

総合科学技術会議
第 3 4 回評価専門調査会議事概要（案）

日 時：平成 1 6 年 3 月 2 6 日（金） 9：00～11：50

場 所：中央合同庁舎 4 号館 第 4 特別会議室（4 階）

出席者：大山会長、阿部議員、薬師寺議員、岸本議員、黒田議員、
松本議員、黒川議員

秋元委員、江崎委員、大見委員、加藤委員、國井委員、国武委員、
末松委員、中西委員、馬場委員、畚野委員、増本委員

欠席者：石田委員、伊丹委員、市川委員、大石委員、鈴木委員、谷口委員、
寺田委員、藤野委員

招聘者：平澤 冷（東京大学名誉教授）

デービッド・トリンクル（米国行政管理予算局プログラムエグザミナー）

ルーク・ジョージュー（英国マンチェスター大学工業科学技術政策
研究所所長）

議 事：1. 評価における今後の課題と改善方向の検討について（議題 1）

（1）海外研究開発関連機関等の調査報告（1）

海外の公的支援研究開発の評価システムからの含意

（2）海外研究開発機関等からのヒアリング

デービッド・トリンクル氏

（米国行政管理予算局プログラムエグザミナー）

ルーク・ジョージュー氏

（英国マンチェスター大学工学科学技術政策研究所所
長）

（3）評価における問題点、課題等に関する自由討議

2. 評価専門調査会（第 3 3 回）議事録について（議題 2）

（配布資料）

- 資料 1 海外研究開発関連機関等の調査報告（1）
－海外の公的支援研究開発の評価システムからの含意－
- 資料 2－1 U.S. R&D Program Evaluation: Methods and Status
- 資料 2－2 Instructions for the Program Assessment Rating
Tool(excerpts)
- 資料 3 Assessing Research and Innovation Policies and
their Impact

- 資料4 これまでに提起された主要課題
資料5 評価専門調査会（第33回）議事録（案）

（机上資料）

- 国の研究開発評価に関する大綱的指針（平成13年11月28日）
- 科学技術基本計画（平成13年3月30日）

議事概要：

【大山会長】

ただいまから第34回評価専門調査会を開催いたします。

本日の議題は、お手元の議事次第に示されておりますように、2つの議題があります。1つ目は、評価における今後の課題と改善方向の検討であります。2つ目は、第33回の議事録（案）の確認であります。

1つ目の議題では、海外研究開発機関からのヒアリングも予定しております。本日は、お2人のゲストをお迎えしております。まず紹介をさせていただきます。デービッド・トリンクル氏とルーク・ジョージュー氏であります。

デービッド・トリンクルさんは、1992年からRAND社において研究開発データ及び政策分析に従事されまして、産学連携モデルの効率や、30余りの研究開発プロジェクトの基礎的な分析を担当されました。2000年からは大統領府行政管理予算局に移られまして、現在までプログラム分析官として、科学及び宇宙プログラムを担当され、2002年からは副部門長をなさっております。担当されている業務は、国家研究開発の調整、マネージメント及び政策立案、科学技術予算策定、管理、プログラムマネージメント、それから、国家研究開発のポートフォリオ分析と多岐にわたっております。本日は、研究開発資金を管理する行政管理予算局の立場から、米国における研究開発プログラム評価の手法及び現状についてご講演をいただきます。

また、ルーク・ジョージューさんは、マンチェスター大学において研究開発評価や科学技術政策の研究に従事されまして、1990年からは同大学の工学、科学及び技術における政策を研究するPRESTの執行役員、そして1992年からは教授に就任され、2003年にはイノベーションリサーチ研究所の代表、また、マンチェスタービジネススクールの統率者の一人として任に当たっております。研究開発評価の分野においては、EU各国を中心に研究開発政策等の評価委員、これを歴任されるなど、報告書、著書も多数出版されております。研究開発評価の手法、歴史等に通曉され、研究開発評価の分野で世界をリードされている方の一人です。本日は、リサーチ及びイノベーション政策及びインパクトの評価について、フィンランド及び英国の事例の紹介を交えて

ご講演をいただきます。

それでは、よろしく願いいたします。

議題 1：評価における今後の課題と改善方向の検討について

前回の評価専門調査会に引き続き、評価における今後の課題と改善方向の検討のため、①海外研究開発関連機関等の調査報告－海外の公的支援研究開発の評価システムからの含意－、②海外研究開発機関等からのヒアリング [デービッド・トリンクル氏（米国行政管理予算局プログラムエグザミナー）、ルーク・ジョージュー氏（英国マンチェスター大学工学科学技術政策研究所所長）]、③評価における問題点、課題等に関する自由討議、が行われた。

【大山会長】

それでは、議題 1 に入ります。

本件については、前回に引き続き、評価における今後の主要な課題の抽出と、それらに対する改善方向の検討のためのヒアリングを行います。本日は、お手元の資料のとおり、海外研究開発関連機関等の調査報告、海外研究開発機関等からのヒアリング、評価における問題点、課題等に関する自由討議を行ってまいります。

まず、海外研究開発関連機関等の調査報告であります。この調査につきましては、財団法人政策科学研究所に委託し、調査を行ってまいりましたので、本日は、その概要の報告の 1 回目といたしまして、プロジェクトの評価と機関評価に関する海外の状況についてご報告をいただきます。

