

「研究開発評価システムの在り方に関する
検討ワーキンググループ(H24.2.20～H24.7.3)」における主要な意見等

平成 24 年 8 月 31 日
評価専門調査会事務局

(1) プログラムの概念とプログラムに向けた手順

- 政策・施策を具体的に実現するための一体的な論理的なまとまりをプログラムと称し、それを具体的に実行していくに当たって、その中でプロジェクトを実施するという形になるのではないか。個々のプロジェクトを足し合わせればそれがプログラムとして成立するわけではない。
- プログラムは施策の最小単位であると捉えられるが、プロジェクトは期間内に研究を遂行し成果を出すという研究者が主体なので、施策の単位ではない。
- プログラムや制度を設計するとき、そのプログラムや制度のアウトカムを考えておいて、そのプログラムに最適なプロジェクトの採択なり形成をするというプロセスを入れていかなければいけない。
- 科学技術的な価値を目指すプログラムと社会的・経済的な価値を実現するところまでを目指すプログラムは、タイプが違うので、この2つを共通で扱えるような枠組みを考える必要がある。
- イノベーションは経済的な価値を生むという意味で使われているが、基礎研究のように直ちに価値を生むものではない研究もあるのだから、体系的、プログラム化しないと基礎研究もできないと思われぬように配慮が必要ではないか。
- プログラムといっても、科学的な価値の創出を目指すタイプと社会経済的な価値の実現を目指すタイプがある。これまで基礎研究の場合は、プログラムとしてなかなか評価されてきていないが、基礎研究を国として振興するという明確な政策目標を持っているわけで、そういう意味ではどのタイプも政策的な目標を持っているというとらえ方をするのが妥当。
- プロジェクト単位での評価が難しい基礎研究の場合難しいところはあるが、プログラムという見方をすることによって制度全体の健全さをきちんと図る、個々のプロジェクトの成否にとらわれないということで、むしろ基礎研究こそプログラム評価が必要ではないか。

- 基礎的研究の意義は、ある確率での大発明や真理の発見、人の育成などであり、それらを長期間モニタする必要がある。そういうことを事前に設計しておき、制度そのものが長いレンジで見たときに本当に初期の目的を達成しているかというプログラム評価的なものがあり、それがもっと長いレンジでは最終的な実用的なものに繋がっていくので、短期的なものの見方より基礎研究ほどプログラムとか制度そのものをしっかり見ていくことがより大切になる。そういう意味では経済価値と科学的価値を峻別する必要もなくやれるのではないか。
- 現状では、今あるプロジェクトをどういうふうにプログラムとしてまとめていくかという作業が必要となり、プロジェクトのプログラム化というプロセスが入ってくる。逆にプログラムというものを初めから設定するのであれば、スタートする前に議論をする必要があり、その中にプロジェクトがきちんと入っていくといいのではないか。
- プログラムというのは研究資金制度ではないことを明確にした方が良い。
- 資金配分制度の下で採択される個別課題とこの事業そのものとの連動性があまりないのが多い。個別の課題の優劣だけで採否を決めてしまうと、同じ傘の下にいるにも係らず、プロジェクト間の相関性が薄くなる。プログラムとしての改善の方向性と、それと連動する形でその下にある個別プロジェクトが選ばれないといけませんが、日本の場合は必ずしもそうっていない場合がある。
- プログラム評価をやることによって重要なのは、それに向けていい研究が集まり、そのプロセスがきちんと事業の目的に向かって進んでいるかということシステム化できること。事業化の確率を上げることは、その結果として上がることが必要であって、初めから事業化できるものだけを採択すればいいということにはならない。
- 研究資金制度のプログラム化というのは、例えば、CRESTならCRESTの一つのテーマは、当然ながらプログラム化していかなければいけない。もしそのプログラム同士に相互の関係があった場合に、全体としてどういう方向に行くかということ自体もプログラム化しておく必要がある。
- 今ある国家プロジェクトみたいなものをプログラム化するというのは、そこにどうやって成果を届けますかという、成果を実現しますか、そこまでちゃんとプロジェクトとして考えていますか、その考える仕組みをきちっと折り込むということがプログラム化ということ。
- プログラムをゴールまでもっていくためには、そのプログラムの中で全部やってしまうのではなく、さらに他の施策に有機的に結びつけていって、最終的なゴールに到達

