅の広い専門知識が必要になったから

2) 提案の質や新奇性を高めるには、第 2 段階での競争をより強化した方がよいと判断されたため、技術的な詳細を判断できる評価者がこれまで以上に必要になったからである。

英国の「環境・交通・地域省（DETR）」の Construction Research and Innovation Program は、「内部担当による全過程方式」（UK モデル）を原則とする英国では、例外的な外部支援方式により運営されている。その理由は、このプログラムへの応募数が毎年 600 件程度ありきわめて多数であることと、ニーズ型プログラムであるため、パネリストが多岐にわたり、その選定や評価基準のあり方等のマネジメントの負荷が大きいことによる。

プログラムは 5 分野に分けられている。その内の 1 つである Technology & Performance 分野を例にすると、この分野は 80 件程度の応募があり、各プロポーザル毎に 4～5 人（内部 2、外部 2～3 人程度）の査読評価者を配置する。内部からのパネリストは 1 人 2 件程度をニーズ側主担当者として担当し、外部からのパネリストは 1 人 8～10 件程度をシーズ側主担当者として担当する。この分野では、内部からのパネリストは 45 人にのぼり、外部からは 8～10 人となる。UH とアシスタントは担当を分割し、各 40 件程度ずつを担当する。

外部からの評価者は、大学、産業界、そして数は多くないが RC 等に関わるポリシーメー

注2　この Plans は 5 つの部門に分かれている。sustainable construction, safety and health buildings, technology and performance, construction process, best practice の 5 つである。

カー等から選ばれる。総数 35 人程度必要となるが、リストにはその数倍が必要である。外部パネリストの最終決定も UH とアシスタントが行う。

評価プロセスは評価パネルにより 2 段階で絞られ、最終決定する。第 1 段階では、アイデア、ポテンシャル、インパクトを主な評価項目として 1/3 程度に絞り、第 2 段階では、方法論やアプローチの適正さ、チームの構成、保有リソースの状況等を考慮し、さらに 1/3 程度に絞る。最終的には 70～80 件程度となる。第 1 段階のプロポーザルは 3～4 ページ程度のものであるが、第 2 段階ではフル・プロポーザルが必要で、第 1 段階での評価パネリストのコメントを踏まえ、政策目的に適合するよう、さらに改善されたものが求められる。この改善の度合いも評価項目として重要となる。評点は、各プロポーザル毎に集計され、分野毎にランキングをつける。評点分布の中間部分を占めるプロポーザルを中心にして、分野毎に全体パネルで検討し、選択候補を決める。評点リストと個別コメントにこの分野毎の全体会議でのコメントを付して、行政内部の担当プログラム責任者である PD に送り、PD が最終的に決定する。

3.1.3 米国メイン州における評価の事例

メイン州では「科学技術計画」の策定や同州にとってのプライオリティ設定等が行われている。また、その枠組みの中で、メイン州科学技術財団が補助する公募プログラムへの提案プロジェクトを評価選定する作業も実施されている。

このプログラムは大きいものではなく、年間 30～40 件程度のプロポーザルの中から 10～15 件程度を選定するというものである。このプログラムは、メイン州の科学技術政策の中心に据えられ、プライオリティ設定に従ってそれに適合したプロジェクトを選定する、という一体的な枠組みになっている。

RCS のチームは、州の側のいくつかの担当者グループと共同で、科学技術分野での州の政策的な優先度を決める作業にあたった。

州の「科学技術計画」を策定する作業において 1999 年には、メイン州が研究開発能力の強化と生産性の向上のために投資する Challenge Grand（約 1,500 万ドル）への研究公募の選定作業が行われた。29 件のプロポーザルが受理され（総計 5,800 万ドル相当）、この中から 9 つに対して実際の助成が決まった。AAAS のチームはメリット・レビューならびに社会的インパクト評価の観点からの選定作業にあたった。採用された公募の関連分野は、水産養殖、バイオテクノロジー、海洋生物学、分子細胞生物学、合成素材、情報技術、高度農業技術であった。

プログラムは、ミッション･オリエンテッドな性格を持ち、メイン州のニーズに直結した課題への取り組みが期待されている。

評価プロセスは、やはり 2 段階になっているが、第 1 段階ではテクニカルメリットを評価することが目的で、各提案当たり 2 人の評価者がそれぞれの評価に当たる。テクニカルメリットに関しては、約 50 項目程度のクライテリアを各々5 段階で評価する。分野が、海

洋開発や森林開発等、多岐にわたっているため、第 1 段階の評点によるランキングはしない。また、件数も少ないので、提案全てを第 2 段階に送る。第 2 段階での評価は、メイン州にとってのインパクトや行政課題との関連性を中心にして評価する。この評価者は外部からの識者 3 名で、3 名が全ての提案に目を通し、各提案 20 分ずつの質疑応答 Q&A を 2 日間に渡って行う。提案内容についての説明は、プロポーザルにあるので省略し、第 2 段階評価者の目から見た問題点についてただす作業が中心となる。第 2 段階の評価では、各提案を総体として 3 カテゴリーのいずれかに評定する。3 カテゴリーとは、それぞれ「強く推薦する」、「予算があれば採用する」、「推薦しない」であり、その結果と予算額を勘案し、財団内部で採否の最終決定を行う。その際、行政課題との関連性について考慮し、分野間の分布に配慮する。

3.1.4 米国ATP（先端技術プログラム）における評価の事例

① 概要－商業化段階以前にある産業技術を対象としたプロジェクト選定モデル

先端技術プログラム ATP（Advanced Technology Program）は 1988 年に議会によって創設され大統領の承認を受けて発足した技術開発支援制度である。このプログラムの主たる目標は、商業化段階以前にある技術、つまり潜在的なインパクトは大きいが企業セクターが自前で投資するにはリスクが大きすぎるとみなされる技術開発を支援し育成していくことである。利益を生み出す産業に的を絞ってなされる政府の研究開発プログラムとしては ATP は最初のものである。他の政府機関や研究開発プログラムでは、研究開発は公益的な機関のミッションやそのプログラムの担当部局の目標が達成されることを支援しなければならないものであり、そのことを通じて公益性のある財の増進と市場への波及効果を生み出すことをねらっている。ATP は基礎研究（公的な資金に大きく支えられている）と開発（主に私企業によって担われている）の間で一般的にみられる出資面でのギャップの問題と取り組むために設立された。プロジェクト助成の最大半分まで政府が出資するというコストの共有化がはかられた支援が特徴である。

商務省 DOC・技術標準研究所 NIST の中の ATP 担当室が“独立機関”として管理運営に当たっている。担当室スタッフは 80～90 人（うち約 10 人が経済担当）である。

発足当初から ATP は政策論争の的になってきた。何人かの下院の有力議員を含む保守派の国会議員たちは、ATP は政府の単なる補助金制度に過ぎないとみなしている。1995 年から毎年、下院は ATP を廃止すべきだという法案を提出している。ATP プログラムが存続しているのは上院とこれまでの行政官の力のおかげである。しかし保守派の議員の抵抗にあって、ATP プログラムは当初計画された規模には至らず、最近ではかなり規模が縮小された。2001 会計年度での予算は 1 億 4,570 万ドル（ピーク時は 1995 年度の 3 億 4,000 万ドル）である。ブッシュ政権下では、ATP の価値を再考するために 2002 年度の予算は計上されていない。

最初の数年間、ATP プログラムではプロジェクトの選定をどうするかが仕事の中心であった。技術的専門家に対して NIST、プロジェクトの担当機関、他の政府の研究機関などを通して接触がはかれるとの自信はあったのだが、それと同じく重要なのが、技術が商業的に生き延びていけるかどうかを評価する必要があるということである。すなわちビジネス計画が大切になる。これは連邦政府の研究開発プログラムにとっては新しい課題であった。過去 10 年にわたって、ATP の評価にあたった人々の大半は、ATP のプロジェクト選択過程に対し肯定的な見方をしている。

これまでの 10 年間では、2 つのタイプの助成制度が使われてきた。1)「重点領域研究」として知られている、個別重点領域ごとに公募し助成する、2)「一般研究」と呼ばれる、領域も定めないで公募する助成方式、である。

② 重点領域研究

1990 年代の半ばに至ると ATP の担当者たちは、特定のテーマを決めて助成対象の選択にあたったほうが、研究開発活動の間によりよい相互影響をもたらすし、技術もよりよく拡散していくかもしれない、と感じるようになっていた。加えて、1990 年代初めの頃は ATP の規模も拡大していたので、「重点領域研究」を導入することで多数のプロジェクトやそれに応募された多数のプロポーザルをより効果的に管理できるようになったのである。

「重点領域研究」で扱われたテーマはおおむね産業関連団体から提案されたものだった。それぞれのテーマについて、ATP の支援がそうした分野に対してなされることが適当か否かを、研究者の側から評価を加えてもらうために、公開のワークショップが開かれた。そして、研究テーマ選定の最終的な責任は ATP の長官にあった。

1994 年から 1998 年の間に ATP が扱った「重点領域研究」は次のとおり。

- 適応的学習システム
- 触媒ならびに生物触媒技術
- デジタルデータ記憶装置
- 情報ネットワークにおけるデジタルビデオ
- 健康管理での情報インフラ
- 複合構造体の製造
- 重工業のための材料加工技術
- 電子技術の製造インフラ
- モーター駆動車輌の製造技術
- 光技術製品の製造技術
- プレミアム・パワー（携帯、移動体、隔絶サイト等の用途のための新高度電源）
- 選択的透過膜の基盤技術
- 製造過程への応用を目指した統合技術
- 薄膜組織工学（生体膜の臨床応用等）
- DNA 診断装置

・気体圧縮冷凍技術

「重点領域研究」はこの時期の助成の大部分を占めていたが、「一般研究」の公募も同様に続けられた。1999 年には予算の減額のためもあって、ATP の運営は「重点領域研究」を廃止して「一般研究」の公募だけに絞られることになった。2001 年会計年度では、締め切り日を決めた一斉公募でなく、プロジェクトの見直しスケジュールに合わせて応募を受け入れることを始めた。

③ プロジェクトの選定

ATP は技術とビジネスの専門知を注意深く組み合わせることをねらった評価システムを開発した。プロジェクト選定過程を図式化したのが、下図である。2000 年には評価過程でいくつかの“関門”を設けて、さらに先の評価を受ける前に通過しておかねばならない評価点を明示した。

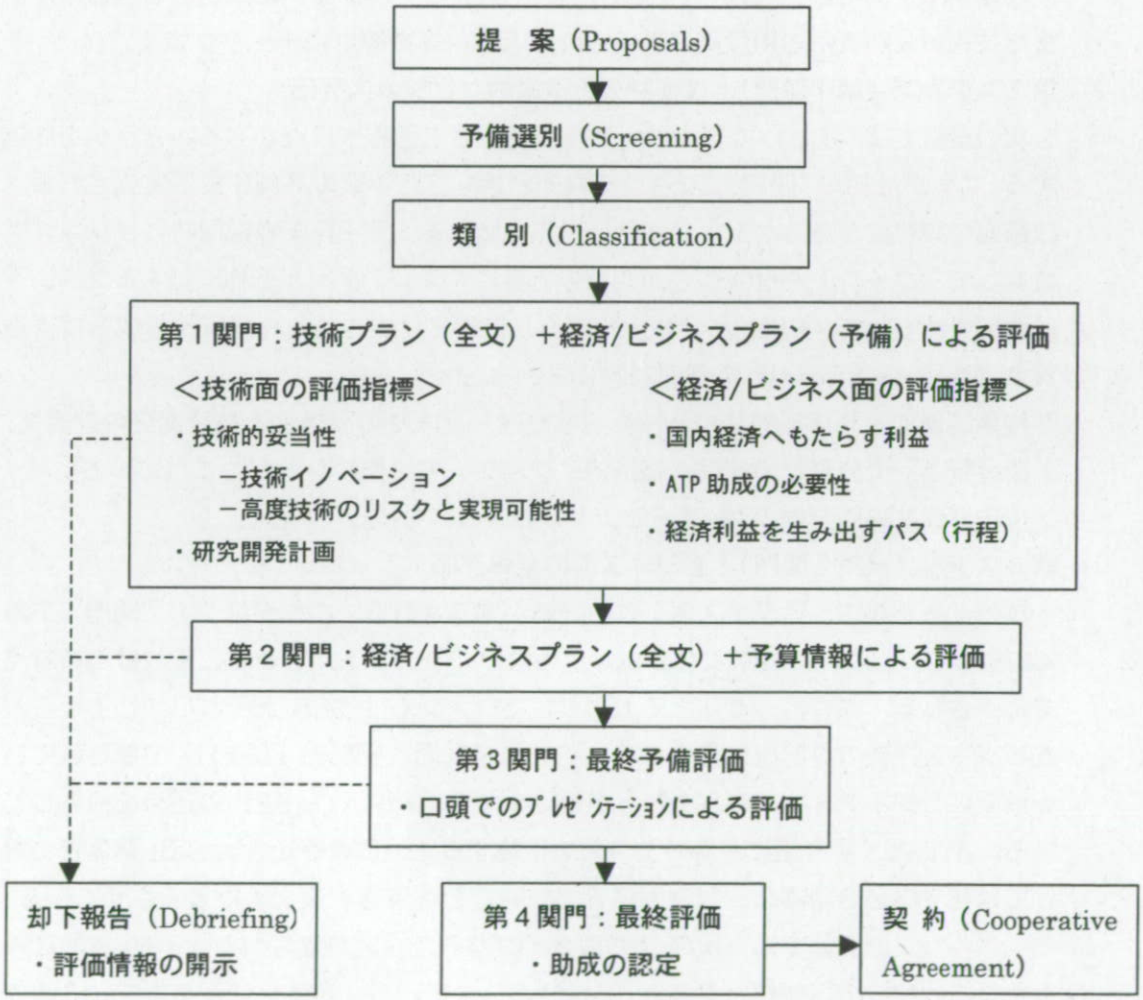

図　ATP のプロジェクト選定過程

1) **第1ステップ：公募評価ボード（公募資料評価外部委員会 SEB：Source Evaluation Board）の形成とランキング**

SEB は公的セクターと企業セクターの両方から選ばれた技術分野とビジネスの専門家グループで構成される。技術専門家は多くの場合 NIST もしくはいずれかの国立研究所から選ばれる。国防総省と NIH の研究者が採用されることはよくある。ビジネスの専門家は、定年後の重役経験者、ベンチャー企業への投資家、コンサルタント、大学の経営学の専門家などから選ばれる。審査員にはプロポーザル審査過程での得られた情報の秘密厳守が義務づけられている。SEB はまず予備的な評点によるランキングと分類を行う。

ATP は必要に応じてそれ以外に技術とビジネス両面の評価専門家を SEB の評価者として補充することができる。SEB の評価過程を補佐する運営コンサルタントとして、外部のビジネス・コンサルタントを ATP は常時雇用している。

このステップでは、プロポーザルの内容が ATP のミッションにふさわしいものか、また技術の成熟度が応用段階にあるものかどうか等を簡単にチェックする。

2) **第2ステップ（第1関門）：技術評価と予備的なビジネス評価**

次の段階では、主としてプロポーザルが技術面で見るべき点があるかどうかを評価する。これが最初の“関門”である。それぞれ独立した複数の評価作業が実施される。技術面での利点があるかないかが評価の焦点ではあるが、評価者は同時にビジネス面に対してもコメントを付すことが求められる。プロポーザルとそれに対するコメントは全部集められて、SEB メンバー及びそれ以外に付加された専門家から成る技術委員会（Technical Board）の検討材料にされる。

技術評価での評価指標には、技術の革新性、リスクの妥当性、技術開発の計画性、上流分野と下流分野との連関、経済的インパクトの可能性などが含まれている。

技術評価は評価全体の 30 パーセント分のスコアを持っている。

3) **第3ステップ（第2関門）：ビジネス面の全体評価**

技術評価を経て、ビジネス面全体の評価が実施される。これが第 2 の“関門”である。もしこのビジネス評価に値しないようなプロジェクトがあれば、それはこの時点で却下される。プロポーザルとそれに付したコメントは SEB とそれ以外に加えられたビジネス評価の専門家に送られる。ATP には期間限定契約（180 日）で雇われている前記のビジネス・コンサルタントが常時 8 から 10 名いて、SEB の運営を補佐している。ATP は、所有権にかかわる情報を保護するために彼等を含め SEB 関係者に対しては秘守協定を結ばせ、利害関係者でないことを確認する文書に署名させている。