なお、2 回目の報告として、次回の専門調査会で、主に政策を対象とした評価について、海外の状況について報告をお願いすることにいたしております。

それでは、政策科学研究所の大熊主任研究員から説明をお願いいたします。

【（財）政策科学研究所】

政策科学研究所の大熊です。海外の研究開発関連機関等の評価の実態調査を幾つかのアプローチでしている途上です。今日の資料の副題にあるように、海外の公的支援研究開発の評価システムからいかなる含意を引き出すかというのが我々のミッションですが、特にこの専門調査会が大綱的指針の見直しの準備に入っていると伺っておりますので、これに供するようになりたいと思います。そういう点で、なかなかこの対象は広いですし、いろいろな文脈の中で考えなくてはいけない。あるいは変貌中でもあるということではあります。今日と、次回にあわせて、海外事例の中にどういう示唆が考えられるだろうかということについて少し問題提起をさせていただきたいと思っております。

スライド 2 に、評価の全体の枠組みとこれまでの検討範囲を示した図があり

ます。今回、この大綱的指針を見直すうえで、特に評価対象の政策区分、それから評価フェーズの組み合わせにおいて、今までの大綱的指針では何が扱われ、あるいはそこでどういう問題が起こっているか、また、どういうものが扱われていないかということを整理しておくことが必要だと思いました。そのことが、すぐこの後にお話しされるおふた方のお話のポジショニングにもなると思いますので、調査方法とともにその説明をさせていただきます。

まず、私どもの調査方法の一つは、訪欧調査です。スライドの11に今日おいでになるルーク・ジョルジュ先生を含めて訪問先がありますが、3月8日から12日まで、政策研究大学院大学の丹羽教授、今日は大変ご多忙で、私がかわりにその内容を踏まえて話をさせていただきますが、その方と私どものメンバーと、それから総合科学技術会議の事務局メンバーのご同行をえて調査に行っていました。調査項目とともに掲げてあります。

それから、調査の方法の二つ目は、今日、欧米の専門家を迎えて集中的なインタビューをするということで、これを素材といたします。三つ目に、国内で既に各調査機関等が各省などの委託調査等の絡みで海外調査を重ねておりますが、その報告書などから、意味ある情報を整理してみるということでもあります。最近、海外の調査レポート、特にEU関連、あるいはOECD関連でかなり総合的なレポートが出るようになってまいりました。これらも踏まえて状況を整理します。主に次回にお話させていただきますが初めのスライド2に戻っていただきます。

研究開発評価のシステム全体を見るときは、通常、たて軸にありますように、政策体系ということを経層的にとらえるということがされております。そこでは、政策、プログラム／制度、あるいはプロジェクトがあります。このプロジェクトの中には、プログラムのもとに発している従属型プロジェクトと、それから独自に、走る独立型プロジェクトがあります。また、プロジェクトの組織的な集積とでもいうべき機関という対象もあります。それから、よこ軸には評価のフェーズとして事前評価、途上評価、追跡評価という枠があります。

我が国で研究開発評価に絡んでかけられている制度的枠組みとしては、大綱的指針があり、これがⅠ期、Ⅱ期と走り、その後Ⅲ期に入ることです。これに加えて、現在政策評価法と、関連するものとして独立行政法人通則法等が同時にかけているわけです。この大綱的指針では、Ⅰ期、Ⅱ期に課題と機関、それからⅡ期に加わった施策（と研究者の業績）という階層が扱われておりました。これを階層とフェーズの枠組みのスライド2図では、ローマ数字のⅠというのが主としてⅠ期で扱われているもの、ⅡというのがⅡ期で扱われてきたもので示しました。問題は、ここで網がかかっていない未だ取り扱っていない部分で、これが難しいものとして残されながら、現実には重要な評価の対象になってきているというところです。海外の先進国では、特に政策

の質的な改善のために取り組まれ、この間蓄積が始まっているという点で注目されるわけです。それで、主として海外調査を含めて、この空欄の部分について調査を重ねました。それから、同時に、今日は展開しませんが、Ⅰ期、Ⅱ期で扱われた対象に対しても、質の向上を図る上での問題点、あるいは国際的に見ても新しい環境の中で生まれて直面している課題というようなもの、そういったものを取り上げて、先進国の取組みの整理をしておきたいと考えております。繰り返しますが、大綱的指針が扱っていない白い部分が新たな問題で、訪欧調査も重点を置きましたし、この後のルーク・ジョージ先生とトリンクル先生のお2方のお話もかんでくるという位置づけになると思います。