するという設計が必要ではないか。

- 生まれてくる中間的なアウトカムについて、そのままに放っておいたのでは成果が意図したところにつながらない。行政施策を埋め込んでいくという形で、補助的に新しい助成を行ったり、優等的な措置として税制とか規制といったもので補ったりしていくことで意図した成果までつなげていくことが必要。
- 途中の行政施策を入れていかないと、最終的なアウトカムは出てこないが、そうってから入れるというのではなく、できたらプログラムを開始するときにそれを想定して、そこまでやらないと最終的にはいけないというのを埋め込んでおくことが必要。
- イギリスの Alvey Programme の追跡評価で得られた教訓は、IT産業の再活性化という商業的な目的を実現しようと思うのならば、本来総合的な産業政策として対応する必要があったが、その一部を担うにすぎない研究開発制度だけを試みた。つまり、目的と手段が乖離しているところを埋めるための工夫がなされていなかった。具体的には、補助装置として、その他の研究開発投資や政府調達、税制、規制といったようなことを行政施策として組み合わせる必要が本来あったのではないかということ。それから、評価可能な形で目的、目標を最初の段階で構築しておく必要があったのではないかということ。
- 研究開発とほかの行政施策をカップルにすることはプロジェクトの研究者ではできない。従って、補助金を出すなり産業政策に組み入れるなりして、上位の政策あるいは施策の責任者である部分に背負ってもらうことが、今回のイノベーション施策の大きなポイント。
- 他省と連携をするようなときにはもう少し大きなプログラムが必要で、そのプログラム同士に相互の関係があった場合には全体としてどういう方向に行くかということもプログラム化すべき。
- アウトカムを考えるときに政策パッケージみたいなことを考えないと、実現できないことが多いが、省庁をまたがるような話を一体誰がコントロールできるのかという最初のとっかかりが難しい。

(2) プログラム評価の視点

- プロジェクト及びプログラム各々の評価に当たっては、プロジェクトの場合、元々決められていた目標が内在的な要因に基づいて達成されたかを見ればよいのに対して、プログラムの場合は政策・施策の責任者ではない人がその成果を受けてイノベーショ

ンにつなげていくので、外的要因というのをきちんと踏まえた上で、そこへの達成に向けた活動というのが妥当であったかどうかを見る必要がある。

○プログラムレベルでなぜ評価をするのかということ、研究の現場の責任ではなくて、それに助成して政策目標を何かしら達成しようとした、お金を出した仕組み自身が妥当だったのかどうかというのをみるのは、政策としての評価をするというのが意味のあるところ。○重要なことは個別のプロジェクトの成果は研究者が評価されるが、プログラム評価になる、研究者ではなく、そういう制度を設計する行政の担当者であり、決定的に違う。従って、プログラム化といったときに、誰が評価されるのかという被評価者の対象が大きく変わるということを明示的に入れておく必要がある。

○プログラム評価は、そのプログラムをつくった行政の責任。アウトプットは、研究者のアクティビティの評価なので、評価対象は研究者であり全く違う。そう考えると、行政施策は必然的にうまく機能しないとアウトカムにつながらない。

○例えば、規制緩和など、法律を改正しないといけないようなことは簡単なことではできないので、行政施策で言われる内容によってどう評価するかというのは、運用の中で個別に考えていく必要がある。

○そのプログラムだけではなく、いろいろな外部の要素を効果的にジョイントしてこういうのができましたというものをインパクト評価やアウトカム評価のときに入れていくことが重要となる。

○基礎研究であってもプログラムとして扱うことは可能であり、評価もできる。例えば科研費の場合でも、すぐれた研究に対してファンディングすることで研究の裾野を広げるとか、それを下支えするといったプログラムとしての目的はある。それが本当にできているのか、それを達成するためのプログラムとして事前、採択審査の仕組みをきちっと持っているのか、それがまずければ見直す仕組みを持っているのかといった観点でのプログラム評価というのはできる。その場合のアウトカムというのは、裾野をきちんと広げているかというようなところを表現するような指標の設定が考えられる。