この時点での評価では、次の 2 つのカテゴリーで査定がなされる。 i)広域的に考えた経済的利益の可能性（全体の 20%分のスコア）、 ii)商業化に至るまでの計画の妥当性（全体の 20%分のスコア）、である。

4) **第4ステップ：SEB の全体評価**

SEB は技術評価とビジネス評価の他に、いくつか他の点をも考慮に入れて総合し、

最終選考の一つ前の段階（最終予備評価）のための評価リストを作成する。ここで言う“他の点”とは、ｉ)提案者がどの程度責任をもってかかわることができるのかという点と提案内容が構成組織とどれくらい見合っているかという点（これで 20%分のスコア)、ii)提案者が所属する機関がどれくらいの経験と実績や資格をもっているかという点（これで 10%分のスコア)、の 2 つの評価指標である。

このうちｉ)の評価指標については次のような点が判断されることになる。企業が当該のプロジェクトにどれくらいの優先度をもたせているか、企業のスタッフの質、成功をもたらすための十分なスキルを確保するために当該企業において他部門との密な連携を持っているかどうか、中小企業の参画、そして経営計画とその実行へのかかわり、である。

5) **第 5 ステップ（第 3 関門）：最終選考の前段階での口頭発表による評価**

SEB は候補者を最も高い評価を受けた一群のプロポーザルに絞り、NIST でそれぞれの提案者による口頭発表の機会を設定して評価作業を実施する。これが第 3 の“関門”(最終予備評価) である。口頭発表の数は多くても 1 グループ 4 人か 5 人である。SEB は最高の質を備えた一群のプロポーザルを助成対象候補として推薦する。一般的に言って、この口頭での評価の段階まで残るのは提出されたプロポーザルの約 15%である。

6) **第 6 ステップ（第 4 関門）：最終選考**

最終選考は NIST の事務官で構成された「公募選考委員会」(公募資料選考内部委員会：Source Selecting Official) が担当する。ATP の長官と上級委員たちが最終決定に関する責任を負う。多くの場合、口頭での評価に至ったプロポーザルの約 3 分の 2 が助成を受けることになる。この助成認定が最後の“関門”である。

④プロジェクトの中間・事後評価

採択の後、助成対象先は、補助金支給開始時、進行中には 2 年に 1 回、そして終了時に聞き取り調査をうける。各企業は聞き取り調査の回答義務がある。

なお、審査員が年 1 回、参加企業を現地視察し評価する。また、プロジェクト参加者への調査等を通じて、リーダーシップや管理、参加企業や組織に対する効果、プロジェクトからの技術・知識の移転、プロジェクトの結果や困難だったこと等の分析を行う。

⑤ ケーススタディ調査

終了プロジェクトから対象を選択し、大学研究者や経済学者などと委託契約して（これまでは主に経済的インパクトに焦点を置いた) 分析・将来予測を行う調査を実施している。参加企業はこのような調査にも協力の義務がある。

1992 年以降の ATP 資金供与プロジェクト対象には追跡調査(民間委託)が制度化された。社会経済影響、技術進歩・実用化進展普及状況やスピルオーバー効果の調査が活発化して、評価研究にも貢献している。

⑥ 比較対照調査

ATP に参加していない同業種・同規模の企業と比較した、プログラムの企業・米国経済に対する影響調査も委託実施している。

⑦ プログラム全体に対する評価

ATP の評価は激しい政治的な争点ともなり、例外的とでもいえる多様な評価調査研究にも晒されてきた。従って、ATP の実施部署でも活発な評価とその委託調査を推進してきた。すなわち、ATP は当初からプログラム評価に力点が置かれ、内部スタッフに加え、学会、コンサルタント企業のエコノミスト、評価専門家の活用（評価計画、評価モデル化、評価データベース開発、評価調査、評価のケーススタディ、評価の統計的経済的分析、成果の尺度の質や信頼性向上に寄与する検討などが行われてきた。

現在の多元的な評価システムは Multi-component Strategy for ATP Program Evaluation と称されている。評価結果を短期に必要とするという緊急な要請と、経験的に長期にわたるプログラムのアウトカムを認識し価値づけするという忍耐を要する現実とを適合させるために、ATP はそのプロジェクト評価に Multi-component Strategy を採用した（Rosaile T. Ruegg, Director, Economic Assessment Office, ATP ）とされる。その構成調査研究を下に示す。

a. portfolio profiling　b. interium progress measures　c. real-time project monitoring
d. status reports on completed projects　e. micro- and macro- economic case studies
f. development of evaluation models, tools, and techniques　g. "special issue" studies
h. econometric and statistical analyses

このように、内部評価も、ATP の経済担当者、NIST 別局の経済担当者により実施されているが、内部「委託」評価の契約研究（調査内容は通常綿密に ATP 事務局と連携調整）が極めて活発に行われてきたことが特徴的である。

とくに経済社会的な視点からの評価は、手法に乏しく信頼性がなく、賛否両論がある。ATP では評価手法自体の開発に資金供給も行っている。

1993 年に成立した GPRA 対応で 1999 年から省庁自身が自らの業績について評価義務を負うことになったことから、ATP の評価結果が ATP そのものの存続に関わることになった。省庁にとってはプログラムの継続のためにはプログラムの有効性がと実証されることが必要になったわけである。

なお、ATP の「外部評価」としては、国家経済研究局 National Bureau of Economic Research をはじめ、大学や独立研究者によるものがあり、また、会計検査院 Generation Accounting Office も数回評価を実施している。また、議会の要請により NSF が ATP の包括的なプログラム評価を現在実施中であり、2001 年秋までには報告書が出される予定とのことである。

3.1.5 米国 DARPA におけるプロジェクト運営評価の事例

DARPA（高等防衛研究プロジェクト局）のプログラムのプロジェクト評価は、主として国立研究所の研究者からなる評価パネルが行う。しかし、プログラム・マネジャーは、NSF などと異なり「強いプログラム・マネジャー」であり、大きな裁量権があるため、評価パネルの判断を覆して決定を行うことがある。DARPA が委託する研究では長期にわたる資金提供が必要であるため、プロジェクトの選定評価時もさることながら、途上評価がより重要となり、そのため SETA と呼ばれる支援スタッフを契約により“調達”し、情報収集や評価を含むマネジメント支援を行う方式がとられている。

① DARPA

高等防衛研究プロジェクト局（DARPA）は、旧ソ連が人工衛星スプートニクを飛ばしたことに対する最初の対抗策として 1951 年に創設された。以来 DARPA の使命は米国が軍事力を強化するために最先端の科学技術を応用することを通して世界をリードできるようにすることであった。より正確に言うなら、想像力に富み、革新的で、しばしばリスクを伴うような研究上のアイデアを発展させて、それによって通常の研究の進展によってもたらされる以上の重要な技術的インパクトをもたらすことが、DARPA の科学技術面での使命である。つまり、原型となるシステムを発展させて実際にそれが使えるかどうかを試すことを通して、そうした新しいアイデアの適否を追求するというのが DARPA の科学技術面での仕事である。

DARPA では高度に創造的で画期的な研究に出資することが最優先の任務である。以下に示すように、DARPA ではプログラム・マネジャー（プログラム・ディレクターともいう）に大きな権限と自主性を与えている。

今日では、DARPA は 240 人の在籍者（そのうちおよそ 140 人が技術要員）から成る組織であるが、約 20 億ドルという巨額の予算を直接管理している。中心となる技術分野は、マイクロ・エレクトロニクス技術、情報技術、バイオテクノロジーである。

② マネジメント上の特徴

創設にかかわる重要な特徴は下記のとおりである。何年にもわたって DARPA はこうした創設時の原則をしっかり守り続けてきている。

- 小規模で融通性に富むこと。DARPA には付属の研究機関は置かない。従って時代の変化に逆らっていつまでも存続させなくてはならないような固定したインフラはない。
- 組織の中に水平的な関係（プログラム・マネジャー、事務担当課長、DARPA 長官）があること。
- 実質的な自律性と自由があり、それを妨げる官僚的な障害がないこと。
- 産業界、大学、国立の研究機関などから世界的な科学技術者が技術スタッフとして入

っていること。

- 技術スタッフは常に新鮮な思考と将来への見通しを保つことができるよう、3～5 年で配置転換されること。
- プロジェクトベースで動いていること：3～5 年で目標を達成するよう仕向けるのが典型となっていること。ただし、大きな技術的課題にはもっと多くの時間があてられるが、その場合も段階を区切った一連の目標が設定される。
- 必要な支援（技術、契約の取り交わし、経営など）はその時々によって“調達”され、元からいるスタッフに支障を生じさせないで、そうした臨時の雇員を自由に活用する体制ができている。これは国防総省をはじめ他の政府機関（軍事 R&D グループ、NASA、NSF など）との合意に基づくものであるとともに、システム工学技術支援（SETA）の契約者と合意している。
- プログラム・マネジャーは DARPA の活動の中心だが、技術的に秀でていることかつ進取の気性に富むことが採用の基準となる。これまで最高の DARPA のプログラム担当者たちはいずれも目標を追求することにかけては“自由奔放な狂信者”とでもいうべき人たちであった。DARPA は、大きな飛躍は委員会で計画をつくることの中からは生まれるものではないと認識している。DARPA のプログラム・マネジャーはプログラムに対する遂行責任と説明責任を共に果たさねばならない。
- 運営にあたって配慮されるのは納税者からの資金を適切に使っていくことの責任であり、それ以外はほとんど規則で縛られるようなことはない。DARPA が担当すべきマネジメントの仕事とは、すなわちプログラム・マネジャーに活動を保証することなのである。
- 成功報酬が充分に高ければ、失敗についても完全に引き受けることになる。

③ 典型的なプロジェクト

典型的な技術プロジェクトの構造は次のとおりである。

- 4 年以上の期間でおよそ 1,000 万～4,000 万ドルの規模。
- プログラムの業務と資金運営には 1 人のプログラム運営担当者があたる。
- SETA の契約者（複数のこともある）がプログラム・マネジャーの支援のために置かれる。プログラム・マネジャーの任務は、個々のプログラムの業務をこなすと共に、議会、国防総省長官官房、各軍担当局、統合参謀本部などに対しプログラム内容の理解を図ることにある。
- 軍事研究開発施設内にプログラム関連部局（1 人から数人で構成される）が置かれ、プログラム予算からの資金により、技術的ならびに契約事務上の支援を行う。
- 統合された特別の目標が定められた場合、それに関する任務を遂行するのは 5～10 の契約機関と複数の大学である。

④ 選定過程

プログラム・マネジャーが活動を実質的に取り仕切ることになるのだが、その場合でも公募研究の採択に関しては形式的なレビューがなされる。すべてのプロポーザルはレビューア（評価者）のグループによって評価される。たいていは 1 グループは少なくとも 3 人の評価者がいるが、それ以上のこともある。ほとんどの場合、DOD の研究所もしくは連邦政府の他の研究所から評価者が選ばれる。選定の第一の基準は技術的な専門性である。DARPA の支援は画期的なアイデアに対して長期の助成をすることにあるので、最終的にその技術がどう使われるのかに関してはそれほど厳しく評価されるわけではない。

しかし、例えば NSF との相違点で著しいのは、DARPA のプログラム・マネジャーはプロポーザルの最終的な承認について実質的な裁量権をもっていることである。NSF のプログラム・マネジャーはどちらかというと控えめで、評価者たちの一致した意見に従う。DARPA のプログラム・マネジャーはそのプロポーザルが充分に画期的だと思えば、評価者たちと異なった判定を下すことができる。

3.1.6 欧州委員会の研究プロジェクト公募・選考方法：STRATA プログラムにおけるプロジェクト評価の事例

① EU 第 5 次フレームワーク・プログラムの概要

第 5 次フレームワーク・プログラムは欧州連合の「研究・技術開発及び実施」分野における 1998～2002 年の研究活動支援を規定し、今後の欧州連合の研究は EU が直面する主な社会経済的諸問題に対応し、技術・産業・経済・社会及び文化に関する問題解決に役立つ研究技術開発を支援する方針を決め、下記の 7 つの分野を対象に研究プロジェクトを公募した。

・テーマ別研究分野

1) 生活の質と生活資源の管理
2) ユーザー・フレンドリーな情報社会
3) 競合・持続可能な発展
4) エネルギー・環境・持続可能な開発

・総合的研究分野

5) 欧州連合内の研究活動の国際的役割の確認
6) イノベーション促進、中小企業の参加推進
7) 人間の研究潜在力と社会経済的知識の改善

ここでは、7) の Improving the Human Research Potential and the Socio-economic Knowledge Base 分野の STRATA （Strategic Analysis of Specific Political Issues）「特殊政策問題の戦略的分析」下のプロジェクトの選考、評価過程を記述する。

② STRATA プロジェクト公募と参加者

プロジェクトは定期的に公募される。公募はヨーロッパ連合の官報に発表され、欧州委員会研究局（Directorate Research）が事務局を勤める。公募は、主に欧州連合加盟国の個人、あるいは私的・公的研究機関、大学、企業（中小企業を含む）を対象に実施されるが、連合国に属する 18 カ国、すなわち次期 EU 加盟候補国、EU 周辺国（スイス等）、等も公募に応じ EU の研究資金援助を受けることが出来る。

③ プロジェクトの選考過程

規定に従い申請されたプロジェクトはすべて評価の対象となる。「評価」は、「透明」、「公平」を基本的理念とし、詳細に決められた評価手順・基準に従い実施される。欧州委員会（EC）が選んだ外部評価者が、後で述べるようなピアレビュー方式に基づきプロジェクトを評価する。評価選考過程は以下のごとくである。

1) 申請前照合

欧州委員会が実施する非公式諮問過程。研究プロジェクト申請希望者は、正式にプロジェクトを申請する前に、申請予定のプロジェクトが EU の公募に添ったものであるかを前もって欧州委員会に問い合わせる事ができる。プロジェクトのコーディネーターは申請プロジェクトの内容を要約し、参加者リストを添え、EU 委員会に提出する。書類受信 5 日以内に EC はファックスあるいは電子メールで検討結果を申請者に送信する。この検討結果は規定の用紙に記され、プロジェクトの概要が EU の要望に添ったものであるか否かを伝えるのみで、プロジェクトの内容の評価ではない。この検討結果がその後のプロジェクトの正式評価に際し参考にされる事はまったくない。

2) 申請前登録

応募プロジェクトの評価、審査準備、すなわち評価者の選考、招集等を迅速に行うため、欧州委員会は申請予定者に、プロジェクト名、要約、及び参加者リストを公募締め切り日の三週間前までにファックスあるいは電子メールで委員会に送ることを依頼する。この仮プロジェクト申請受信後数日以内に委員会は登録番号を申請予定者に送り、評価準備に入る。このプロセスは、EC の事務を円滑に行うもので、申請者がのちに、申請取り消しをすることは自由である。

3) 申請登録

欧州委員会に送られた申請プロジェクトは委員会のデータベースに登録され、受信書が申請者に送られる。締め切り 10 日後、すべての書類はリストアップされ、評価プロセスに入る。

4) 管理上のチェック

申請されたプロジェクト書類が、委員会の公募規定に添ったものかどうかを、委員会が検討する。検討される内容は、

－発信日、受信日と締め切り日との照合

－コーディネーター、コーディネーティング機関代表者、研究参加者それぞれの正式

署名

－プロジェクト参加者最低数とECの規定照合

－申請規約を網羅しているか等

規定にそわない申請者は直ちに除外され、コーディネーターに通知される。

5) 専門家による評価

欧州委員会は8～10人のパネルをエキスパート・リストから選び、評価委員会を組む。この内、最低3人のエキスパートが申請プロジェクトを吟味、コメント、採点する。

外部専門家（エキスパート）は公募により欧州委員会外の専門知識を備えたものが選ばれる。公募は定期的にEU官報に掲載される。エキスパートに求められる能力は、第5次フレームワーク・プログラムの包括する専門分野に精通した知識、特に、プロジェクト管理・評価、研究技術開発プロジェクト成果応用、技術移転、技術革新、科学技術、国際共同研究、及び人的資源開発等の分野で活躍、経験を有する者、国立・私立大学、研究機関、企業、コンサルタント等を対象に募集。また、研究開発の社会・経済的インパクトを測る能力を有し、欧州委員会の方針にそったプロジェクトを選び、評価する能力を有する者が求められる。エキスパートはまた、高度な語学力を要求される。