それで、次のスライドの3番目ではありますが、我が国では評価対象を課題と政策として大綱的指針等でも定義されているわけですが、ここでは政策とプログラムとプロジェクトという階層的で、構造的なとらえ方が重要になってまいりました。より上位の政策レベルのものに照らして評価を行いつつ、政策体系の改善を図るというわけです。特にプログラムということを手がかりに、政策をつくり、新しい改善なり見直しの切り口を探るというメカニズムが進んでまいりました。明確な施策目標単位で政策装置を備えた、いわば政策のプログラム化ということも絡んだ話でもあります。スライド3の真ん中にある四角ですが、追求すべき内容をもった個別の政策が、パフォーマンスが検証可能なような、あるいはオペレーショナルな操作的な概念であるプログラムという格好に展開をしています。政策装置、ポリシー・インストルメンツという言い方を（欧州では）して展開し、集積しています。これを四角の右側のサイクルでいいますと、具体的なプロジェクト群としてプログラムは展開をして、評価のサイクルをつくって見直していく。四角の左側の方では、この施策からプログラムの関係を、一方でより包括的な政策との間でループをつくって新しい循環的な見直しをかける。それぞれの政策対象レベルにおいて見直しの内容等については違っているわけですが、こういった構造で実際に政策の評価のダイナミズムをとらえることが、評価の混乱を避ける意味でも非常に重要なおもっております。我が国の場合は、このプログラム化の度合いが非常に小さい。既に実施されている政策評価の中で評価のやりにくさが（政策の立て方や構造から）生まれているということが言えると思います。

少し前置きが長くなりましたが、スライドの4番目にまいります。今のプログラムということが組織的に集積し展開しているものが、機関の行動といえます。機関評価については大綱的指針の第Ⅰ期、第Ⅱ期においても扱われてきています。機関評価の代表的なものとして大学、それから公共的な研究機関の評価があるわけですが、大学の評価で国際比較をしてみますと、非常に特徴的な異なったタイプが見られます。それがこのスライド4と5の各国の大学における研究評価の状況というもので、その1とその2に分けてあります。評価の哲学の違いが象徴的にでています。

その1のグループに、フランス、オランダ、ドイツが挙げてありますが、これは、評価結果を直接予算等への処遇に直結させない、それをマネジメントの改善なり全体のシステムの改善なりに振り向ける、評価を公開しながらもサポータティブに、大学にある意味の自立的な改革を促すというタイプの運用をしているタイプであります。フランスは、評価機関が、中央のCNEという自立的な独立行政組織であり、その結果は公表するが、資金・資源配分面に反映しないという構造であります。オランダは、最近、大学評価の中では非常に注目されているところではありますが、同様に自己評価書に基づいて評価をするわけですが、これも学生の選択等に供しながら大学の自立的な改革に使っていくということでもあります。ドイツは非常に複雑で多様でもありますが、これも一般にドイツの国内であるように、いわばトラスト（信託）というような評価に関する、あるいは政治的な関係においての理念を生かしたようなことが行われていると言えると思います。

一方で、スライドの5ページには、よく知られているように、イギリスではHEFCSという高等教育資金配分カウンスルにおいて、国際的にも注目を浴びたわけですが、UoAというユニット・オブ・アセスメントの組織単位で、3から5年ぐらいのサイクルで評価を行い、それで実際に、評価が下の方では研究資金を切るという対応もしたということでもあります。実際の経過上では、評価の点数の上位校が急増してきて、評価環境に順応したという面もあり、評価のコストや有効性など、いろいろな議論が起きました。最近の情報では、全体としては、この枠組みは変わらないが、これまでの単一の評価点でなく、一種のプロファイル型といいますか、全体の組織の研究の質の分布状況を反映させる格好で運用を続けるということになったようであります。いずれにしても、評価を資源配分に反映させる格好で刺激を与えているわけであります。

また、非常に変わったタイプに公的な研究評価システムのないアメリカがあります。アメリカでは、よく知られているようにアクレディテーション制度により、教育に関する評価が大学間で自律的に行われています。研究では、研究資金の多くが外部からプロジェクト単位で競争的に獲得される。このように競争的資金を大学自らが獲得をするというための努力、競争があるために、結果として既にパフォーマンスについて反映されているということがありまして、全体としては、幾つかの州でのパフォーマンス・ファンディングを除き、研究評価については、実態的なものはやられていないと思います。日本は、それに対しまして始まったばかりではありますが、評価を行って、評価結果を運営費交付金の配分に反映させようという動きが伝えられているところでもあります。

機関評価としては、今は大学のお話でありましたが、公的な研究機関も同じような違いがありまして、スライド6には例示をしておりませんが、アメリカの研究機関がGPR Aのような循環型の評価にさらされているということを除きますと、大体評価については同じような動きがあります。ここには機関評価

における論点を掲げました。まず、評価結果を資源配分に直接的に反映することがいいか悪いか、あるいは、このことの裏返しとしては、評価理念としていけば被評価者に対してエクスクルーシブか、インクルーシブか、対置的か、支援的かというような、いわば評価理念の選択でもあります。あるいは、評価が実績評価を中心とするのか、マネジメント評価の中心とするのかとかということでもあります。この点で、我が国のシステムが組織文化とも関連してどう進めるべきかという議論があります。

もう一方で、資金提供の性格がブロックファンドかプロジェクトベースのファンドか、また、どんな組み合わせか、という論点があります。おのおの安定性とか効率性という傾向を伴っているわけですが、この両立をどう追求するか、この中で研究人材の確保策にかかるフルコスト・プロジェクトのような制度的な可能性があるかどうか。

異なった論点として、また、評価を予算に反映させる場合にも、どのような活動単位で競争をさせるか、という論点もあります。単位が大きすぎればその中のマネジメントが改めて問題になり、多様な活動単位を持ちながら、その競争を促すような仕組みが必要ですが、これは一方で、信頼性とコストのトレードオフの関係になったりしているわけです。この辺が各国の機関評価の拮抗からみて大きな選択であろうと思います。