○重要なのは、プログラムの目標が達成できることであるが、そこに至るシナリオをきちんと設定してそれを評価する必要がある。事前評価においては、そのプログラムの意義と目標、アウトカムを明確にしているか、シナリオがきちんと論理的な整合性をもって書かれているか、ロードマップやマイルストーンが提示してできているかを評価する。中間評価においては、ロードマップに沿った研究推進になっているか、シナリオやロードマップの見直しは必要ないかを評価する。終了評価においては、アウト

カムにつながると考えられる研究成果が得られているのか、目標の設定やシナリオ、ロードマップは適切であったのかを評価する。追跡評価においては、アウトカムができたかどうか、できなかったとしたら、その理由は何なのか、その検討がフィードバックできたかということの評価する。

○研究開発自体不確実性が高いので、あくまでも戦略計画を目的、目標として設定する際に、仮説としてそれをきちっと持っておく。仮説が間違っていれば修正する。決められた枠組みについていかに達成したかといったようなことだけを見ていくと、現実が変わっていくのに変化する前の目標に固執してしまい、間違っただけの問題を解くという事態になってしまう。あくまでも仮説としてきちっと持っていることが重要である。

○アメリカのPARTの場合は、成果を50%見るというところと、戦略計画をきちっと立てているのかということを含めて、マネジメントなど成果を実現するための仕組みがどれぐらいきちっと設定されているのかといったことが50%となっている。

(3) アウトプット、アウトカム等の定義とこれに係る指標の設定の考え方

○アウトプットは、研究開発の現象的ないし形式的側面であり、例えば論文ができたとか、設計図ができたとか、規格原案を提出したとか、特許出願をしたということで、ある意味、数等で評価できる。

例えば、プログラムとしてどのような活動の水準を達成していればいいのかということでアウトプットを見ていくわけであるが、基盤研究の場合は、応募資格者数のうちどれぐらいの応募数があったのか、それによって多様性がどれぐらい確保されるための努力をしているのか、研究分野のポートフォリオが適切だったのかという判定もある。

○アウトカムは、研究開発成果の本質的ないし内容的側面ということで、例えば、学術論文については科学技術コミュニティでの評価を経た内容になっているのか、あるいは、科学技術コミュニティが発展したのか。社会的・経済的な効果の場合には、具体的な製品あるいはサービス、あるいは、その価値的な内容となる。

プログラムとして設定した目標を具現化したもの、それを代理的な指標としてあらわしたものが、アウトカムととらえるのがよい。この場合、パフォーマンスの結果としての指標ではなくて、指標につながるようなプロセスをアウトカムとしてとらえるといった場合もある。

○研究開発型プログラムにおけるアウトカムの例としては、

①基盤研究のような場合、助成された研究プロジェクトと助成されなかった研究プロ

ジェクトを比較した場合、生産された論文の被引用度の割合の差によって、どれぐらい質の高い研究にお金を投じているのか、それによって多様性を確保される可能性がどれぐらいあるのか。

②若手研究者の自立支援ということであれば、賞を受けていない若手研究者との比較において、テニュアトラックを獲得した研究者の割合がどれぐらい高いのかといったものによって、プログラムの意図した結果が達成されているか、次のステージにステップアップするような助成金を獲得できた割合がどれくらいあったのか。

③新領域では、新領域への研究者の参入数の推移やその領域発展にどれぐらいこれが寄与しているのかといった度合い。

などが考えられる。

○インパクトは、研究開発者の手を離れた後、間接的成果による波及効果である。但し、アウトカムを比較的近い、アウトカムを研究開発者が直接的に寄与している直接アウトカムと、少し時間がたって研究開発者の寄与が間接的である間接アウトカムというのに分けることもある。

○NSFが用いているアセスメントフレームワークでは、ニアタームとミッドターム、ロングタームということでアウトカムを設定している。また、アメリカのエネルギー省の風力エネルギープログラムでは、4つの研究開発プロジェクトが走っていて、それを関連づける形でプログラムとしての体をなしている。ショートターム、インターミディエット、ロングタームという形でアウトカムが展開していくという形になっており、ショートタームだと初年度から3年間の間、中間だと3年後から約10年後、ロングタームだとそれ以上の間という形で、マイルストーンを設定している。