応募方法は、委員会の応募要項を記入、送付する。要項は、数十項に及ぶもので、詳しい履歴・能力・特技等を記入する。選ばれたエキスパートは、委員会のデータベースに登録される。プロジェクト審査の際、欧州委員会は必要な数のエキスパートをデータベースより選択し、パネルを形成。エキスパートの任期はフレームワーク・プログラムの6年間の内3年（または3公募以内）の評価に当たる。

プロジェクト選考の任に当たるエキスパートは、プログラムの要求する専門知識を有する他、国や性別のバランスをも考慮に欧州委員会が任命する。エキスパートは秘密保持契約、及び公平評価誓約書に署名し評価の任に就く。エキスパートは有給である。

パネルが組まれ、プロジェクトの検討作業は開始される。第一にパネル内から選ばれた3人のエキスパートのプロジェクト吟味が全体の基礎となる。エキスパートは一カ所に集められ、評価は第5次フレームワーク・プログラムで規定されている評価基準に従い実施される。エキスパートは各プロジェクト内容を吟味し、コメント及び採点を記す。エキスパートはプログラム全体に共通する評価基準に沿ったコメントの他、個々の公募内容に沿った評価コメントも要求される。申請プロジェクトの中で、時に、必ずしも、その時の公募に添ったものではないが、研究の質が高く評価されたプロジェクトは、他の公募に移転することが検討される場合がある。また、時に、フレームワーク・プログラム参加国以外の国が加わるプロジェクトに関しては、第三国の参加がEUにとって利益になるかの検討も行われる。

評価過程は委員会の監督のもとで実施され、エキスパートは定められた評価基準に

従って作業が進められる。

申請書の「科学技術の質、及びイノベーション」に関する項目は、匿名で作成されている。エキスパートは、まず、この項から評価を開始し、コメントを付し、採点する。その後、プロジェクト参加者の評価に入る。評価項目毎にコメントを付し、0～5の採点をする。0—書類はプログラムの評価基準に不適合、ないし情報不足。1—可（poor）、2—並（fair）、3—良（good）、4—優（very good）、5—秀（excellent）。

また、公募ごとに定められた「特別評価項目」にも 0～5 の採点をし、全体の点数を計算する。

6) パネル・ディスカッション

エキスパートの評価、採点、終了後、パネルが招集される。3 人の評価エキスパートを含む、8～10 人のエキスパートによりパネルは構成される。ここでは、3 人のエキスパートの評価が検討され、プロジェクトが選ばれる。パネルの合議後、評価サマリー・レポートが作成され、パネル全員が署名する。

ここで、パネルの合意が得られない場合、他のエキスパートが加わることもある。評価サマリー・レポートは、パネルの合意点を記すとともに異なる点も明記される。

7) プロジェクトの選考

評価の最終段階では、エキスパートを含むパネルによりランク付けされたプロジェクトが、全体としてプログラムの趣旨に基づき一貫したものかが再検討される。同得点を獲得したプロジェクトの再検討、優先順位付けが行われる他、合格した幾つかのプロジェクトの内、併合できるものの有無等も吟味される。

評価結果はプロジェクト・コーディネーターに通知される。

その後、欧州委員会はエキスパートとパネルの評価をもとに、採用プロジェクトの順位付けをする。この際、委員会は全プログラムの方針を優先させる為、時にエキスパート／パネルが作成した順位と異なるリストができることもある。この場合委員会は、エキスパート／パネルの評価との相違を文章に明確に表すことが義務づけられている。

最後に委員会は採用外プロジェクトのリストを作成する。不採用決定及びその理由をプロジェクト・コーディネーターに通知する。

8) 採用プロジェクトの契約

採用プロジェクトの順位付けが委員会により行われた後、プロジェクト・コーディネーターは評価結果（エキスパートのコメント・採点）とともに仮合格通知を受ける。仮採用されたプロジェクトのコーディネーターは、エキスパートのコメントに従い、疑問点を明確にし、修正や補足説明を加える等の作業に取り組まなければならない。又、事務的手続きに必要な追加書類の依頼を受ける場合もある。いずれの場合にもコーディネーターは期限内に委員会の要求する情報を提供する。期限内に委員会の要求に応じられなかったプロジェクトは、この段階で却下される場合もある。

すべての情報がそろった上で欧州委員会は契約書を作成し、コーディネーターに送

る。契約書の署名が終わった段階で、プロジェクトは活動開始する。

9) 評価レポート

各公募の評価終了後、評価要約レポートが欧州委員会により作成され公開される。レポートには、申請プロジェクトのテーマ、申請者の職業、申請予算等の統計データ他、採用されたプロジェクトの数、予算、申請者内容等の統計を掲載している。又、評価を担当したエキスパートに関するデータ（人数、専門分野、国籍、性別等）も公開される。

個々のプロジェクト評価ごとの担当エキスパートの名は公開されないが、委員会はある期間内に評価作業に当ったエキスパートをまとめて、一つのリストを作成し公開している。

10) オブザーバーによる評価活動の点検

上記のプロジェクト評価活動中、年に一回評価活動の監視が専門家により実施される。オブザーバーと呼ばれる数名の専門家がこの任務を担当する。オブザーバーは欧州委員会が、「評価・監視エキスパート」及び「研究開発プログラム顧問」の公募に従って採用された専門家である。オブザーバーの任務は、プロジェクトの評価が委員会の定めた基準に従って正しく行われているかを検討し、評価活動が効率よく進められるよう改善すべき点があれば指摘、提案することである。

3.1.7 フランス国立保健医学研究所（INSERM）の公募プログラムにおけるプロジェクト評価の事例（匿名外国人評価委員）

1) 国立保健医学研究所（INSERM）の概要

INSERM（インセルム）は、本来ならば病院において行われるべき研究を強化するために設けられた研究機関で、医療行為を実践すべき病院の補完的機関として研究機能を担っている。

2) プログラムの設置

インセルムは毎年テーマ別に研究プログラムを立案し、研究プロジェクトを公募している。公募は主としてインセルムの研究者を対象としているが、時に他の公的研究機関（CNRS、大学、病院等）とインセルムの共同プログラムを設置することがあり、インセルムの枠を超えてプロジェクトを募集する場合もある。公募対象はプログラムによって異なり、一律ではないがインセルム外の研究者の参加を募る場合には、インセルムの常置評価審議会の承認を得る必要がある。

プログラムの規模、各プロジェクトの研究期間も様々で、年間設置されるプログラム数も決められていない。このようにインセルムでは、プログラムはケース・バイ・ケースのゆるい枠組の中で必要に応じて計画される。

研究プロジェクトの公募は、各研究所、大学共同研究所等インセルム関連研究者に通知される他、ウェブサイトにも常時掲載される。

3) 評価委員会の設置

プログラムが公募されると、評価委員がインセルム局長により任命され、委員会が設置される。評価委員会とは、インセルムに常設されている評価審議会とは異なり、プログラムの設置の度に組織されるアドホックな委員会である。

プログラムの規模により委員数は異なるが、通常 12 人以上、24 人以下で組織される。インセルムの研究者約 5,000 人は、インセルムのプロパーな研究所に 2,000 人、病院／インセルムとの共同研究所に 1,800 人、他の公的研究機関（CNRS、INRA 等）との共同研究所に 1,000 人と全研究者の半数以上はインセルム外の研究所と組んでいることから、評価委員も広くインセルム外の分野の専門家が加わることが多い。

評価委員は、論文等を通して研究活動、専門性を吟味される他、他の委員会ですでに活躍している者は、その活動成績も合わせて調べられ、十分な科学的知識を備え、評価委員としても適任と判断された者が選ばれる。

今日まで、評価委員に企業からの人材が加わった例は、ほとんど皆無である。

4) 匿名評価委員の任命

匿名評価委員は専門知識を有するフランス人以外の人が任命される。インセルムの評価関係データ・ベースには、分野別に外国人専門家が 1,000 人程度登録されている。プログラムごとに専門家はリストアップされ、評価委員はそのリストを検討、承認する。承認後、評価事務局は、そのリストの中から 1 プロジェクトに 2 人の匿名評価委員を指名し、プロジェクト書類を送る。匿名委員の数は、応募プロジェクト数によって異なる。例えば、あるプログラムに 60 件のプロジェクトの応募があった場合には、一プロジェクト審査に 2 名任命、合計 120 名の非フランス人専門家が任命される。これらの匿名委員は事務局のみ名を知り、専門家リストを承認した評価委員でさえ最終的には、誰が匿名委員に任命されたのか知らされない。

匿名委員の仕事は、プロジェクトのオリジナリティー、予想成果の適切さを評価するほか、研究チームの評価も行う。一つのプロジェクトのコメント及びプロジェクトの採用の資格を述べる。

匿名委員は、このように評価期間中は匿名でインセルムの評価作業に参加し、無給の奉仕をするが、インセルムの「評価委員」という肩書きが与えられ、研究キャリアにプラスになることから、評価依頼を拒む人はほとんどいない。

5) 事前評価課程

プロジェクトの応募書類は英／仏両方で記される。この英仏 2 ヶ国語書類作成はプログラムの応募のみならず、インセルムでは研究者評価、機関評価の書類作成にも適用される。フランス人以外の評価者の参加を可能にするための政策である。提出書類は 10 数枚に及ぶ。その中には、特に匿名委員として自分のプロジェクトの評価に当るのを避けて欲しい研究者がいる場合には、その旨を記すことができる。これは、競

合者が不適当な評価を下す可能性を避け公正を期すためであるが、申請者の希望を評価委員会が取り上げるか否かは、委員会の判断に一任されている。

集ったプロジェクトは第一に匿名委員に送られる。各申請プロジェクトにつき、2人の委員が評価し、それぞれの結果がインセルムに送り返される。

次に一プロジェクトにつき2人の割でインセルムの評価委員が評価作業を行う。プログラムの規模により、各評価委員に割り当てられるプロジェクト数は異なる。例えば12人の評価委員により60のプロジェクトが評価される場合、各評価委員は10プロジェクトの評価を担当することになる。こうして、一つのプロジェクトにつき合計4人の専門家が評価する（2匿名評価委員、2評価委員）。

評価委員は、各プロジェクトに当てられた2名の匿名委員の評を吟味した上で、自分の評価を下す。評価委員はプロジェクトをコメントする他、ABCDの4段階採点を行う。

評価委員が自分に割り当てられたプロジェクト批評、採点を終えると、評価委員全員の会合が開かれる。各委員の評価結果が討議され、プロジェクトが選ばれる。一日平均30プロジェクトが審議される。評価日数はプログラムの規模によって異なる。

すべてのプロジェクト審査終了後、評価委員会は各プロジェクトの批評をまとめる。評価結果とプロジェクト批評はインセルムの局長に送られ、局長の承認後、各プロジェクト・リーダーに送られる。公募から評価終了まで約4～6ヶ月要する。

6) プロジェクトの中間・事後評価

プロジェクトの規模により、中間評価を行う場合と、プロジェクト終了後評価のみを行う場合に分かれる。3年のプロジェクトは18ヶ月後に中間評価が施行され、プロジェクトの今後の継続が審議される。

中間評価も、プロジェクト終了評価も、評価委員が担当する。いずれの場合も、プロジェクトが予定の成果をあげられたか吟味されるが、プロジェクトのインパクト、社会への影響といった科学的側面外の調査は行われない。このような評価は今後の課題として、インセルムで目下検討中であるという。

インセルムではプロジェクト選考評価課程では、外部、特に外国人の援助を受けるが、プロジェクトの中間評価、プロジェクトの事後評価は、内部の評価委員が担当する。

3.2 プログラム評価の事例

以下では研究開発の評価手法が実際の評価でいかに用いられているか具体例を紹介する。事例は、公的研究開発のプログラム等の評価を受託している海外コンサルティング機関であり、米国のSRI International、Arthur D. Little、英国のTechnopolisである。ただし、これらはいずれもプログラムレベルの評価であることには注意が必要である。

3.2.1 SRI International の評価事例

SRI International 内部の組織である Science and Technology Policy Program では、研究プロジェクト、プログラム、政策といった各レベルの評価を主要業務の一つとして行っている。この部門はジョージア工科大学の公共政策学科の教授も務める David Roessner 氏が中心的なメンバーであり、実際のプロジェクト評価でも同大学の Bozeman 教授らとの共同のもとに評価活動を行っている。

(1) Engineering Research Center プログラムの評価

この評価では、米国 NSF が 1985 年から設置を行っている Engineering Research Center が産業にいかにインパクトを与えてきたかを、「大学研究者と産業とのインタラクションのパターン」に焦点を絞って分析した。通常考えられるような「ERC によってどの程度の経済的効果が得られたか」といった計量経済学的分析は、効果が得られるまでに長期間かかることやその他の要因とも関係するために分析することは困難な場合が多い。そのため、SRI ではそのような連鎖のステップの中間にある、「大学と産業とのインタラクションのパターン」に焦点を置き、そのメカニズム自体や重要度の分析を Center の顧客である産業側へのインタビュー、郵送・電話アンケート調査により行っている。

ERC プログラムの目的は大学と産業との結びつきを強化することによって米国の競争力を強化することであり、具体的には次のような目標が設定されている。

①工学システムアプローチによる学際的研究・教育の実施

②「次世代発展」に関連し、その成果が短期的すぎるフォーカスでなくて産業界に有用な研究の実施

③米国産業の国際競争力の獲得

④センターで開発された知識・技術の市場への移転を強化する新たな方法の提供

⑤継続的かつ相互に利益を生むインタラクションを形成するためのリエゾンプログラムの設立

⑥実験室から市場へと製品を移転させるのに何が必要かについての幅広い理解の学生への提供

これまでに多くの ERC は助成期間を終了して独り立ちのフェーズへと入っている。調査が行われた時点では助成を受けている ERC は 18 あり、産業との 700 以上のパートナーシップを形成していた。

評価ではまず、大学と産業との関係の形態について、考えられるものを列挙することが行われた。ERC プログラムスタッフは次の 10 のメカニズムを認識していた。

①ERC の研究アジェンダの共同による戦略的プランニング

②ERC に産業から出向する技術者

③共同プロジェクト（実証試験を含む）
④テクニカル・アドバイザリー・グループ
⑤特許（特に特許やソフトウェアのライセンス）
⑥ERC の卒業生の産業への就職
⑦ERC の成果を基にしたスタートアップ企業
⑧特定問題に対応した産業と ERC のコンソーシアム
⑨学生：共同プロジェクトにおいて産業の技術者とともに研究を行う、産業界からの指導教官、論文審査委員会メンバー、サマージョブ、企業への訪問、産業界からのセミナースピーカー
⑩産業からのサポート：会員費、研究委託、装置寄付、人材交換

このような予備調査をもとにして、実際に ERC によって大学研究者と産業との間のインタラクションのパターンが分析された。特に、このようなインタラクションの結果としてどのようなタイプのインパクトがおきているかを分析することに焦点がおかれた。分析では、まずドラフトのモデルを産業界の参加者グループに提示することで改善し、次に ERC データベースにおけるデータの分析を行った。その上で、調査の方法としては ERC 参加企業への郵送でのアンケートが行われた。対象は 18 の大学の ERC に関与している企業で回答数は 355 である。また、補足的に対面インタビューや電話インタビューを行った。