ところでスライドの7番になりますが、プログラムのもとにあるプロジェクト評価では、プログラム自身のあり方にかかなり左右されているわけですし、このプログラムのあり方が、プロジェクト評価のパフォーマンスの死命をむしろ決めている面があります。そこで、プログラムの設計と運用の仕方について、いろいろ進展があり、やはりこれも欧米においてかなり先行的な動きがあります。ここのスライド7番にありますように、A l v e yプログラムの追跡評価研究をきっかけに、ここにいらっしゃるルーク・ジョージ先生もこれに貢献されたと同っておりませんが、リアルタイムな評価制度の必要性が認識され、また、ここにあるROAME型のシステムが導入されたことがありました。事前にこういった評価サイクルを十分に折り込んだ格好でプログラムを立てている。また、検証可能な格好で、プログラムの目的等も明確に立てていくというような構成をとったわけですし、これがヨーロッパを中心に影響を与えました。

そこで、実際にそういったプログラムのもとで走っているプロジェクトでは、その目的に照らして個々の課題の評価を行うということになります。この課題評価の体制においても、幾つかの重要な違いが出てまいります。スライドの8番ではアメリカにおけるファンディング機関のパターンだけを示してありますが、プログラムマネージャーの位置や権限について、研究者等の入ったパネルとの関係でタイプを分けてあります。プログラムマネージャー自身は、いわば庶務的な事項の責任者というような言い方もありますし、制度設計の事実上の責任者であります。プロジェクトディレクターというような人たちと一緒にパ

ネルリーダーやパネリストたちを選び、パネルの運営での共同関係も行うケースがありますが、強力な権限をもつものが多いのが米国です。こういったプログラムマネージャーの役割や、どういった人たちから選抜をされ、どういう機能を発揮するかということに関わって、評価の目的や性格、評価専門人材の所在に関わるパターンがあります。今日は時間がありませんので、改めてこの機関なりの性格とあわせた格好で整理しようと思っておりますが、特徴的なバリエーションを持っています。

それから、スライド9番ですが、アメリカ以外の各国のプログラムの運営責任体制をまとめたものであります。ここでは多くの場合、プログラムマネージャーが専ら調整役であり、行政内部の職員によって担われているということが特徴であります。ドイツとか、一部違ったタイプのものもあります。これもやはりプログラムの全体の特徴とか、特に研究開発者コミュニティとの関係ということの絡みで選択されるわけで、この辺の論点も整理しておきたいと思っております。

スライド10番にとくに課題評価の中心的な議論の一つであるレビュー体制の質的改善の問題を掲げました。我が国であいまいなピアレビューとエキスパートレビューの峻別をしながら、実際に全体のレビューシステムの設計と運用をしなくては行けないという点であります。ここで1番から6番までの囲みに考えなくては行けない論点を提起をしております。海外の、特にEUであるとか、あるいはNIH、NSFというようところでは、非常に詳細なガイドラインが生まれ、それがしかも不断に見直されています。この辺の見直しの過程等も、いわば一種の制度合理性というような知見の集積でありますので、次回を含めて、報告書でまとめたいと思っております。

とりあえず与えられた時間が来ましたので、簡単ですが、一応海外の公的支援研究開発の評価システムからの含意を整理する枠と一部、機関評価、従属型プロジェクト評価の論点をご紹介します。この枠を用いてあとの両先生のお話しされるところがどういうところであるかということをお伝えしたところです。よろしくお願ひします。

【大山会長】

ありがとうございました。

ただいまの説明に対しまして、何かご質問がありましたらお願いいたします。如何でしょうか。よろしいでしょうか。

先ほどご案内しましたとおり、海外研究開発関連機関等の調査報告につきましては、次回もまたご案内いただくことになっていきますので、その中でまたいろいろご議論いただきたいと思います。本日はありがとうございました。

それでは、続きまして、海外研究開発機関からのヒアリングに移ります。

まず、デービッド・トリンクルさんから、米国の状況についてご説明をいた

だきます。30分ほどでご説明をお願いいたします。

【トリンクル氏】 David Trinkle:

I am grateful for this opportunity to speak to you about research evaluation efforts in the United States. It's in tune with the evaluation efforts I will be discussing that are still being held, so I am interested in brief discussions and would like to repeat that. I think we all benefit from sharing our experiences in S & T policy. Just a little background to start off with. I am with the Office of Management and Budget, OMB, which is a large office within the executive office of the President. OMB has multiple roles in the government, including developing the president's budget request for all agencies each year, and also overseeing the management of the government both operationally from day to day, and also through own return management implementation and coordination roles. All of these roles extend to research and development. As you have heard, I have a cue role with them, the Office of Management and Budget, both through the over-sighted National Science Foundation, coordinating research initiatives across the government, and hoping to write the budget chapter each year on research involvement. In these roles our office works closely with the Office of Science and Technology policy, another White House office, and we interact frequently with various members of the academic community. On the second slide, I mentioned that just before President Bush took office, he described the hollowed fellows of results and accountability throughout the government. I can say from my personal experience as staff within room B. I was there during the Clinton administration as well. Both were very focused on management. I have to say this administration is that much more focused on formalizing items, and having insistent processing across the government. On the third slide, I just wanted to mention that when this administration came in and the policy officials that I work for wanted to start making decisions they were somewhat disappointed that our budgeting appeared to focus mostly on marginal increases in the previous year, rather than on how our programs did, whatever it was that they were meant to do. And an intended focus more on effectiveness and efficiency really only is that much greater as the US economy moved from surplus to deficit. On the first slide, the challenge then is how to incorporate more program results into