○例えば5年のプロジェクトの5年後に、さらに5年たった後に出てくるアウトカムをどれくらい設定するかというのは現実問題として難しいところがあるが、FSの議論を踏まえておいて、その時点で一定の判断をする、10年後にどうなるかなど誰もわからないのだから、その時点でこれだけの議論をしておいたということをしちんとしておく必要がある。

○アメリカ流(PARTモデル)の実績概念と欧州流の実績概念の区分は若干異なっている。アメリカでは、自分たちが確実に責任を持つというところをアウトカムと設定しているのに対し、欧州ではアメリカでいうアウトカムみたいなところを含めて成果の全てを包含する概念としてインパクトといった表現を使う場合がある。

○アメリカの PART の枠組みの中では研究開発は非常に不確実性が高いということがあって、特に定量的な指標を用いなくてもいいという言い方をしている。ただし、用いない場合はなぜ用いないのか。他にそれよりも価値を測る良いやり方があるということを示さないといけない。

(4) 追跡評価について

○プログラムの見直しやプログラムの仕組み自身を改善していくために追跡評価をやるということであり、研究開発上の課題だけではなく、いろいろな実現を阻害するような要因とか課題みたいなことを発見したり、そのことによって施策を見直していくというために使われるものもある。意図していなかった重要な成果の芽をきちっとそこでも見つけていくといった意義もある。そういった意味で何のために追跡評価をするのかというのをきちっと明記したほうがよい。

○海外の研究機関は、それぞれの組織が自分たちの存在意義や能力を外にアピールすることで、自分たちの組織の存続のためには必須のものだという認識を持って評価を行っている。そのために、各自、各機関が工夫をしている。透明性をいかに確保できるか、自分たちの仕組みを見直すための実効性をどうやって確保するか、各自非常に工夫をしている。

○アウトプットがどうアウトカムとしてつながっていくかというところを評価していく際には、実際にだれがどうやっていくのか、外注に出すといった点も含めて人的・予算的な面からも考える必要がある。

○研究者と所属する機関との関係が非常に曖昧なまま成果が研究者個人に帰属してしまっていて、研究者の人事異動があると後で追跡調査できないということは重要な検討課題である。

○どの評価の場合でもとっておくべき情報、データは共通するものがあり、そういうデータを当初の段階からある程度定型化し、データを蓄積していけるような体制になれば、そのデータを基に評価をしてもらうという形で効率化はかなりできてくる。

○海外では、プロジェクトやプログラムの実施期間中にどのようなデータを集めていけばいいのかを事前の段階で十分に検討しており、プログラム終了後も情報収集に協力してもらうことを助成や委託の条件にする、といったような取組も行っている。

○通常、追跡調査は、実用化を目指す研究を対象とすると思うが、イギリスの事例ではそれのみだけではなくて、基礎研究も対象になる。海外では、ファンディングプログ

ラムも個々のプロジェクトの評価の行く末を見ることによって資金運用の妥当性を評価する仕組みがある。

○追跡評価を設計するときに、どこまでをプログラムの責任として予め問うのかということをしちんとしておいた方がいい。基礎研究の場合には、あまり明確なアウトカムがない代わりに、インパクトを評価するというのが1つの考え方。プログラムをある施策の実質的な実施と考えた場合には、やはりアウトカムを明確に定義して、もちろんインパクトのプラス α はあってもよい。

○イノベーションに至るパスの設計という意味では、追跡評価のモデルとか考え方というのは、そのままプログラムの設計あるいはイノベーションに至るシナリオの設計とある種対をなすようなものであって、事前の段階である程度こういう方法論に基づいてプログラムを作り、同じ方法論の土俵の上で追跡評価で評価するという形になると、プログラム、プロジェクトを実施する人の中にも頭と終わりのところの一貫性がある、非常にいいのではないか。その際、きっちりしたゴールだけでなく、確率的なものも含めて、そういった幅を持ったようなシナリオの中で同様のモデルで事前と事後を対照させていくというような、例えば企業経営でもバランスコアカードとか、同じようなモデルに基づいたようなPDCAの一種のモデルだと思うが、そのようなやり方もあるのではないか。