分析結果の要点は次のようなものである。

・ERC の技術分野としては、デザイン・製造が 30.4%と多く、バイオテクノロジー・バイオエンジニアリングが 20.0%と続く。参加企業の業種は電気・通信機器が 21.6%と多く、化学が 15.8%。
・ERC 参加企業のメンバーであった期間は平均 4.4 年。回答者は、教授や博士号以上の研究者とは平均年 3～5 回電話や E-mail でコンタクトするとともに、年 3～5 回実際に対面している。
・ERC に参加したことで得られる利益に影響する要因については、「ERC と企業との技術分野の焦点の近さ」が評点平均 3.6 点（5 点中）と最も高く、「ERC 教授の企業ニーズへの反応」（3.4 点）、「ERC が参加者とコミュニケーションをとる努力」（3.4 点）と続く。「地理的な近さ」は 2.4 点と 14 の選択肢の中では最も低い。
・利益を得るのを阻害する要因としては、「他の会社の問題事項に ERC が占有されてしまう」（2.4 点）が最も高いが、総じて評点はどの項目も低く、大きな問題は存在していないことが示されている。
・ERC のメンバー期間が長い企業、教授らとのコミュニケーション頻度が高い回答者は利益が高かったと回答している。
・ERC の研究アジェンダ設定に企業が強く影響したと考えている企業ほど、利益が高かったと回答している。
・メンバーシップ期間にかかわらず、成果は、「新たなアイデア、ノウハウ、技術へのアクセス」が最も高く、「直接の技術支援」がそれに続く。一方で、「他の企業とのイン

タラクション」や「製品・プロセスの改善」はメンバーシップ期間が長期であるほど顕著になる（図１参照）。

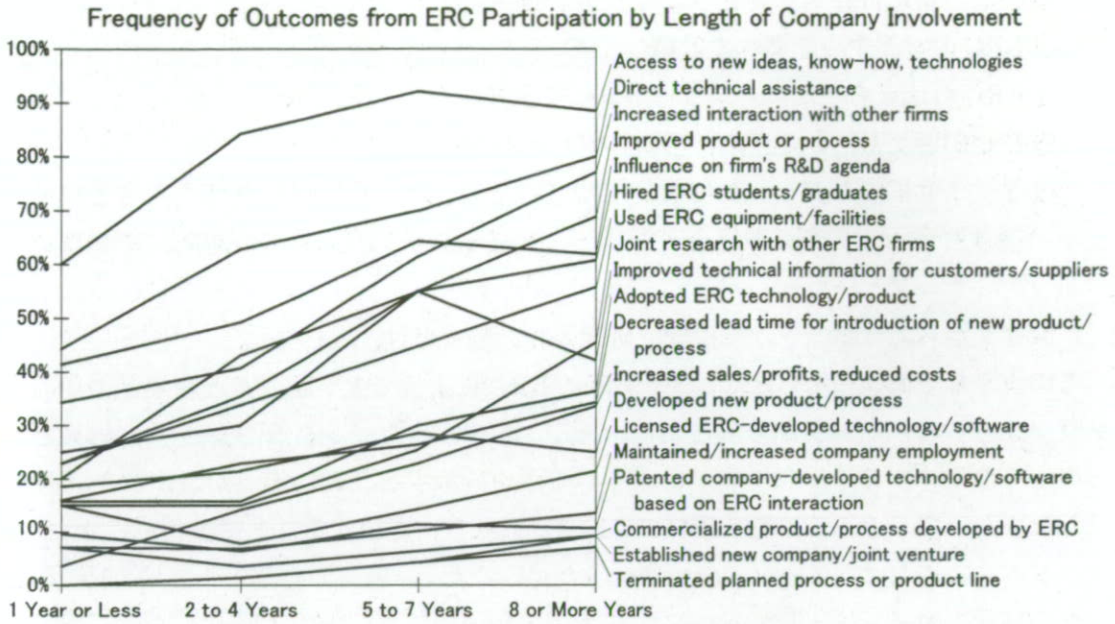

図１　ERC 参加企業の成果の表れた頻度

(2) 技術イノベーションに対する NSF サポートの評価

この評価は、NSF が工学研究に対して行ってきた支援が、最近の 12 の重要なイノベーションの開発や商業化にいかに貢献してきたかを、4 年間かけて分析したものである。すなわち、これはイノベーションに対する NSF サポートの貢献度の評価であり、1960～70 年代の TRACES や Hindsight などで行われてきた遡及的評価の最近の事例の一つである。ただし、その中では後述のように文献分析、インタビュー、ビブリオメトリックス、特許分析なども行われている。分野は初年度は、インターネット、MRI（磁気共鳴映像法）、ポリマー混合物の反応射出成型（RIM）の 3 つを、2 年次は、電子回路への CAD の応用、テレコミュニケーションのための光ファイバー、携帯電話を取り上げている。

調査ではまず、選択されたイノベーションについて、主要なプレイヤー、年表、技術変化、その他の特徴を文献調査によって把握した。次に NSF スタッフ及びプログラムに関係した科学者・技術者にインタビューを行い、個人や組織の貢献、マイルストーン、特許などの詳細情報を得た。また、NSF の助成金データベースにより、プリンシパル・インベスティゲーター、助成機関、博士学生へのグラント、旅費、ワークショップなどの情報を 1995 年から得ることができた。その後に、キーとなる人物へのインタビューを行い、知識や技術の流れを把握した。同時に特許・論文分析も行っている。その上で、NSF 及びその他か

らの支援のタイプや影響が把握された。特に、NSF の研究、教育、施設、装置、ワークショップへのサポートのタイミングと影響に注意をおいている。

結論として得られたのは、各事例で機関間の関係も、重要な要素も異なっているということであった。共通して言えることは、機関間のネットワーキングが重要であったことであり、いくつかの事例ではこれはキー研究者のキャリアパスに現れている。事例別の NSF の貢献は、反応射出成型（RIM）では直接の研究サポートが初期に影響しており、MRI では NMR spectroscopy 研究、技術的・分析的コアへのサポート、将来 MRI に関与する大学院生の教育への影響が強い。インターネットでは、組織的なリーダーシップが大きいという結果が出ている。

以下では、具体的に MRI の事例についてその評価内容をまとめる。

MRI（磁気共鳴映像法）は CT スキャンにまさる診断装置として 1988 年までに多くの医療機関に導入されていったものである。MRI（磁気共鳴イメージ）は技術的にはそれ以前から研究開発が行われていた NMR（核磁気共鳴）に大きく依存しているが、その市場は重複が極めて少ないという特徴を有していた。SRI による評価では、まず MRI の要素技術を明確にすることが行われた。すなわち、磁石、グラディエント・コイル、高周波トランスミッタ及びレシーバ、コンピュータ、ディスプレイ、空間符号化と画像再構成技術といった要素技術について、その開発過程の概要や NMR との関係を含めて分析された。また、NMR 分光器、CT スキャン、MRI それぞれの歴史が簡単に分析された。

評価の結論は、次の 4 つの観点から示されている。

① 官産学の役割と関係

MRI の研究開発の最初の 10 年間（1971～1981 年）においては、特に英国の大学の役割が重要であり、物理学者と医学者によって研究が行われてきた。英国政府は特に Medical Research Council からのサポートを研究者に提供していた。一方で、産業界は 1970 年代中ごろには EMI のみが MRI の研究開発に関与した。米国では 1970 年に、Raymond Damadian が NMR により癌を探知する研究を行い、1973 年に Paul Lauterbur が試験管を画像化する研究を行った。その後、これら技術が MRI として確立された。だが、1980 年代半ばに英国の大学による MRI 機器の臨床への応用によって、医療への MRI の応用の可能性の大きさが明らかになると、産業界が主導的役割を果たすように変化した。急速な発展は General Electric や Siemens によってもたらされた。多くの場合には大学との共同研究ではなく、企業は独自に研究を行った。1985 年に政府の医療保障で MRI スキャンの使用が認められ、次の年には、MRI は最も急速に成長している医療画像技術となった。

② 基礎研究と技術開発との関係

MRI の基礎的要素である NMR は核物理学に依存していたが、スタンフォード大での基礎研究から Varian Associates の Russell Varian によって急速に商業化の道へと進んだ。この時点から企業による開発が中心となった。さらに、ノーベル賞受賞物理学者 Richard Ernst が 2D イメージングの開発をすすめた。核共鳴の現象が理解された後は、

NMR と MRI の開発は基礎研究よりは技術開発に依存した。

③ 知的所有権

MRI に関しては多くの特許が出されたが、それによって開発が遅れたという明らかな証拠はない。パテント・サイテーションの分析によれば、特許は、他者の研究を明確に把握したうえで出願されている。

④ NSF の役割

NSF は、NMR の研究や装置への研究助成金によって、米国の参入者が MRI 開発において依存するインフラストラクチャーの多くの部分を提供した。1955 年から現在までに、NMR 研究への助成額は 9000 万ドルになる。これには装置購入や大学院生の教育が含まれる。さらに NSF は関連分野へのサポートも行ってきた。しかしながら、これらの MRI の開発への NSF の直接的役割の追加的な証拠は 1980 年代終わりまではない。

このように分析では、技術自体の開発及び商業化の歴史が詳細に分析された上で、各要素技術ごとに NSF のサポートの貢献を分析し、全体としての役割を上記のように結論づけている。この事例は特定プログラムの評価ではないために、プログラム目的に基づく評価という形態をとるものではない。この評価では、技術開発の歴史的展開、及び NSF のサポートの種類の双方について、幅広い観点からの分析を行うことで、妥当性の高い評価を行おうとしている。

参考資料 Roessner, D., B.Bozeman, I.Feller, C.Hill, and N.Newman (1997), The Role of NSF's Support of Engineering in Enabling Technological Innovation, First Year Final Report

Roessner, D., R.Carr, I.Feller, M.McGeary, and N.Newman (1998) *The Role of NSF's Support of Engineering in Enabling Technological Innovation Phase II, Excective Summary*

C.P.Ailes, D.Roessner and I.Feller (1997), *The Impact on Industry of Interaction with Engineering Research Centers, Final Report*

3.2.2 Arthur D. Little の評価事例

ADL は 1886 年にマサチューセッツ工科大の Arthur D.Little 氏により設立された民間受託研究機関であり、ボストンやブリュッセル、東京を主要拠点に、国際的な経営や政府機関のコンサルティング活動を行っている。この一環として、政府や民間機関における研究開発や技術関連の戦略策定や評価の支援の受託事業がある。

(1) 事例 1 ：DOE のエネルギー効率・再生可能エネルギー部門の評価

米国エネルギー省 (DOE) のエネルギー効率・再生可能エネルギー部門 (Office of Energy Efficiency and Renewable Energy: EERE) が GPRA (行政機関業績・成果法) に基づいて行う研究開発プログラム評価の支援を ADL は 6 年前より行っている。2000 年の報告書

では、EERE 内部の以下の 5 つのセクター（部署）の計 14 の Planning Unit（プログラム）についての評価を ADL は行った。この評価事業の委託費は 1 年で$250,000（3 千万円程度）である。

① Office of Power Technologies（OPT）
- Hydrogen Research and Development
- Power System Integration – Transmission Reliability
- Solar Buildings
- Energy Storage Systems

② Office of Industrial Technologies（OIT）
- Steel Vision
- Industrial Assessment Center
- Inventions & Innovations

③ Office of Building Technology, State and Community Programs（BTS）
- Community Energy Program
- Technology Roadmaps and Competitive Research and Development
- Weatherization

④ Office of Transportation Technologies （OTT）
- Materials Technologies
- Technology Deployment
- Vehicle Technologies - Heavy Vehicle Systems

⑤ Federal Energy Management Program （FEMP）
（内部の全プログラム）

これらセクターと Planning Unit の評価において、ADL は 2 つの定量的な測定を行っている。これらは社会的な効果の定量的測定である。

① 各技術によるエネルギー使用削減、炭素排出削減量の 2001～2030 年度における見積
② Planning Unit のパフォーマンス測定（短期目標及び次の 5 年間へのマイルストーンを含む）

この 2 つに加えて 5 つのセクターそれぞれについて、セクターレベルでのパフォーマンス測定項目（PMs）の設定についても成果とともに焦点を置いている。

すなわち、評価結果として記述されているものは、まずは各セクターそれぞれについて測定項目としていかなるものを設定する必要があるのか、実際にはどのような成果が得られたのかである。次に各 Planning Unit それぞれについて上述の排出量見積もり、測定項目の設定、DOE の反応などが列挙されている。このように、GPRA に対応した評価では、目標達成を示す指標の設定そのものがその値とともに評価活動の重要な焦点となっている。下表 1，2 はそれぞれ OPT のセクターレベルでの評価結果と、Hydrogen Research and Development Planning Unit の評価を示したものである。

表１：OPT(Office of Power Technologies)のセクターレベルでの評価結果

OPT Sector Level PMs/Accomplishments

MAJOR FINDINGS FOR SECTOR LEVEL PM

- The goal of tripling installed U.S. renewable generation capacity by 2010 is an appropriate Performance Measure （PM）.
- OPT needs to develop additional quantitative sector-level PMs that reflect more than one program and that reflect OPT goals that are not directly related to energy and emissions savings.

MAJOR FINDINGS FOR SECTOR ACCOMPLISHMENTS

- OPT did not develop sector-level accomplishments for the FY2001 GPRA submission. They will need to be developed for FY2002.

表２：OPT の Planningy Unit の評価

Planning Unit

Hydrogen Research and Development

	Total Primary Energy Displaced （Trillion Btus）						
	2001	2005	2010	2015	2020	2025	2030
Preliminary Draft	0	45	165	363	588	Not avail.	Not avail.
Final Submission	0	1	43	143	303	460	620

MAJOR FINDINGS FOR QM

- Market size estimates were initially too high and occurred too soon for the applications considered given current product status and the timing of planned demonstrations.
- Fuel cell mini-grids require more detailed analysis to strengthen the rationale and the benefits calculation.

MAJOR FINDINGS FOR PLANNING UNIT PM

- OPT has presented several useful trended PMs that need to be added to the budget document. Several additional, complementary PMs may be beneficial.
- The links between the QM, PM and Planned Accomplishments need to be strengthened within the GPRA documentation.
- A single, self-contained GPRA document needs to be developed that includes QM, PM and Planned Accomplishments. This document would draw upon, and be consistent with, other Hydrogen program documentation.

DOE RESPONSES AND ACTIONS

- OPT has appropriately modified the FY2001 QM analysis, and has agreed to several other recommendations for future GPRA analyses.
- OPT provided information showing the link between the PM, QM and Planned Accomplishments, and agreed that a self-contained GPRA document needs to be developed.