funding management decisions. As you have heard, government performance and results after GPRA has required, in law that agencies report on the performance. While the intent was a good one, our officials felt that the GPRA process fell short, however yielding useful information in a way that would be helpful for making decisions, and on a timing that would allow us to use that information each year, in order to make those decisions. So, on page five the administration put into effect, defined and published the president's management agenda, which includes a number of initiatives for improving government management. I am not going to go into most of them, but it's within this framework that the PART and other things I will be talking about came about. On page 6 is one of those, this actually predates the PART itself, and pretty much defined the philosophy that we used for the R&D components of the PART. And this was an initiative to develop better criteria by which to judge research and development in general, in order to better understand what are investments are intended to yield, and have yielded, to link information about potential benefits, improving performance and to decisions we make, and also to, because our office was intended to be effective by laying all this down, we essentially were able to articulate the government's expectations to the agencies and research community. By 2002, on page 7, we refer to a pilot attempt at the US Department of Energy, which have brought in, the research investment criteria to apply to all US agencies. In doing this, we were faced with a number of challenges. Research and development has many different characters, many different flavors, and how to make a set of guidelines that can apply to basic research and also long mission driven research, can apply to operational research facilities, but also is founding to the universities. And one thing that we wanted to make sure we did is not set such high expectations for reporting in performance of basic research that we will drive basic research to take fewer risks to essentially be more applied in its focus. So with that, page 8 is, ah, OK, in your package it's now on page 7 instead of page 8. The outcome was with three fundamental criteria, these are based in part on existing factors, but also on recommendations of the National Academy of Sciences in the United States, we chose to focus on three fundamental principles: the relevance of the research, the quality of the research, and the performance of the research. And, as you see,

this has both a prospective component, components in advance, of the research and a retrospective component. I am not going to go into great detail on these, but somewhere in your package the criteria are listed in full. But briefly, by relevance in combination of well-defined research, and I think on page 8 now in your packet, and research that can demonstrate how it's relevant to the intended customer, whether that's fields of science such as chemistry, other R&D programs, or some development attempt. Quality, by that we mean essentially a research process that itself promote quality research. Competition, as you have heard, is a large component of this, a process based on merit, and also a process that has independent or external review by experts in the field. And then finally, there is performance. And this is, of course, in a sense we'd like to always be able to show how well we are doing. But essentially relevance and quality are to help us show whether we are on track, or we can't necessarily show performance on an end basis. From page 11, I just want to add that we have a few more specific criteria for industry-related programs, where they would, in order to justify government support, would have to justify the value of the effort at all, the need for the government funding, and the need for a program design in order to design. These grew out specifically of concerns for research funded through the Department of Energy, but we have a problem this same of the rest of the government. To do this, some must be asked at the government. Now as you know, I was asked here to discuss the PART, this is all by way of background, but the philosophy is how looking at the PART of the research in general. So I thought that was important.

The PART, as the result of another of the initiatives of the president's management agenda, It was the intent was to be able to use performance information in making, funding or rather decisions. And this is meant to apply not only to research, but to all government programs. And research was only one of those type programs. In order to do this, we found that we needed a new tool to uniformly and consistently report on performance. And my outline of these tools, now on page 13. PART is an acronym for Performance Assessment Rating Tool, and it's essentially a large Excel spreadsheet. It's a tool that OMB and the agency can work on together to characterize each program in government. It has four sections, which is somewhat important because it characterizes the four components of a program

which was to look out. First of all is, the way of program's focus is defined, and the program is designed. Second is the way the program does its strategic planning. Third is a more day-to-day program management aspect. And fourth are results and accountability. Now I will discuss each of those actually I discuss each of these on the slide. Performance and design, these are types of characteristics that, essentially it's how the program was defined to begin with. Many of our programs are defined in law, while others are defined throughout the standard processes within the government agencies at their discretion. But the first section of the PART gets out, even what the program is supposed to do, how well the program is designed to do what it's supposed to do, and there are, now there are how many questions are each section, but fails to take questions in that section. Strategic planning is the second section. It sets up the program result section, the last one. And it is the prospective, the in advance characteristics of the program. For example, what are the program goals? What are the measures that are defined to help track the program to performance towards those goals? And what are the targets that the program has to try to hit its protocols. Finally, the third section is program management, which as I said has to do with the more mundane operations program, can include financial management and also how well the program promotes and maintain efficient operations. And then finally, program results. And that's where we assess how well programs do, they are work, in case not their research to address the goals that we have set up in this strategic planning section. On page 14, essentially we, resulting from this process we have a number of results for this PART. In each section we actually score them numerically, that's the focus of a lot of attention, but in many ways that to me is the least important outcome of all this. The numeric scored role come up with a single rating, just broad terms, whether the program is effective, or moderately effective, or ineffective. But I think the important thing is that, because of the four sections of the PART tool, we have a greater understanding if the program isn't perfectly effective, what are its relative strength and weaknesses, and what types of recommendations come out of this whole process that inform what we do next. I think, early in the design of the PART we had said, you know, the PART would be a tool we could use to reward effective programs, and perhaps cut ineffective