○海外では、評価結果の活用に関しても、評価そのものは意思決定ではないが、その結果をどのように扱ったのかということについて明示し、それを公表していくということで活用を促進するような仕組みにしている。

(5) 評価体制(外部評価)について

○「外部評価」という言葉は、外部を活用した評価という大きな意味でとらえるべきと考える。1つは、評価の内容に専門性を持つ外部の有識者から構成される委員会などが実際に評価を行う場合、もう一つは、評価に必要な調査・分析に専門性を持つアナリストのような方々に関わる場合である。前者の有識者の場合は、評価対象についての専門性を持つということなので、これらの専門家は評価者として特段の訓練を受ける必要はなく、育成の対象にはならないが、一方、後者の調査・分析の専門家は、育成をする対象になりえると思う。

○被評価者が事業推進主体であれ、研究開発実施主体であれ、目的が自己学習の場合以外、被評価者が自ら評価を行う自己評価のみでは完全な評価にはならない。

○アメリカのPARTに基づく評価は査定というのを目的にした評価であり、被評価者

が連邦政府機関となる。これは被評価者が外部の専門家等を活用して行った自己評価をもとに、第三者である一査定部門である行政管理予算局OMBがメタな立場で評価をする。具体的には、プログラムの目的と設計、戦略計画、プログラム・マネジメント、プログラムの結果と説明責任といったようなカテゴリーに対して、各連邦機関が申し開きをしたもの、自己分析をしたものについて、それが妥当かどうかということをも別の角度から検証していくというものになっている。

○カナダのNSERC(自然科学・工学研究会議)の事例では、まとめられた勧告(外部評価結果)に対して受け取った側が同意するのかもしれないのか、どのような行動をそれに対して起こすのか、その結果を起こす責任者は誰なのか。それをいつまでにちゃんとやるのかといったようなアクションプランを作成して公開している。また、その評価のレポートも用いた方法論からデータまで公表されているので、外部の第三者から検証可能となっている。

○海外事例から得られた示唆は以下の3点である。

- ①被評価者は行政が中心である。研究成果の評価ではなく、政策目標に対する寄与の観点から、研究への助成、委託の仕組み等の評価すべき。
- ②チェック&バランスが重要であり、多様で階層的な評価が前提となる。例えば、採択審査をするところで1つプロジェクトの評価が入り、その評価に対して、評価のやり方が妥当だったのかというメタな評価というように、階層的な評価が重なっている。何か1回やれば十分というわけではなくて、別のタイプ、別のフェーズの評価がきちんと入っている。
- ③低次元のアカウントビリティから高次元のアカウントビリティ、レスポンシビリティへと拡大する必要がある。1つは評価結果の活用状況については、海外ではきちんと公開する。もう一つは、透明性として、外からの検証可能性が確保されていることが重要で、あわせて体制の独立性も必要。ただ単に結果を公開したり、外部を活用すれば透明性が確保されるということではない。

○海外のシンクタンクと日本のシンクタンクというのは位置付けやステータスみたいなところは若干違うところがあり、海外のシンクタンクは非常に高度な専門性を持った人たちがやっている。

○シンクタンクによる評価は1社ないし2社が(独占的に)やるので、中身がブラックボックスにならないような工夫もされている。例えば、そこで使ったデータやまとめた報告書全てのものについては開示をしているので、検証が可能。

○海外では、機関評価のようなものを行う場合でもプログラムというのが単位になっており、プログラム評価というところをある意味積み上げていく。その累積がいわばそ

の機関の機関評価につながるというこの括りが非常に大きなプログラム評価の意味合いである。我々の頭の中は、施策、プロジェクト、プログラム・制度とそれを実際に推進している機関とを明確に分けたがるが、実態は明確に分かれるものではなく、そういった施策制度やプログラムを推進する主体としての機関があるわけなので、そこはある意味では密接不可分である。