(2)オランダ経済科学教育省におけるエネルギーR&D ポートフォリオ分析

オランダ経済科学教育省 the Dutch Ministries of Economic Affairs and Science and

Education の委託のもとに、ADL はエネルギー分野全体の研究開発全体の評価をポートフォリオ分析によって行った。

このような調査が行われた背景としては、1995 年にオランダ政府がエネルギー政策の変更を取り上げたことにある。すなわち、それまでの石炭を基盤とする技術から、ガス及び再生可能エネルギーの技術へとシフトした。また、市場の規制撤廃へのシフトという背景も存在する。そのため、政府支援によるエネルギー関連の研究開発活動と、新たな政策や市場の要求との間のマッチングを知るために、R&D ポートフォリオの理解が必要になったのである。

このような背景のもとに ADL では第一に 5 つの契約研究機関と 7 つの大学の全てのプリンシパル・インベスティゲーターへのヒアリングを行った。調査は、次のような全てのエネルギー・バリュー・チェーンをカバーしていた。

・資源採掘、生産
・資源の転換（精製、代替燃料の製造など）
・輸送と分配
・エンドユース

次にこのエネルギーポートフォリオを国のエネルギー政策のもとに分析し、オーバーラップする分野や相乗的共同研究の可能性について同定した。次に、エネルギー使用やエネルギー装置の企業に対してエネルギーR&D のニーズについて上記それぞれの分野についてインタビューした。それをもとにエネルギーR&D ポートフォリオが産業ニーズや政策に合致しているかやそのバランス（新規技術と成熟技術のバランス、大学と国研のバランス）を分析した。最後に各キー技術分野についてワークショップを開催し結果を議論した。

評価によるインパクトとしては、研究開発行為の調整や再焦点付けの改善、新たな共同研究の開始へと展開したと言う。

参考資料 Arthur D. Little, Inc. (2000), *Summary of Findings – Peer Review of the FY2001 GPRA Assumptions, Report to National Renewable Energy Laboratory*

3.2.3 Technopolis の評価の事例

Technopolis は 1989 年に設立した科学技術政策を専門とするコンサルティング会社であり、英国サセックス大学 SPRU 出身の研究者により設立された。Technopolis はこれまで多くの研究開発プログラムの評価を行ってきている。表 3 はこれまで Technopolis が行ったプログラム評価の主要なもののリスト、及びその評価で用いた分析方法をまとめたものである。

表 3 で示されるように、Technopolis がプログラム評価において有効であると考えている手法は、ケーススタディとユーザー・サーベイである。前者は実際に行われた研究開発活動を理解する上で最も効果的な方法であり、後者はプログラムの成果やそのインパクトを

分析する上で必須のものと考えられている。一方で、書誌計量学的手法は個人の研究成果やその重要性を理解するには用いられるが、プログラム全体のレベルの評価については使用することが困難という特徴を有している。

表3　Technopolis が行ったプログラム評価において用いた手法

プログラム名	歴史的分析	ケーススタディ	書誌計量学的手法	ピアレビュー	ユーザー・サーベイ	コスト・ベネフィット分析
技術移転・普及プログラム						
試験的						
CEC – IRISI（国際的情報社会イニシアティブ）	–	●	–	△	△	–
CEC SPRINT – SPAL（特定プロジェクト・アクションライン）	–	●	–	△	●	–
TEKES – ESV	△	△	–	●	●	–
Forbairt – TTPP（技術移転パートナーシップ・プログラム）	△	●	–	–	●	●
KFB – Biofuels	●	△	–	●	●	–
成熟的						
研究プログラム						
試験的						
ESRC – Face Processing Initiative	●	●	△	●	●	–
NUTEK, BFR, SBUF – IT – BYGG	△	△	–	●	●	–
Forbairt – Basic Research	●	●	△	–	●	–
成熟的						

●…主に用いた　△…いくらか用いた　–…用いていない

(1) 事例１：アイルランドの技術移転・パートナーシップ・プログラム（TTPP)の評価

上記の表の中でコスト・ベネフィット分析を用いた唯一の評価が、アイルランドのTTPPを対象としたものである。コスト・ベネフィット分析は、コストの算出は比較的容易であるが、ベネフィットをいかにして分析するのかが難しいという問題を有している。すなわち、プログラムに参加した企業や機関については、どの範囲までがプログラムからの効果と言えるのかを線引きするのが困難であり、さらにプログラムには直接参加していない機関への波及効果（インパクト、あるいは外部性）を考慮するとさらにその範囲の設定は難しくなる。

TTPP はフィージビリティ調査から投資までが急速に進むシンプルなプロジェクトが多く、企業も追跡しやすいという特徴を有していた。そのため、コスト・ベネフィット分析を比較的に行いやすいという判断がなされた。分析では最初の 3 年間のみのベネフィットを調査し、可能な場合には全プロジェクトコストにおける TTPP のシェアをベネフィットにかけて計測した。詳細な分析方法については評価レポートの本文が未入手であるために不明であるが、分析結果では、ベネフィット：コスト比が 4：1 から 5.5：1 という結果が得られている。また、得られた結論のうちの主要なものを列挙する。

- プログラムはよりイノベーティブなメンバーやハイテクセクターに焦点を置いている。
- プログラムは大企業よりも小企業（50 人以下）が技術移転に関わることを増進させることに成功している。
- インパクトは長期でかつ他のサポートシステムからのインパクトと切り離すことは困難である。
- プログラムの対象企業への支援額はおどろくほど大きく、1991～96 年で一企業あたり 30 万ポンドである。それにもかかわらず、企業は資金面でないサポートに重きを置いている。
- プログラムの企業の主要な焦点は、アイルランドやそのほかの EU 諸国におけるマーケットアクセスを増加させることである。
- プログラムは、新しいネットワーク、新しい製品、新しい顧客、新しいプロセスを形成する企業の能力に対して効果を有している。
- プログラムは全体として 450 の新たな仕事を形成し、一つあたり、I£24000 のコストがかかっている。

(2) 事例２：EU INFO2000 プログラムの評価

INFO2000 は欧州マルチメディアの使用を刺激するとともに、欧州のマルチメディア産業の世界的競争力を支援することを目的とした EU のプログラムである。実施期間は 1996～99 年であり、全体で 65 百万 EURO が用いられた。Technolopois はイタリアの Databank

Consulting、ベルギーの LENTIC、フランスの IDATE と提携し、このプログラムの終了時点での評価を行った。

プログラム自体は次の 4 つのアクションラインが設定されたものである。

・ディマンドを刺激し、認知度を上げる
・欧州の公的セクターの情報を活用する
・欧州のマルチメディアのポテンシャルを誘発する
・サポートアクション

このようなプログラムに対して、評価は次の基準から行われた。

・INFO2000 はその目的に合致するのに効果的であったか
・プログラムの背後の「介入のロジック」は効果的であったか
・プログラムは効率的にマネジメントされ組織化されていたか
・プログラムの原理、シナジー、維持は調整されていたか
・プログラムはターゲットとする顧客のニーズに合致していたか
・中間評価における提言は実行されていたか

評価は 6 ヶ月もの期間をかけて行われた。評価の方法は次のようなものである。

①セクトラルオーバービュー、構成分析
プログラム期間中に欧州のマルチメディアが発展してきた科学的・技術的・経済的・政治的背景の分析。

②プログラムから生まれた文書の分析

③ケーススタディ
EU 委員会スタッフやプロジェクト参加者、産業のオブザーバー全 44 名に対する構造化されたインタビュー。

④500 以上の参加者へのアンケート。
これらを通じて、上記の評価基準それぞれについて評価を行うとともに、将来のプログラムについての提言を行っている。

参考資料

Arnord, Erik （1997）, "Uses and Limits of Cost-Benefit Analysis in Evaluating Industrial Modernisation Programs: The Irish Technology Transfer Program", Paper prepared for the *American Evaluation Association Conference*

Arnord, Erik and Guy Ken （1997）, "Technology Diffusion Programmes and the Challenge for Evaluation", *Policy Evaluation in Innovation and Technology*, OECD

Final Evaluation of the INFO2000 Programme, Final Report, Technopolice Ltd, Databank Consulting SpA, IDATE, and LENTIC (2000)

参考資料４．　主要文献

4.1　主要総説文献

◆ 総合

1. OTA, "Research Funding as an Investment: Can We Measure the Returns? -A Technical Memorandum", NTIS (1986.4)

本報告所は米国科学技術下院委員会科学政策タスクチームが提議した「研究投資を他の経済投資と比較可能な形で測定できるのか」という問いに対して OTA（技術評価局）が 1986 年当時の研究評価の研究動向をレビューしたものであり、翌年に発刊された OECD の報告書とともに研究評価が体系化される初期段階でのレビュー報告書の一つである。報告書は主に、経済的リターンの測定、非経済的アウトプットの測定（ビブリオメトリックス、科学指標）、産業及び政府での評価の意思決定プロセスへの利用に分けて事例を交えて説明されている。既にこの報告書においても、公的研究開発の利益は国防や衛生など多様なものでありそれを経済的価値（金銭）で見積もることの困難さと不適切さが指摘されている。それでも経済的価値を測定しようとするのであれば、科学を技術やイノベーションと連結する要素に注目することが必要であり、技術移転や産業ニーズ主導の研究、予測とプラニングの必要性が述べられる。また、ビブリオメトリックス手法については NSF や NIH の分析事例を引きながら今後の利用可能性を述べるとともに、これらは指標にすぎず評価と同等ではないことも指摘される。また、産業や政府での評価手法の利用状況の調査では、経済的手法は産業においても懐疑的であり、基礎研究では主観的なスコアリングモデルが主流であり、応用研究でも質的情報に基づく点数付けが主流であることが示される。産業においても政府においても、R&D の計画や予算作成では多くの部門・組織・階層が集まって反復的な情報交流や将来予測による意思決定を行うことこそが重要であり、経済的モデリングなどの評価はそれを喚起するための２次的な役割しか果たさないことが述べられている。

2. Bozeman,Barry (ed.), Melkers,Julia (ed.), "Evaluating R&D Impacts: Methods and Practice", Kluwer Academic Publishers (1993)

本書は研究評価に用いられる方法論と実際の評価事例について、米国及びカナダの著者らによって 1993 年に書かれたものである。そのため、近年の米国の動向（GPRA による影響や ATP 評価など）や欧州における評価の動向を把握するには向かない反面、ピアレビュー、事例研究、ビブリオメトリックス、企業における経済性評価、オペレーションズ・リサーチ手法などの基盤的方法が順に解説されており、入門書として適している。方法論を紹介した幾つかの章では、主にプロジェクトレベルの評価について、それぞれの方法論の内容と長所・短所を説明しており、いずれにおいても方法論には限界が存在するために他の複数の方法と組み合わせて用いることが推奨される。また、実際の評価事例を報告した複数の章では、定量的手法が実際には企業・政府において使用頻度が低いこととその理由の調査結果、公的研究所の産業との関係の評価、スピンオフ効果の評価、ポートフォリオなどが議論される。

3. Office of Naval Research, "Handbook of Research Impact Assessment. Edition 5. Summer 1995", NTIS (1995)

本書はONRのKostoff氏がそのホームページ上で公表している研究評価関連の報告書（ハンドブック）の中の1冊である。本書は米国連邦政府による研究のインパクト評価に焦点をおいて、事前・中間・事後の評価について、遡及的手法（Hindsightや TRACESなど）、定性的手法（ピアレビューなど）、定量的手法（費用対効果分析、文献計量学的手法、モデリングなど）の各方法論について説明を行う。本書は 3000 以上の参考文献リストを含んでおり、今後進んで学習するための足がかりとしても使える報告書である。本書ではピアレビューは最も用いられている手法であるために、その質の維持・向上のための条件が示される一方で、ピアレビューだけでは完全でなく、その他の手法をも用いる必要が指摘される。

4. Callon,Michel, Laredo,Philippe, Mustar,Philippe, "The strategic management of research and technology", Editions Economica (1997)

本書はEUの評価研究者のネットワークにより書かれたものであり、出版時点での欧州での研究評価の最新動向を示したものとなっている。本書では特に研究開発プログラムの評価に焦点がおかれている。最初の部分で欧州各国及びEUの研究評価システムの特徴が紹介され、続いてプログラム評価において必要となる様々な視点が議論されたのち、ビブリオメトリックス手法、経済的効果の測定手法（計量経済学的手法、BETA 法、コントロール・グループ・アプローチ、前後比較）、アンケート・インタビューの手法、ネットワーク効果の評価手法が具体的な評価事例を基に記述されている。本書に含まれている内容は多様であるが、評価システムの設計からプログラム自体の概念化、そして具体的な評価作業の設計までを包括的に把握することに最適なテキストブックである。

◇ Program の社会経済効果

5. Georghiou,Luke, Roessner,David, "Evaluating technology programs: tools and methods", *Rsearch Policy*, Vol.29, 657-678 (2000)

本論文は、技術開発プログラムの評価の手法について、欧米の主要なプログラムの評価事例を基にして包括的にレビューしたものである。そのため、欧州及び米国の研究評価ならびに評価研究の現状を把握するには最適の論文である。論文では大学・公的機関で実施される研究プログラム、産学・官民などの連携促進プログラム、普及・エクステンションプログラムの3種に区分して評価の現状を報告している。ここでは生産関数あるいは投資利益率や内部収益率の分析は実際にはプログラムの正当化や意思決定には有効でないことが強調され、新たな動向として知識生産者と利用者のネットワークを重視する試みが指摘される。また、より具体的にアンケート対象の設定による結果の分散の問題や比較群の設定の必要などが指摘される。

6. National Research Council, "The Advanced Technology Program: Assessing Outcomes", National Academy Press (2001)

本書はNRCが行ったAdvanced Technology Program（ATP）の評価の報告書であり、主にNRCが主催したATP評価に関するシンポジウムでの議論をとりまとめたものである。ATPは民間企業による研究開発プロジェクトを競争的に選択して公的資金を提

供する制度であるために、市場経済に政府が介入することの正当性への批判が特に議会から継続的になされてきた。そのため、経済学者やシンクタンク・コンサルタント、GAO などによる多種多様な評価が行われており、シンポジウムではそれら評価の取り組みが紹介されている。

ATP の評価で問題とされることは、プログラム内の研究開発プロジェクトが経済効果をもたらしたかという単純なものではなく、民間企業によるプロジェクトに公的資金を提供する正当性があるのか、さらには、正当性を認めるための社会的利益がどのようなメカニズムによって得られているのかを把握することにある。この点において ATP の評価の取り組みは、依然として研究開発が実施されその効果が得られるメカニズムをブラックボックスとし、インプットとアウトプットのみからの費用対効果を求めようとする日本に多くの示唆を与えるものとなっている。

例えば前者の正当化については、GAO が 1997 年に行った評価では、ATP からの資金が得られなくても申請者が他の助成を受ける可能性があったか、助成を受けなかった場合に国民経済上の著しい損失となる可能性があったかを検討した。これは事後・追跡評価で行われるものであるが、その結果はプロジェクト選定（事前評価）のクライテリアを提供するものである。また、ATP では民間企業が行いにくいハイリスク・ハイリターンのプロジェクトを選定するためにプロジェクトの失敗割合が高くなることを許容する必要があることと、民間企業からもほぼ同額の資金拠出を求めることにより実現可能性がある程度存在するプロジェクトが申請されることのバランスがとられているというプログラム・マネジメントの特徴も評価されている。その他にも、私的利益と社会的利益の関係、ベンチャーキャピタルと ATP との関係、ATP における大企業の役割と必要性、プロジェクトの審査方法の信頼性が議論されている。

また、社会的利益のメカニズムについては、Jaffe（1996）はスピルオーバーのメカニズムを「知識スピルオーバー」「市場スピルオーバー」「ネットワークスピルオーバー」に区分し、いかなる特徴を持つプロジェクトが波及効果が大きいかを分析している。これは、プロジェクト選択・形成の指針を提供するものでありプログラムの中間評価の主要な役割と言える。

7. **科学技術政策研究所第 2 研究グループ『米国における公的研究開発の評価手法』2002 年**

本報告書は、主に ATP の評価に用いられた手法の紹介とビブリオメトリックス手法の紹介の 2 つから構成される。ATP の手法については長年 ATP の Economic Assessment Office のディレクターを務めた Ruegg 女史によりまとめられている。報告書では ATP の評価に用いられた方法を 1)ポートフォリオ作成のための統計的プロファイリング、2)参加者及び非参加者へのサーベイ調査、3)叙述的な事例研究、4)費用便益分析等を用いた事例研究、5)計量経済学的手法にまとめ、それぞれについて具体的な評価事例（大学の経済学者やコンサルタントなどが行った実際の評価のレポート）を引きながら説明する。また、ビブリオメトリックス手法の部分については米国 CHI 社がこれまで省庁などからの委託によって行ってきた評価・分析事例を紹介している。

8. ETAN, "ETAN Working Paper, Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes", (1999.1)

本報告書は EU のフレームワーク・プログラム（FP）の評価（及びモニタリング）

システムを今後いかに改善すべきかについて、EU の研究評価研究者によりまとめられ EU 第 12 総局へ提出されたものである。そのため評価システムの具体的な改善方法については FP 特有の内容であるが、その前提として議論されていることは、研究評価全体の近年の問題や今後の方向性、あるいはプログラム評価の一般的な要求事項を含むものとなっている。特に、FP 自体が 5 次に以降した際に社会・経済的目的への対応を強めたことにより、社会・経済的インパクト（特に、金銭的な計測にそぐわない社会的インパクト）をいかに評価するかの問題が生じた。ここではインパクトの種類やそのタイムスパンを区分したのちに、インパクトの測定には限界があることや Additionality 概念の必要性が議論される。また、FP は異なる期のものが重なって展開されるという特徴を有するため、過去のプログラムの事後・追跡評価や現在のプログラムのモニタリングから新たなプログラムへのフィードバックを得るようなシステムの体系化が提案されている。

9. PREST, AUEB, BETA, ISI, Joanneum Research, IE HAS, Wise Guys, "Assessing the Socio-economic Impacts of the Framework Programme", (2002.2)