programs, and we quickly realized that's actually not what we normally do, and while we may want to be able to do that, the tool gives us a lot more information to make whatever the right decision is, whether it's an ineffective program, for example, might benefit from more funding to help fix some of the problems that, you know, are its weaknesses. Or conversely, with an effective program it's possible that we may have achieved its goal, and we can now move on. This whole system allows any number of decisions to be made. And through the PART assessment process, we come up with recommendations, and then the following year we will go back to that same program and understand what the program has done, whether it's program management, or re-defining some aspect of the program to address those recommendations and address any weaknesses the program might have. Just started our effort, on slide 15, we have assessed about 400 programs to date, about a trillion dollars of government funding. This is the second of five years, that by the fifth year we have to have assessed the entire government. I do provide a link to any of the information on the PART, it's all on the Webpage, that includes our guidance, but I think it's in the packed actually also. But it also includes every assessment we have done to date, and you can actually look at the details of any of these programs. Specific to R&D, on page 16, I just want to note that most of what I just said applies to all programs, research and development is one of seven specific types we defined, and is in many ways the most difficult to assess. And that's why, I think we're all here to discuss how to do these things. As I said before, we did rely on the principles of the R&D investment criteria, where for example, with basic research results uncertain, the research has a long-term character where the payoffs won't happen for many years, and there are a number of risks involved. And so we wanted to make sure that in detailed assessments we took those factors into account. So far, of the 400 programs we have assessed, 58 of those have been research and development programs. On page 17, just for those research program, nearly half of them, assessed to be effective, now they say that that itself doesn't necessarily mean anything, or it mean as much as I should say as for the programs is not perfectly. We now have some recommendations for how to address those. On slide 18, I do input some specific examples of the assessments we have made, for example some of the NASA or National Science Foundation

programs that have been assessed to be effective, all the way down to ineffective programs. There are a couple of fossil field programs of the Department of Energy, for example, that were found ineffective due to a combination of measures that aren't well defined, and not well used, and performance that just doesn't seem to exist in a way that justifies these programs. To address those, we are focusing first of all more on how to get a measure of the performance, and track it, and how to focus the program more on activities where they have a better chance of having better performance. Next steps, essentially this is work in progress, but we continue to assess more programs for the 2006 budget. Using the PART we do have more work to do, I think, in clarifying our expectations for what a good performance measure is, for what the expectations are of research programs. And I think, we do have more work to do in communicating with our legislative branch, the Congress, in what it is that we are trying to do and what our assessments show, and in getting their acceptance of the processes we now use. That ends my prepared talks, so if there are any questions on any of these things, I'll be taking them. Thank you.

本日は、皆様に、研究評価がどのように行われるかということについてお話しする機会を与えていただきまして光栄に思います。本日お話しする評価方法というのは、まだ進化途上の状況でありますので、ぜひ皆様からのフィードバックもいただきたいと思っております。科学技術政策においてお互いに経験を共有することは有用だと思っております。

背景説明をまずしたいと思っておりますが、私、行政管理予算局（OMB）に所属しております。これは大統領府の大きな局でありまして、複数の役割を果たしております。その一つとしましては、大統領によるあらゆる省庁の予算要請・要求について全体像をつくるということと、また、日常のオペレーションについて政府全体を見ていくということと、より長期的な行政管理の計画立案、そしてチェックをしていきます。こういった役割は、すべて研究開発と関連しております。先ほど紹介がありましたとおり、行政管理予算局におきましては、NSFのオーバーサイトを行っております。研究イニシアティブを政府間におきまして調整する役割を担っております。また、毎年研究開発費の予算策定に貢献しております。ホワイトハウスの科学技術政策局と緊密に協力しております。また、学会のいろいろな方たちと接触しているところであります。

2枚目のスライドですが、ブッシュ大統領が大統領に着任する前に、政府におきまして成果主義にするということを言っております。OMBのスタッフとしまして、クリントン政権のときもOMBにいました。両大統領ともマネージメントにかなりフォーカスしておりますが、今回のブッシュ政権は、一貫し

たプロセスを一貫したガイドラインにおいて必ず踏襲して確保するというところに力を入れております。

3番目のスライドですが、ブッシュ政権が樹立されたときに、私が勤めているところの政策決定者は落胆しました。というのは、予算編成におきまして、以前、前年比で若干のプラスにすればいいというような予算編成だったのです。プログラムの内容について余り精査しないで、前年比で若干増加すればいいという形でありました。そして、効果と効率を重要視するということは、アメリカの経済が財政黒字から赤字化になって、ますます重要になってきました。

4つ目のスライドですが、それでは、プログラム成果をどのように予算に反映させるか、管理に反映させるかということですが、G P R Aと呼ばれている政府成果法の法律の中で要件化されております。各省庁が成果について報告するということが義務化されております。よい意図ではあったのですが、政府関係者は、この法律のプロセスではまだ不十分であって、十分に予算で使えるような有用な情報を確保できない。そして、その情報を毎年予算編成に反映させるようなタイミングで出してもらうことができていないということになったわけです。