本書は欧州の研究評価研究者らによる、EU のフレームワーク・プログラム（FP）の社会経済的インパクトの評価方法に関する報告書である。本書はプログラム評価に用いられる方法論を具体的な事例を基に包括的に比較検討したものであるだけでなく、それらの背景にある評価の理念あるいは公的研究開発プログラムの理論的根拠をも深く検討している点で優れた報告書となっている。具体的には、第 2 章では FP により公的支出を行うことの経済的合理性を経済学、経営学、イノベーション研究、科学技術政策研究など様々な学問的視点から考察している。これはそもそも公的プログラムの評価においてどのような点に焦点をおくべきであるか、どのような効果が得られていればプログラムは望ましいと考えるかといった理論的根拠を与えるものである。第 3 章では、FP に対して行われてきた評価研究自体をメタ評価する。比較群などを設定せずに期待される影響を考える「模索的評価」と、事後的に実際の影響を測る「堅い（solid）評価」とを区別したうえで、特に後者について FP 及び EUREKA と ATP の評価事例を基にいかなる次元の効果・影響がどのようなタイムスパンで評価されているか、及びその利点と欠点を検討する。第 4 章ではケーススタディとして、社会科学研究の評価やサービス・セクターでの研究の評価といった評価研究で今後に課題とされる分野が示されるとともに、BETA 法・費用便益分析・オプションアプローチと言った経済効果分析手法の比較検討などが行われている。このように、本書はこれまでの評価研究における方法の比較・総括と今後に課題なる分野を明らかにした良書である。

10. Ken Guy, "Strategic Options for the Evaluation of the R&D Programmes of the European Union", Technopolis Ltd. (1998.11)

◇ 基礎研究の評価

11. Arnold,Erik, Balazs,Katalin, "Methods in The Evaluation of Publicly Funded Basic Research –A Review for OECD" Technopolis Ltd. (1998.3)

本論文（報告書）は Technopolis 社により行われた基礎研究の（経済・社会的効果の）評価手法のレビューである。本論文では、これまでの先行的な分析・研究は基礎研究と

経済効果の正の関係を認めているが、いかなる研究が経済効果につながるかは示していないと述べ、今後の基礎研究の評価の方法は科学と経済の間のリンクを理解し、戦略的計画と評価を結合させることが必要であることを指摘する。このようなスタンスから、本論文では基礎研究のアウトプットの次元の類型、基礎研究からの利益の類型、これまでイノベーション研究で発展してきた理論モデルの類型、あるいは分野ごとの基礎研究からの社会経済的利益への展開の違いなどを先行研究をレビューしつつ概念整理する。このように、単にアウトプットやアウトカムの計量でなく、研究開発からそれら社会・経済的効果へと展開する過程自体の理解を一層深めるべきであるという考えは特に欧州の研究評価では現在の中心的課題として提議されているものであり、本論文はそのための出発点として優れた報告であると言える。また、全ての対象に用いることができる単一の評価システムはないとし、個別のシステムを形成するための一般的な評価システム・プロセスのデザインの手順とそこで必要となる変数及び評価項目をまとめる。また、手法については、定量データ、ケーススタディ、ビブリオメトリックス、共語分析、ピアレビュー、特許分析、質問紙・サーベイ調査、費用便益分析についてその概略と長所・短所をのべ、そのパッチワーク的な複合的利用が必要であることを述べる。

12. National Academy of Public Administration, "A Study of The National Science Foundation's Criteria for Project Selection", (2001.2)
    本報告書は、NSF のプロジェクト選定（メリットレビュー）のメタ評価である。NSF はメリットレビューの基準を 1997 年に変更したが、その変更が実際にどのような効果を選定プロセスや結果に生んでいるかを調査し改善のための提言を行う。このため、報告書の内容は NSF に固有の内容とはなっているが、他方で実際の現場レベルでピアレビューのシステムを改善するためには、どのようなことが必要かを示している。

◇ 国際会議講演集

13. OECD, "Policy Evaluation in Innovation and Technology –Towards Best Practice", (1997)

14. Buhre,Susanne (ed.), Kuhlmann,Stefan (ed.), "Evaluation of Science and Technology in the New Europe", *Proceedings of International Conference*, Berlin (1999.7-8)

15. Philip Shapira (ed.), Kuhlmann,Stefan (ed.), "The 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation", Bad Herrenalb (2000.9)

16. European Commission Directorate General for Research, "The Contribution of Socio-economic Research to The Benchmarking of RTD Policies in Europe", Brussels (2001.3)

17. National Science Foundation, "Research Assessment: What's Next?", Washington. DC. (2001.5)

各国の研究評価研究者を招いて米国で開かれたワークショップであり、その概要はCozzensによりまとめられWebで公開されている。このワークショップでは、評価研究者、評価の実務家、評価のユーザーが一同に介し、タイトルにもあるように、研究評価研究に今後何が必要であるかを探ることを目的とした。概要で強調されている点は米国と欧州での研究評価の動向の違いである。米国ではGPRAによりパフォーマンス指標が重視され、いかなる指標をいかに測定するかが評価における中心課題として捉えられている。しかし、いかなる指標が必要であるかは研究プロセスあるいはその市場への展開プロセス自体の深い理解が必要とされるのに対し、実際にはそのような洞察はGPRAでは注目されないという矛盾を抱えていることが報告されている。他方で、欧州では指標ではなくいかなるシステムで評価を行い意思決定へと結びつけていくかが焦点とされている。米国はGPRAシステムと評価をいかなる関係におくかを検討すべきという議論が報告には記されている。また、評価自体の品質管理、あるいは意思決定と結びついた「戦略的評価」のための条件、研究評価で今後焦点となるトピックスなどがワークショップでは議論されている。

4.2 社会経済性総説文献：

◇プログラム効率の評価

National Audit Office(1995) The Department of Trade and Industry's Support for Innovation, Report by the Comptroller and Auditor General, HG 715, 1994-1995

◇研究開発評価論

Bozeman,B. & Melkers,J.(eds.)(1993) Evaluating R&D Impacts: Methods and Practice, Kluwer: Boston

Van Raan,A.(1988) Handbook of Quantitative Studies of Science and Technology, Elsevier: Amsterdam

◇行動研究

Argyris,C, and Schon D.A.(1974) Theory in practice: increasing professional effectiveness, Jossey-Bass

Clark, P.A.(1972) Action research and organisational change, Harper and Row

◇ＥＵにおける評価

Bobe,B. & Viala,H.(1997) "Panels of experts of the European Commission" in Callon,M., Laredo,P. & Mustar, P. (eds.) The Strategic Management of Research and Technology - Evaluation of Programmes", Economica: Paris

Massimo,L.(1997) "Lessons from the experience of the European Commission" in Callon,M., Laredo, P. & Mustar(eds.) The Strategic Management of Research and Technology - Evaluation of Programmes", Economica: Paris

Melkers,J. & Cozzens,S.(1997) "Developing and Transferring Technology in State S&T Programs: Assessing Performance" Journal of Technology Transfer, Vol.22, No.2, pages 27-32.

Luukkonen,T. (1998) "The difficulties in assessing the impact of EU Framework Programmes" Research Policy, No.27, pages 599-610

Luukkonen,T.(2000) "Additionality of EU Framework Programmes", Research Policy, No.30, Vol.6, pages 711-724

Bach,L. & Lambert,G. (1992) "Evaluation of the economic effects of large R&D Programmes: the case of the European space programme" Research Evaluation, Vol.2, No.1, pages 17-26.

Dale,A. & Barker,K.(1994) The Evaluation of EUREKA: a pan-European collaborative evaluation of a pan-European collaborative technology programme, Research Evaluation, Vol.4, No.1, pages 66-74

Sand,F. & Nedeva, M.(1998) The EUREKA Continuous and Systematic Evaluation Procedure: As assessment of the socio-economic impact of the international support given by the EUREKA initiative to industrial R&D co-operation

Cunningham,P. & Nedeva,M.(1999) Towards a system of continuous evaluation and monitoring for European co-operation in scientific and technical research (COST), Research Evaluation, Vol.8, No.1 pages 142-154

◇ベンチマーキング

Rush,H. et al(1995) "Strategies for Best Practice in Research and Technology Institutes - An Overview of a Benchmarking Exercise" R&D Management Vol.25, No.1, pages 17-31

Lundvall,B. & Tomlinson,M.(2001) "Learning by Comparing: Reflections on the Use and Abuse of International Benchmarking" in Sweeny,G.(ed.) Innovation, Economic Progress and the Quality of Life, Edward Elgar: Cheltenham

Walgenbach,P. & Hegele,C.(2001) "What Can an Apple Learn from an Orange? Or: What Do Companies Use Benchmarking For" Organization, Vol.8, No.1, pages 121-144

◇コストベネフィット分析

Averch,H.(1994) "Economic Approaches to the Evaluation of Research", Evaluation Review Vol.18, No.1, pages 77-88

◇計量文献分析

See Van Raan,A.(1988) Handbook of Quantitative Studies of Science and Technology, Part II, Elsevier: Amsterdam

Rao,I.(1996) "Strategies for Best Practice in Research and Technology Institutes - An Overview of a Benchmarking Exercise", Scientometrics Vol.25, No.2, pages 265-270

◇メタ分析

Rosenthal,R.(1991) Meta-Analytic Procedures for Social Research, Sage: Newbury Park

Wachter,K. & Straf,M.(1990) (eds.) The Future of Meta-Analysis, Russell Sage: New York

◇ケーススタディ

Yin,R.(1994) Case Study Research: Design and Methods, 2nd Edition, Sage: Newbury Park

Luukkonen-Gronow,T.(1987) "Scientific Research Evaluation: A Review of Methods and Various Contexts for Their Application", R&D Management, Vol.17, No.3, pages 207-221

◇科学計量学ーサイエンティフィックマッピング

Rip,A.(1988) "Mapping of Science: Possibilities and Limitations" in Van Raan, A.(1988) Handbook of Quantitative Studies of Science and Technology, Elsevier: Amsterdam

◇フォーサイト

Irvine,J. & Martin,B.(1984) Foresight in Science: Picking the Winners, Frances Pinter: London

◇ＥＣ文書

Comm (97) 553 Final "Fifth Framework Programme for Research and Technological Development(1998-2002) Commission Working Paper on the Specific Programmes: Starting Points for Discussion"

Comm (2001) 94 "decision of the European parliament and of the council Concerning the multiannual framework programme 2002-2006 of The European community for research, technological Development and demonstration activities aimed at Contributing towards the creation of the European research Area

SEC (2000) 1973.

◇フレームワーク・プログラムの評価

Bach,L. & Lambert,G.(1992) "Evaluation of the economic effects of large R&D Programmes: the case of the European space programme" Research Evaluation, Vol.2, No.1, pages 17-26.

Bozeman,B. & Kingsley,G.(1997) "R&D Value Mapping: A New Approach to Case Study-Based Evaluation" The Journal of Technology Transfer, Vol.22, No.1, pages 33-42

Rossi and Freeman (1993) Evaluation - A Systematic Approach, Sage: London

Piric,A. & Reeve,N.(1997) "Evaluation of Public Investment in R&D - Towards A Contingency Analysis", OECD Conference Policy Evaluation in Innovation and Technology Towards Best Practices

Georghiou,L. & Meyer-Krahmer,F.(1992) "Evaluation of socio-economic effects of European Community R&D Programmes in the SPEAR network", Research Evaluation, Vol.2, No.1, pages 5-15

Williams,D. & Rank,A.D.(1998) "Measuring the economic benefits of research and development: the current state of the art", Research Evaluation, Vol.7, No.1, pages 17-30

Roessner,D.(2000) "Quantitative and qualitative methods and measures m the evaluation of research", Research Evaluation, Vol.8, No.2, pages 125-132

Arnold,E. & Balazs,K.(1998) Methods in the Evaluation of Publicly Funded Basic Research, Overview for OECD

Link A. & Scott,J.(1997) "Policy Evaluation in Innovation and Technology Towards Best Practices - Lessons from the National Institute of Standards and Technology" OECD Conference Policy Evaluation in Innovation and Technology Towards Best Practices

Link A. & Scott,J.(199?) Hugh ? "Evaluating Public Sector R&D: A Retrospective Examination of US Approaches" APEC Conference

Dietz,J.(2000) "Building a social capital model of research development: the case of he Experimental Program to Stimulate Competitive Research", Science and Public Policy, Vol.27, No.2, pages 137-145

Georghiou,L. & Roessner,D. (2000) "Evaluating technology programs: tools and methods" Research Policy, Vol.29, pages 657-678

Papaconstantinou,G. & Polt,W. (1997) "Policy Evaluation in Innovation and Technology: An Overview", OECD Conference Policy Evaluation in Innovation and Technology Towards Best Practices

Wallace H. & Wallace,W.(2000) Policy -Making in the European Union, fourth Edition, Oxford University Press: Oxford

Nicoll, Sir W. & Salmon,T.(2000) Understanding the European Union, Pearson: Harlow

Salter,A. & Martin,B.(2001) "The economic benefits of publicly funded research: a critical review" Research Policy, Vol.30 pages 509-532

Georghiou,L.(1999) Socio-economic effects of collaborative R&D - some European Experiences, Journal of Technology Transfer Vol.24 No.1 pages 69-79

◇Ｒ＆Ｄ査定への財政的アプローチ

Perlitz,M., Peske,T. & Schrank,R. (1999) "Real options valuation: the new frontier" R&D Management, Vol.29, No.3, pages 255-269

Trigeorgis,L.(1996) Real Options: Managerial Flexibility and Strategy in Resource Allocation, Cambridge, MA: The MIT Press

◇ＲＴＤのインパクト

Bach,L. & Georghiou,L.(1998) The Nature and Scope of RTD Impact Measurement, Discussion Paper for International Worksop on Measurement of RTD Results / Impacts, Brussels.