5ページに書かれているとおり、そこで政権としましては、プレジデント・マネージメント・アジェンダという大統領の行政管理アジェンダを決定いたしました。政府管理のやらなくてはならないことを決めました。詳細は割愛いたしますが、この枠組みのもとで、後ほどお話しいたしますようなPARTなどが出てきたのです。

6ページが、その一環の一つの策でありまして、これはその前からあったのですが、R & D分野における哲学を反映しております。このイニシアティブは、よりよい投資基準を研究開発の判断基準として使おうということが意図でありまして、投資の目的をよりよく理解する、実際にその成果をよく評価する。そして、改善の潜在性について定義して、それを決定に反映させるという試みでした。これは偶然かもしれませんが、こういったことを政策として文書化することによりまして、政府の期待を各研究機関や政府関係当局に伝えることができたわけです。

7ページ、2002年におきましてOMBは試験的な試みをエネルギー省で始めまして、投資基準を導入して、それを拡大して、あらゆる官庁でR & D投資基準を導入してみました。そのときにいろいろな課題がありまして、研究開発にはいろいろな側面、いろいろな要素があって、それぞれの状況によって違います。よって、基礎研究にも、また、より目的型の応用研究にも適用できるようなガイドラインというのはあるのかなのか、また、大学の予算編成に直接反映できるような評価基準というのはあるのかなのかということを検討しなくてはなりません。我々としては、基礎研究に関しては成果報告について余り高い期待をかけない方がいいと。そういうことをやってしまうと、基

礎研究におきましてリスクがとられなくなってしまうからということ念頭に
入れなくてはならなかったところです。

8 ページ、皆様の配付資料では7 ページのようですが、失礼いたしました。
3 つの基本的な基準ができました。部分的には既存の慣行からできた基準なの
ですが、全米科学財団からの提案で決めた基準でもあります。3 つの基本的な
原則、研究の重要性、研究の質、研究の成果であります。これは将来の先を見
通す評価、そして過去を振り返る評価という事前・事後の評価を行います。詳
細は割愛いたしますが、恐らく配付資料の中にフルの基準の資料があると思
います。

8 ページに移りましたが、リレバンスというのは、果たして研究の目的がき
ちんと定義されているかということと、また、研究の中でもいかにその研究が
関連性が高いか、その対象としている最終使用者に対して、例えば化学のよう
な、そのほかの R & D プログラムでもいいのですが、あるいは開発努力におい
て関連性が高いかということの評価しなくてはなりません。

9 ページですが、クオリティーというのは研究プロセスでありまして、品質
の高いリサーチを奨励するのにどうしたらいいか。1 つ重要な鍵は競争であり
ます。また業績ベースの評価方法、また、独立性のある第三者による客観的な
レビューが関連分野の専門家によって必要です。

パフォーマンスですが、ある意味では常にどのように自分たちがうまくやっ
ているかを示したいわけですが、クオリティー、そしてその関連性、意義を示
すことによって、必ずしも実績が毎年出てこない場合に、それでも正しい方向
に向かっているということを示すためには、その意義、関連性を示すことがで
きます。

11 ページですが、もう少し具体的な基準が産業関連の分野について述べて
あります。政府の支援を正当化するために、これだけの努力をするという価値
があるということ正当化するための基準です。なぜ政府の支援が必要なのか、
また、なぜこのようなプログラム設計になっているのかを正当化する理由があ
ります。こういう基準が設けられたのは、DOE（エネルギー省）経由の資金
配分で問題があったので、そういう背景があり、こういう基準がその後設けら
れました。

PART の話をするように依頼がありました。PART というのはプログラム
評価の採点のツールですが、ここまでの背景的になっておりまして、ざっと
背景をご紹介いたしました。PART、これは同じく先ほどの大統領のマネー
ジメント・アジェンダの一部として設けられました。ねらいは、いわばパフォー
マンスに関する情報を活用して資金配分、あるいはそのほかの決定ができる
ように資するということです。これは研究開発のみならず、ありとあらゆる政
府プログラムに適用がされることとなります。研究開発というのはその中の一
つでしかありません。これを実施していくために、新しいツール、何か仕組み

が必要だということになりました。一貫性のある形で整合性のある形でパフォーマンスに関する報告を確認したいと思いました。

そこで、13ページにその概要が書いてあります。略してPARTと言っておりますが、これはプログラム、もしくはパフォーマンス評価採点ツールとなります。基本的に言えば、エクセルの大きなスプレッドシートになります。こういうツールを使うことによりまして、OMB、そして各省庁が協力をして、それぞれのプログラムの特徴づけを見ます。4つのセクションがあります。これは重要だと思います。なぜかといいますと、このセクションを見ることによりまして、どのコンポーネントを重要視しているかわかっていただけだと思います。

第1ですが、プログラムの目的がどう定義されて、どう設計されているのか。第2、そのプログラムの戦略的な計画立案、第3、これが日々のプログラムの管理、第4のコンポーネントですが、実績、そして説明責任です。それぞれの項目に関しまして、このスライドを使いながら紹介していきます。

パフォーマンス、それからその目的と設計です。これはそのプログラムの特徴です。基本的に言えば、そもそも、そのプログラムをどう定義づけているかということなのです。多くのプログラムは法律で定義づけられているものもありますし、あるいは、標準的な手順が各省庁でそれぞれの裁量で設けられている場合もあります。ただ、第1のセクションで何をプログラムは行うべきであるのか、どのようにそのための設計がきちんとできているか、いないか。大体5から10ほどの質問が入っています。