Barker,K.(1994) "Strengthening the impact of R&D evaluation on policy-making: methodological and organisational considerations", Science and Public Policy, Vol.21, No.6, pages 405-413

European Commission / ETAN Working Paper(1999) "Options and Limits for Assessing the Socio-economic Impact of European RTD Programmes

Georghiou,L.(1998) "Socio-economic Effects of Collaborative R&D -European Experiences", Journal of Technology Transfer, Vol.24, pages 69-79

Illinois Institute of Technology(1968) "Technology in Retrospect and Critical Events in Science", IIT Report Series

Kostoff,R.(1994) "Assessing research impact: US government retrospective and quantitative approaches", Research Evaluation. Vol.21, No.1, pages 13-22

Martin,F.(1998) "The economic impact of Canadian university R&D" Research Policy Vol.27, pages 677-687

Narin,F., Hamilton,K. & Olivastro,D.(1997) "The linkages between US technology and public science Research Policy, Vol.27, pages 835-851

Salter,A. & Martin,B.(2001) "The Economic benefits of publicly funded basic research - a critical review" Research Policy, Vol. 30, pages 509-531

Working Group on S&T Financial Management Mechanisms (Canada) (1993) Methods for Assessing the Socio-economic Impacts of Government S&T

◇ピアレビュー

Godlee,F. & Jefferson,T.(1999) Peer Review in the Health Sciences, BMJ Books: London

Barker,K.(1994) "R&D Evaluation - Strengthening the impact of R&D evaluation on policy making: methodological and organisation considerations", Research Evaluation, Vol.21, No.6, pages 405-413

Bozeman,B. & Melkers,J.(eds.)(1993) Evaluating R&D Impacts: Methods and Practice, Kluwer: Boston

Porter,A. & Rossini,F.(1985) Peer Review of Interdisciplinary Proposals" Science, Technology and Human Values, Vol.10, No.3, pages 34-42

◇社会科学研究とそのインパクト

Newby,H.(1994) "The Challenges for social science: a new role in public policy-making" Research Evaluation, Vol.4, No.1, pages 6-11

Shea,M. & Lewko,J.(1994) "Transferring knowledge to influence 'small p' policy development: examples from the implementation of integrated service delivery systems" Research Evaluation, Vol.4, No.1, pages 31-36

Spaapen,J. & Sylvain,C.(1993) "Assessing the value of research for society", Research Evaluation, Vol.3, No.2, pages 117-126

Van Langenhove,L.(2001) "Can the Social Sciences Act as an Agent of Change in Society? An Introduction to the Workshop on the Contribution of Social Sciences to Knowledge and Decision Making", Social Sciences for Knowledge and Decision Making, OECD: Paris

Harre,R.(2001) "Causal Mechanism and Social Practices: What can Social Science Contribute to Social Practice? Social Sciences for Knowledge and Decision Making, OECD: Paris

Bedford,R.(2001) "A Robust Research/Policy Interface: international Migration and Social Transformation in the Asia-Pacific Region", Social Sciences for Knowledge and Decision Making, OECD: Paris

Kastrinos,N.(2001) "Promoting Dialogue Between Research and Policy-Making: The ETAN Experience" Social Sciences for Knowledge and Decision Making, OECD: Paris

Boddington,A.(1993) "Research evaluation systems: sources of policy information and 'evaluation push"' Research Evaluation, Vol.3, No.3 pages 133-208

Shadish, William R.Jr., Cooks,Thomas D., Leviton,Laura C.(1991) Foundations of Program Evaluation: Theories of Practice, Sage: London

Smith,G.(1993) "A comment on the reflexive relationship between social policy and social research" Conference on Evaluation, Social Science and Public Policy, Ottawa, Canada

Tiler C. & Boddington,A.(1993) "Outputs, structure and process in the evaluation of social science research centres" Research Evaluation, Vol.3, No.2, pages 107-116

Willinsky,J.(2000) If Only We Knew - Increasing the Public Value of Social Science Research, Routledge: London

◇社会経済的なベネフィットと生活の質（ＱＯＬ）

McDaniel B.(2001) "The Crisis in Social And Institutional integration" Social Science Journal Vol.38 No.2, Pages 263-275

参考資料５．　欧米の評価研究者ネットワーク

5.1　評価研究者の国際的ネットワーク

（1）The World Research Evaluation Network

http://www.reseval.net

◆ **目的と活動の概要**

世界の評価研究者と評価機関（offices）が経験、方法、データを共有するための組織である。このwebサイトには、新しく公表された論文の紹介、今後開催される会議の広報や過去の会議の記録の掲載を行っている。かなり大量の近年の重要文献資料が蓄積されている。

活動の目標は、研究評価（research assessment）の手法に関する最新状況をユーザーのコンテキストを含みつつレビューすること、そしてその分野の研究のベースとなるアジェンダや評価方法の開発について概観することである。

◆ **活動の内容**

The Res. Eval. discussion list という名前のメーリングリスト（購読制）を運営しており、世界中の研究評価者にリンクできるようになっている。サイトのアドバイザリー・ボードとして、Phillipe Laredo, Philip Shapira, and Gretchen Jordan が関与し、サイト自体のスポンサーは NSF の Program of Research on Science and Technology である。Susan Cozzens（susan.cozzens@pubpolicy.gatech.edu.）がサイトの更新にあたっている。

またこのネットワークのウェッブサイトには、関連する世界中の研究者、雑誌、研究機関、コンサルタント会社、政府機関へのリンクが張られている。

◆ **ウェッブサイトに掲載（リンク）されている資料**

① 研究者及び研究グループ／助成機関等の評価グループ
② 国際会議情報
④ 評価方法に関する重要文献（ダウンロード可能）
⑤ 当ネットワークが関与した評価活動ならびに評価研究活動（ダウンロード可能）
⑥ 評価に関連する情報源（S&Tデータ、連邦政府のプログラム評価関連情報源）
⑦ 専門誌"Research Evaluation"について

◆ **研究者及び研究グループ**

▶ Asia / Pacific

- Ron Johnston（University of Sydney
  ACIIC, Australian Centre for Innovation and International Competitiveness, Australia）
- Tim Turpin（University of Wollongong, Centre for Research Policy, Australia）

▶ Canada

- Benoit Godin（Universite du Quebec a Montreal

CIRST, Centre Interuniversitaire de Recherche sur la Science et la Technologie)

- Yves Gingras (Universite du Quebec a Montreal, CIRST)
- Benlit Godin (OST, Observatoire des sciences et des technologies)
- Adam Holbrook (Simon Fraser University
  CPROST, Centre for Policy Recearch on Science and Technology)

▶ Europe

- David F.J.Campbell (IHS, Institute for Advanced Studies, Austria)
- Stefan Kuhlmann
  (ISI, Fraunhofer Institute for Systems and innovation Research, Germany)
- Research Policy Institute (Lund University, Sweden)
- Luke Georghiou (University of Manchester,
  PREST, Public Recearch in Engineering, Science and Technology)
- Ben Martin (University of Sussex at Brighton,
  SPRU, Science and Technology Policy Research)
- 個人名記載なし(Evaluation Associates Ltd.)
- Erik Arnold (Technopolis, Innovation Policy Research Associates)
- Patries Boekholt (Technopolis, Innovation Policy Research Associates)
- Pirjo Niskanen (VTT, Group for Technology Studies, Finland)
- Terttu Luukkonen (VTT, Group for Technology Studies, Finland) (注)現籍は異なる。
- Juan Campanario (Universidad de Alcala de Gebares, Spain)

▶ Latin America

- Judith Sutz (Republic University of Uruguay
  CSIC, Comision Sectorial de Investigacion Cientifica)
- Sandra de Negraes Brisolla (State University of Campinas, UNICAMP Brazil
  Geosciences Institute
  DPCT, Departiment of Scietific and Technological Policy)
- Lea Velho (State University of Campinas, UNICAMP, Brazil)

▶ United States

- Ronald Kostoff (Office of Naval Research)
- Barry Bozeman (Georgia Institute of Technology,
  RVM, The Research Value Mapping Erogram)
- Juan Rogers (Georgia Institute of Technology, RVM)
- Susan Cozzens (Georgia Institute of Technology
  TPAC, Technology Policy &Assenssment Center)
- Gordon Kingsley (Georgia Institute of Technology, TPAC)
- Alan Porter (Georgia Institute of Technology, TPAC)
- David Roessner (Georgia Institute of Technology, TPAC)
- Philip Shapira (Georgia Institute of Technology, TPAC)
- Irwin Feller (Penn State University, Institute for Policy Research and Evaluation)
- Diana Hicks (CHI Research Inc.)
- Francis Narin (CHI Research Inc.)

• Michael Crow (CSPO, Center for Science, Policy & Outcomes)

◇ 助成機関等の評価グループ

▶ Canada

• Conseil de la Science et de la Technologie, Gouvernement du Quebec
http://www.cst.gouv.qc.ca/cst_English.html

▶ Europe

• DTI, Department of Trade and Industry, United Kingdom
TESE, Technology, Economics, Statistics & Evaluation Directorate
http://www.dti.gov.uk/tese (John Barber)

▶ United States

• NIST, National Institute of Standards and Technology
http://www.nist.gov
ATP, Advanced Technology Program/Economic Assessment Office
http://www.atp.nist.gov/eao/eao_main.htm

• National Science Foundation
http://www.nsf.gov
Office of Integrative Activities
http://www.nsf.gov/od/oia/start.htm (Nathaniel Pitts)
REC, Research, Evaluation & Communication
http://www.ehr.nsf.gov/EHR/rec (Eric Hamilton)

• U.S. General Accounting Office
http://www.gao.gov Robin (M.Nazzaro)

◇ 国際会議情報

• "Seventh International S&T Indicators Conference", Karlsruhe, Germany (September 25-28 2002 予定)
• "Evidence for Better Decision-Making", Canadian Evaluation Society, Annual Conference, Halifax, Nova Scotia (May 5-8, 2002 予定)
• "International Conference on Socio-economic Evaluation of Public R&D", Tokyo, Japan (April 18-19, 2002 予定)
• "Evaluation and Transition: devolution, diversity and evidence based change", United Kingdom Evaluation Society (UKES) Annual Conference Belfast (December 6-7, 2001)
• "International best practices in evaluation of research in public institutes and universities", Brussels, Belgium (November 12-13, 2001)
• "Mainstreaming Evaluation, American Association Evaluation (AEA) Annual Conference", St.Louis, Missouri (November 6-10, 2001)
• "The Fifth Iberoamerican, Interamerican Workshop of Science and Technology Indicators", Montevideo, Uruguay (October 15-18, 2001)
• "China's Economy: Confronting Restructuring, Stability, and International Competitiveness", University of Wollongong, Australia International Business Research Institute Department of Economics (July 14-15, 2001)

- "Strengthening the Capacity of Science and Technology Evaluation in APEC Region Science and Technology Evaluation in the 21st century", (May 30 - June 3, 2001)
- "Powering Up for Evaluation: The Odyssey Continues", (May 21-23, 2001)
- "Research Assessment: What's Next?" (May 17-20, 2001)
- "EC conference on the contribution of socio-economic research to the benchmarking of RTD policies in Europe", (March 15-16, 2001)
- "Learning from S&T Policy Evaluation". (September 11-14, 2000)
- "Evaluation of Science and Technology in the New Europe",. (June 7-8 1999)
- "Civilian Technology Policy in the European Union and the United States: Recent Experiences and New Directions", (April 9-10, 1999)

◇ **評価方法に関する重要文献**

- Erik Arnold, Katalin Balázs, "Methods in the Evaluation of Publicly Funded Basic Research", Technopolis, March 1998.
- Philippe Laredo, "The development of a reproducible method for the characterisation of a large set of research collectives", Armines/CSI, March 1999.
- Ken Guy, Katalin Balazs, James Stroyan, Erik Arnold, "Strategic Options for the Evaluation of the R&D Programmes of the European Union", Technopolis, November 1998.
- "Policy Evaluation in Innovation and Technology. Towards Best Practices OECD", Organization for Economoic Co-Operation and Development.

◇ **専門誌『Research Evaluation』について**（http://www.scipol.demon.co.uk/re.htm）

ピアレビューによる国際誌。個々の研究プロジェクトの評価から研究活動の国際比較までを幅広く扱う。公立機関も民間機関も対象とし、自然科学も社会科学も扱う。4 月、8 月、12 月の年 3 回の発行（発行は Beech Tree Publishing）。

- 編集主幹：Professor Anthony van Raan (University of Leiden, The Netherlands)
  Susan Cozzens (Georgia Institute of Technology, USA)
- 書評編集主幹：Michael Keenan (University of Manchester, UK)
- アドバイザリーボード：Andy Boddington (Evaluation Associates, UK),
  Bruce Campbell (Servier Laboratories, UK),
  Luke Georghiou (PREST, Manchester University),
  John Holmfield (formerly US House of Representatives),
  Ron Johnston (ACIIC, Sydney, Australia),
  Wilhelm Krull (Max-Planck Gesellschaft, Germany),
  Philippe Laredo (Ecole des Mines/CSI, France),
  Terttu Luukkonen (Tech Research Centre of Finland),
  C Marciano da Silva (New University of Lisbon),
  Ben Martin (SPRU, Sussex University),
  Frieder Meyer-Krahmer (Fraunhofer Institute, Germany),
  Francis Narin (CHI Research, USA),
  Erkki Ormala (S&T, Helsinki, Finland),
  Giorgio Sirilli (National Research Council, Italy).

(2) Members of the "6 Countries Programme" —The International Innovation Network

http://www.6cp.net/

◇ 目的と活動の概要

1975年に設立された、イノベーションの研究と政策形成にかかわる専門家、政策形成担当者、政策実施者などからなる国際的なネットワーク。設立にかかわったのはオランダ、ドイツ、フランス、英国の4ヵ国だが、現在はメンバーが、オーストリア、ベルギー、カナダ、フィンランド、ハンガリー、アイルランド、スウェーデンと合計11ヵ国にまたがる。

1年に2回開催される運営委員会（Steering Committee）には各国から少なくとも1名が入り、ワークショップを企画する。秋と春の2回のワークショップが最も重要なイベントとなる。インフォーマルで開かれた雰囲気の中で、各国の最新情勢に関する情報と意見を交換し、この領域の専門家たちの研究を刺激することをねらいにしている。成果は6CPのウェッブサイト、ニューズレター、そして雑誌論文や出版物の形で公表される。

◇ the 6 Countries Programme のメンバー（The International Innovation Network）

- Austria: Mr. Georg Panholzer, Mr. Wolfgang Polt
- Belgium: Mr. Paul Zeeuwts
- Canada: Mr. Geoffrey Nimmo, Dr. Micheal Kelly
- Finland: Mr. Torsti Loikkanen
- France: Mr. Remi Barre
- Germany: Prof. dr. Stefan Kuhlmann
- Hungary: Dr. Andras Siegler
- Ireland: Mr. Dermot O'Doherty, Mr. Gerry Sweeney
- The Netherlands: Mr. Arie van der Zwan, Mr. Jaap van der Vlies, Prof.dr.Ruud Smits
- Sweden: Mr. Lennart Elg
- United Kingdom: Mr. Tony Pedrotti

◇ 最新回・次回のワークショップ

- Innovation policy and sustainable development
  Can public innovation incentives make a difference?
  Flanders, Six Countries Programme Conference,
  Brussels, 28 February- 1 March 2002
- New Governance for Innovation
  14-15 November 2002, Karlsruhe, Germany:
- Governance of Innovation: Policies and Strategies within Organisations, Countries and Regions
  Spring 2003, Canada

◇ 過去のワークショップ

49. Merging Innovation Cultures - Leveraging Creativity for Competitive Advantage (4-5 October 2001, Dublin)
48. Innovation Policies for a New Era (15-16 January 2001, Stockholm)

47. Services in Innovation - The Role of Knowledge Intensive Business Services in the Innovation Process of Firms (11-12 May 2000: Utrecht)
46. Services Innovation and the Knowledge Economy (22-23 April 1999: Manchester)
45. Knowledge Transfer in the Information Society (19-20 November, 1998: Budapest)
44. R&T Co-operation with Central and eastern European Countries (25-16 May, 1998: Vienna)
43. Local Innovation - New Social, Economic and Technological Dimensions (November 1997: Dublin)
42. The past and Future of EU Research Funding - Framework programmes under consideration (16-17 June 1997: Helsinki)
41. Innovation and Sustainable Development: Lessons for Innovation Policies (November 1996: Eastern Germany)
40. R&D Subsidies at stake: In search of a rationale for public funding of industrial R&D (18-19 April, 1996: Gent)
39. Re-Inventing the Future? Foresight at the edge of chaos November 1995: Stockholm.
38. Technology, Innovation and (un) Employment 15-16 June, 1995: Rotterdam.
37. Research Co-operation with Countries in Transition (1-2 December 1994: Vienna).
36. Innovation - Applying New Ideas for Profit (25-26 May 1994: London).
35. Financing the Early Stage technology Company in the 1990's an international perspective (8-9 November 1993: Montreal).
34. National Networks to Global Links (21-22 June 1993: Dublin).
33. User-producer relations in the innovation process (26-27 November 1992: Helsinki)
32. Lean Company: Model and Practice, Context and Policy (11-12 May 1992: Stuttgart).
31. Technical Competence and Firm Strategy - Implications for Public Policy (5-6 November 1991: Stockholm).
30. Innovation and Society: in Search of New Linkages (16-17 April 1991: Noordwijk aan Zee).
29. Technological Innovation and the Service Sector (23-24 October 1990: Vienna).
28. R&D Organizations in the 1990's (14-15 May 1990: London).
27. The Internationalization of R&D (16-17 October 1989: Ottawa).
26. The future of the Rural Economy (22-23may 1989: Shannon).
25. Interfirm Innovation Dynamics (October 3-4 1988: Stuttgart).
24. Regional Innovation Supporting Services (December 7-8 1987: The Hague).
23. Investment in training (April 23-25 1987: Paris).
22. Home Information's (November 3-4 1986 London).
21. Engineering education for Industrial Development (June 9 - 10 June 1986: Stockholm).
20. Financing of Entrepreneurial Ventures and Innovation (30 September - 1 October 1985: Vienna).
19. International technology transfer: Promotion and Barriers (May 6-7 1985:

Ottawa).

18. Regional Innovation Policies and Programmes (November 19-20 1984: Limerick).
17. Policies for industrial Innovations - Situations and Perspectives (May 7-8 1984: Bonn).
16. Technological Information from abroad: scientific attaches and other public institutions, industrial needs and public policies (December 1983: The Hague).
15. Technological culture and success in industry (6-7 June 1983: Paris).
14. Evaluating the effectiveness of government innovation policies (22-23 November 1982: London).
13. Industry university relations (10-11 May 1982: Stockholm).
12. A systematic approach to innovation (23-24 November 1981: The Hague).
11. Regional innovation policy: technology policies or regional policies (1-2June 1981: Sophia Antipolis).
10. New entrepreneurship and the smaller innovative firm (9-10 June 1980:Limerick).
9. National innovation policies (1-2 December 1980: Berlin).
8. Trends in collective industrial research (26-27 November 1979: London).
7. Industrial Innovation and Government regulation (11-13June 1979: The Hague).
6. Technical change and employment (13-14 November 1978: Paris).
5. Government procurement policies and industrial innovation (6-7 June 1978: Dublin).
4. Small and medium sized manufacturing firms: their role and problems in innovation, government policy in Europe, the USA, Canada, Japan and Israel (21-22 November 1977:The Hague).
3. The current international economic climate and policies for technical innovation (November 1977).
2. Government aid to technical change in the mechanical engineering industries (19 April 1977: Karlsruhe)
1. Government direct financial assistance to industry, programmes, experiences and trends (12-13 October 1976: London)

◇ 最近のイノベーション政策関連の資料紹介として

R&D INVESTMENT IS KEY TO INNOVATION AND ECONOMIC GROWTH

英国 DTI (Department of Trade and Industry) が出版した「the Future and Innovation Unit of the UK's 2001 (http://www.c3e.fr/Inteval/)R&D Scoreboard」は、企業が潜在的な成長力を十全に発揮するためには研究開発投資が決定的に重要であることを明らかにした。

(3) International Evaluation Research Group (INTEVAL)

http://www.c3e.fr/Inteval/

◇ 目的と活動の概要

政策・プログラム評価の国際的研究グループである INTEVAL は、1986 年以来、毎年、各

国政府、会計検査機関、大学、民間セクターに属す評価専門家と会合を開催している。現在メンバーは20ヶ国以上にわたる。

国際的な視野に立って、電子メールやサブグループによる年次会合によって研究をすすめている。結果は、共著の書籍や Comparative Policy Analysis Series (Transaction Publishers)のシリーズとして出版される。

15年間の活動によって、研究グループは評価の理論と実践に関してかなりの量の資料を収集している。

◇ 研究グループの会合歴

1986年にブリュッセルで最初の会合を開いた。以後9年間はブリュッセルの International Institute of Administrative Sciences の支援を受けた。政策評価の理論と実践において国際的な比較研究をどう活用するかが焦点であった。

第1期のグループメンバーは、Dr. Ray C. Rist, Prof.dr. H.-U. Derlien, Germany, Prof.dr. R.V. Segsworth, Canada, Prof.dr. W. Jenkins and Prof. A. Gray, UK, Dr. E. Albaek and Dr. S. Winter, Denmark, B. Eriksen, Norway and Dr. G. Arvidson, Sweden plus the only founding mother Dr. M.L. Bemelmans-Videc.である。

以後の会合は次のとおりである。

- Brussels, 1986 hosted by IIAS
- Paris, 1987 hosted by the Ministry of Telecommunication and Post
- Leiden, 1988, hosted by the State University of Leyden
- Aarhus, 1989, hosted by the University of Aarhus
- Stockholm, 1990, hosted by the Swedish National Audit Office
- Bruges, 1991, hosted by the IIAS
- Ottawa, 1992, hosted by the Office of the Comptroller General, Canada
- Oslo, 1993, hosted by the Nowegian Government
- Lyon, 1994, hosted by the Centre for European Evaluation Expertise
- Madrid, 1996, hosted by the Institute of Fiscal Studies, Autonomous University of Madrid
- Seoul, 1995, hosted by the Hankuk University of Foreign Studies
- Washington, 1997, hosted by the World Bank
- Ireland, 1998, hosted by the Institute of Public Administration, Dublin
- Durham, 1999, hosted by the Durham University Business School
- Haifa, 2000, hosted by Haifa University.
- Stockholm, 2001 hosted by the Swedish Court of Auditors

◇ これまでの出版物

- COMPARATIVE POLICY ANALYSIS SERIES ,Edited by Ray C. Rist, Transaction Publishers.
- CARROTS, STICKS, AND SERMONS : Policy Instruments and Their Evaluation, Marie-Lousie Bemelmans-Videc, *Ray C. Rist, and Evert Vedung, Editors.*
- PROGRAM EVALUATION AND THE MANAGEMENT OF GOVERNMENT, *Ray C. Rist, Editor.*
- PUBLIC POLICY AND PROGRAM EVALUATION, *Evert Vedung.*

- MONITORING PERFORMANCE IN THE PUBLIC SECTOR, *John Mayne and Eduardo Zapico-Go, Editors . With a foreword by Joseph S. Wholey.*
- POLITICS AND PRACTICES OF INTERGOVERNMENTAL EVALUATION, *Jacques Toulemonde and Olaf Rieper, Editors . With a foreword by Ray. C. Rist.*
- BUDGETING, AUDITING, EVALUATING : Functions and Intergration in Seven Governments, *Andrew Gray, Bill Jenkins, and Bob Segsworth, Editors . With a foreword by Ray. C. Rist.*
- CAN GOVERNMENTS LEARN?, *Frans. L. Leeuw, Ray C. Rist, and Richard Sonnichsen, Editors.*
- BUILDING EFFECTIVE EVALUATION CAPACITY : Lessons From Practice, *Richard Boyle, Donald Lemaire, Editors.*
- EL SEGUIMIENTO DE LOS RESULTADOS DE LA GESTION EN EL SECTOR PUBLICO, *J. Mayne, E. Zapico.*
- CARROTS, STICKS, AND SERMONS : Policy Instruments and Their Evaluation *Marie-Louise Bemelmans-Videc, Ray C. Rist, and Evert Vedung, Editors .*
- PROGRAM EVALUATION AND THE MANAGEMENT OF GOVERNMENT, *Ray C. Rist, Editor.*
- PUBLIC POLICY AND PROGRAM EVALUATION, *Evert Vedung.*
- MONITORING PERFORMANCE IN THE PUBLIC SECTOR, *John Mayne and Eduardo Zapico-Go, Editors . With a foreword by Joseph S. Wholey.*
- POLITICS AND PRACTICES OF INTERGOVERNMENTAL EVALUATION, *Jacques Toulemonde and Olaf Rieper, Editors. With a foreword by Ray. C. Rist.*
- BUDGETING, AUDITING, EVALUATING : Functions and Intergration in Seven Governments, *Andrew Gray, Bill Jenkins, and Bob Segsworth, Editors.*
- CAN GOVERNMENTS LEARN? *Frans. L. Leeuw, Ray C. Rist, and Richard Sonnichsen, Editors.*
- BUILDING EFFECTIVE EVALUATION CAPACITY : Lessons From Practice, *Richard Boyle, Donald Lemaire, Editors.*
- Carrots, Sticks & Sermons; Policy Instruments & Their Evaluation Edited by *Marie-Louise Bemelmans-Videc, Ray C. Rist and Evert Vedung,* Transaction Publishers, New Brunswick (U.S.A.) and London (U.K.)
- PROGRAM EVALUATION AND THE MANAGEMENT OF GOVERNMENT, *Ray C. Rist, Editor.*
- The Politics and Practices of Inter-Governmental Evaluation, *Olaf Rieper and Jacques Toulemonde, Editors.*
- Budgeting, Auditing, and Evaluation - Functions and Integration In Seven Governments, *Andrew Gray, Bill Jenkins, Bob Segsworth, With a foreword by : Ray C. Rist.*
- CAN GOVERNMENTS LEARN? Comparative Perspectives on Evaluation & Organization Learning, *Edited by Frans L. Leeuw, Ray C. Rist& Richard C. Sonnichsen.*
- Building Effective Evaluation Capacity: Lessons From Practice, *Richard Boyle, Donald Lemaire, Editors.*

◇ 現在進行中の出版活動

次の４冊の作成が進行中である。

ｉ）THE INTERNATIONAL EVALUATION ATLAS

Jan-Eric Furubo, Rolf Sandahl and Ray Rist, Editors.

・20ヵ国ならびに国際機関を取り上げ、そこでの評価システムを総覧している。

ii）EXPLORING EVALUATION AND COLLABORATIVE GOVERNMENT

Andrew Gray, Bill Jenkins, Frans Leeuw and John Mayne, Editors.

・政府がサービスを提供する際に、公的、民間ならびにボランタリーな組織との共働が次第に求められるようになってきた。そうした流れの中で、政策とプログラムの評価はどう変化すべきなのかを探った。

iii）ASSESSING EVALUATIVE INFORMATION: prospects and pitfalls.

Reuven Schwartz, John Mayne, Jacques Toulemonde, Editors.

・信頼性・妥当性の高い評価作業を実施しようとすればするほど、社会全体に‘査察’の目を張り巡らせ、息苦しさを生んでしまうという矛盾が生じる。この矛盾を解消できる評価のあり方を探る。

iv）EVALUATION UTILIZATION AND KNOWLEDGE MANAGEMENT : WHERE WE ARE WITH THE EVALUATION UTILIZATION DEBATE

Ray C. Rist, Marcus Spinatsch, & Nicoletta Stame , Editors.

＜この先の出版予定＞

- ASSESSING EVALUATIVE INFORMATION: prospects and pitfalls
  Reuven Schwartz, John Mayne, Jacques Toulemonde, Editors.
- EVALUATION UTILIZATION AND KNOWLEDGE MANAGEMENT (tentative title) *Ray Rist, Markus Spinatsch, Nicoletta Stame, Editors.*

◇ INTEVAL のメンバー

- Erik Albaek
  Department of Economics, Politics & Public Administration
  Aalborg University（DENMARK）
- Goran Arvidsson
  SNS Center for Business & Policy Studies（SWEDEN）
- Per Oyvind Bastoe
  Managing Director,ErgoDialog AS (NORWAY)
- Jean-Claude Barbier
  Director of Research（CNRS）Centre d'Etudes de l'Emploi"Le Descartes"（FRANCE）
- Marie Louise Bemelmans-Videc
  (THE NETHERLANDS)
- Richard Boyle
  Institute of Public Administration（IRELAND）
- Hans-Ulrich Derlien
  Professor of Administrative Science, University of Bamberg（GERMANY）

- Stan Divorski
  Assistant Director,Center for Evaluation Methods and Issues, US GAO (USA)
- Jan-Eric Furubo
  The Swedish National Audit Bureau (SWEDEN)
- Andrew Gray
  The Centre for Public Sector Management Research,Durham University Business School (UK)
- Katia Horber-Papazian
  Professor, Institut des Hautes Etudes en Administration Publique (SWITZERLAND)
- William Jenkins
  Darwin College,University of Kent, Cantebury (UK)
- Myoung-Soo Kim
  Professor of Public Administration, Hankuk University of Foreign Studies (REPUBLIC OF KOREA)
- Stefan Kuhlman
  Fraunhofer Institute, Systemtechnik und Innovationsforschung (ISI) (GERMANY)
- Yoon-Shik Lee
  Professor and Vice President for Academic Affairs, Department of Public Administration, Soongsil University (REPUBLIC OF KOREA)
- Frans L. Leeuw
  Chief Review Officer, Open University and Utrecht University (THE NETHERLANDS)
- Donald Lemaire
  Senior Regional Director, Quebec Regional Office, Department of Justice Canada (CANADA)
- Jeremy Lousdale
  National Audit Office (United Kingdom)
- John Mayne
  Principal, Audit Operations, Office of the Auditor General (CANADA)
- Ragnhild Ovrelid
  Directorate of Public Management (NORWAY)
- Dale H. Poel
  School of Public Administration, Dalhousie University (CANADA)
- Olaf Rieper
  Associate Professor, AKF - Institute of Local Government Studies-Denmark (DENMARK)
- Ray Rist
  c/o Operations Evaluation Department,The World Bank (UNITED STATES OF AMERICA)
- Rolf Sandahl
  The Swedish National Audit Bureau (SWEDEN)
- Robert (Reuven) Schwartz
  Department of Political Science,University of Haifa (ISRAEL)
- Robert Segsworth

Department of Political Science,Laurentian University (CANADA)

- Nicoletta Stame
  Universita di Roma "La Sapienza" (ITALY)
- Markus Spinatsch
  Consulting for Public Policy and Administration (SWITZERLAND)
- Hilkka Summa
  Directorate-General for Budgets (BELGIUM)
- Jacques Toulemonde
  Center for European Evaluation Expertise (FRANCE)
- Marit Waerness
  Directorate of Public Management (NORWAY)
- Eduardo Zapico-Goni
  Deputy Director of Policy Analysis, Directorate of the Budget, Ministry of Finances (SPAIN)

5.2 主な評価研究者リスト

◇ Stefan Kuhlmann (Fraunhofer ISI, Germany)
sk@isi.fhg.de
http://www.isi.fhg.de/ti/Lebenslaeufe/deutsch/sk-ll-d.html

◇ Philip Shapira (Georgia Institute of Technology, USA)
philip.shapira@pubpolicy.gatech.edu
http://www.cherry.gatech.edu/ps/

◇ Irwin Feller (Pennsylvania State University, USA)
iqf@psu.edu
http://econ.la.psu.edu/BiographyPages/feller_bio.html

◇ Arie Rip (University of Twente, The Netherlands)
A.Rip@WMW.UTwente.NL
http://www.ctit.utwente.nl/programme/disciplines/pts2.html

◇ Luke Georghiou (PREST, University of Manchester, UK)
Luke.Georghiou@man.ac.uk
http://les.man.ac.uk/PREST/People/Staff/Luke_Georghiou.html

◇ Susan E. Cozzens (Georgia Institute of Technology, USA)
susan.cozzens@pubpolicy.gatech.edu
http://www.spp.gatech.edu/people/faculty/scozzens.htm

◇ Laurent Bach (BETA, University Louis Pasteur, France)
bach@cournot.u-strasbg.fr
http://irist.u-strasbg.fr/dea.htm

◇ Terttu Luukkonen (ELTA, Finland)
terttu.luukkonen@etla.fi
http://www.etla.fi/english/people/searchengine/index.php3?suku=Luukkonen&etu=Terttu

◇ Barry Bozeman (Georgia Institute of Technology, USA)
barry.bozeman@pubpolicy.gatech.edu
http://rvm.pp.gatech.edu/goco/goco_biosketch_bb.html

◇ Ken Guy (Wise Guys, UK)
ken.guy@wiseguys.ltd.uk
http://www.wiseguys.ltd.uk/

◇ David F.J. Campbell (IHS, Institute for Advanced Studies, Austria)
campbell@ihs.ac.at
http://www.ihs.ac.at/index.php3?hr=47

◇ Erik Arnold (Technopolis, Innovation Policy Research Associates)
erik.arnold@technopolis-group.com
http://www.technopolis-group.com/staff/index.htm#ea

◇ Anthony van Raan (University of Leiden, The Netherlands)
vanraan@cwts.leidenuniv.nl
http://leidsewetenschappers.leidenuniv.nl/show_en.php3?medewerker_id=428

◇ Michel Callon (Ecole des Mines de Paris)
callon@csi.ensmp.fr
http://www.ensmp.fr/cgi-bin/whoswho?Qid=679

◇ Philippe Laredo (Ecole des Mines de Paris)
laredo@csi.ensmp.fr
http://www.ensmp.fr/cgi-bin/whoswho?Qid=687

◇ Frieder Meyer-Krahmer (Fraunhofer ISI, Germany)
fmk@isi.fhg.de
http://www.isi.fhg.de/il/institutsleiter.htm

◇ J. David Roessner (SRI International) (Georgia Institute of Technology, USA)
roessner@wdc.sri.com

david.roessner@pubpolicy.gatech.edu
http://www.spp.gatech.edu/people/faculty/droessner.htm

◇ Ben Martin (University of Sussex at Brighton
SPRU, Science and Technology Policy Research)
B.Martin@sussex.ac.uk
http://www.sussex.ac.uk/spru/people/bmartin.html