戦略立案、これが第2のセクションになっています。プログラムの実績は第4項目で確認していくわけですが、これは事前の特徴づけということになります。実施する前の確認ですが、プログラムのゴールは何なのか、どういう指標を決定して、そのパフォーマンスのトラッキングをしていくのか、何が目標とねらいであり、プログラムとして何を到達させようとしているのかということなのです。

第3のセクションがプログラム管理です。このマネージメントというのは日々の管理になりますし、財務の側面も入りますし、どのようにそのプログラムがうまく効率的なシステムを維持し、推進しているかも含まれております。

第4がプログラムの実績です。このプログラムがどれだけうまく進行しているか評価をします。例えば研究開発であれば、当初戦略企画のセクションで設けられたゴールに向けてどのように進んでいるかです。

14ページ、このようなプロセスを経ていきますと、幾つかのものが出てきます。このPARTの手順に従って幾つかわかることがあります。それぞれ点数を数的に勘定します。これは数字に注目しがちではあるかと思うのですが、本当は一番重要性が低いかと思います。ともかくスコアをつけていきますと、大ざっぱにいいまして、そのプログラムが効果的なのか、その効果は中等度な

のか、効果が出ていないのか、それはおおよそわかります。しかし重要なのは、先ほど4つのセクションに分かれていると言いました。ですから、そのセクションごとに見ることによって、もしその効果が高いレベルではないときには、どこが弱いのか、強いのかかわかることになります。どういうタイプの提案を出すことができ、それに基づいて次に何をどうするべきかが見えてくるということが重要です。そのPART、この中身を考えていく段階でどういう考え方をしたかです。これをツールとして活用していきまして、もし効果的なプログラムがあれば、見返りをきちんと提供するべきですし、効果がないということであれば中断が必要かもしれません。

ただ、それは通常、必ずしもそうしているわけではありません。本来であれば、効果がないということであれば打ち切りはしたいわけですが、そう単純にするのではなく、正しい決定がどういうものであろうと、もし効果が低いのであれば、場合によってはもっとお金を追加したら効果が高まるという可能性もあります。ですから、その弱点を改善できるかもしれません。あるいは効果が高いプログラムの場合には、もうゴールは達成ができたということであれば、次のということで視点を変える必要もあるかもしれません。ですから、スコアが低い、だから中止とは限りません。その結果、いろいろな顛末があり得ます。このPARTを使ったアセスメントを行っていく中で提言を出すこともあり、場合によっては次に同じプログラムを再び検討して、プログラムがどうなっているか、管理に問題があるのか、一部再定義をした方がよいのか、その提言に従って何か直すべきなのか、弱点についてどう対応するべきか振り返ります。

15ページ目、現状です。これまでのところ、大体400ほどのプログラムをここまで評価しています。金額的には1兆ドルほどの公的資金が投入されたプログラムに相当します。5年間のうちの2年目に入ります。5年間の5年目に達したときには、すべてのプログラムを評価が終わっている予定です。下のところにアドレスが書いてありまして、ここにガイダンス、実は皆さんの資料にも入っていると思いますが、これまで行ったアセスメントの結果も入っていますので、よろしければ詳細をご覧くださいことができます。

16ページ、研究開発に関連したところですが、今申し上げた大半の話は、全部のプログラムに適用される話です。その中の研究開発というのは、我々が特定している7つのバージョンのうちの1本です。といっても、あらゆる意味で評価が一番難しいものが研究開発だと思います。難しいからこそ、皆さん、このように集まって協議をしているのだと考えています。申し上げたように、研究開発投資基準の原則を使っています。例えば基礎研究であれば、結果というのははっきりしないものですし、期間は長目になる傾向がありますし、投資の回収といっても遠い将来かもしれませんし、リスクもたくさんあるものです。そこで、詳細な評価をするときにも、こういう基礎研究の性格は反映しております。これまで400のプログラムを評価して、うち58が研究開発絡みです。

17ページ、その研究開発のプログラムに関しまして、およそ半分のものが効果があるという結論が出ています。だからといって何だというわけでもないのですが、効果的だからといって、これがいいと言い切るわけではなく、一方で効果のレベルが低いということであれば、提言を出して試みるということで、それで何が決まるというわけではありません。

18ページですが、これまでのアセスメントの具体的な結果の例が紹介してあります。例えばNASA、全米科学財団、一部のプログラムは見たところ効果があるという評価になっていますし、一方で下の方にいきますと、効果がないというものもあります。例えば化石燃料に関するもの、これはエネルギー省関連のものが一例です。効果が十分ではないという評価になっています。理由は幾つかありまして、例えば測定の方法が明確ではない、あるいは、その手法を評価したところ、このプログラムが十分に正当化されないものがあるということで、どのようにすればパフォーマンスをよりよく評価し、トラッキングができるか再検討しております。また、プログラムのフォーカスをどのように絞り込めば活動内容としても成功する、あるいはパフォーマンスを上げる確立が高くなるのか、そういう見方をしています。

次のステップ、これは現在進行中の段階です。引き続きより多くのプログラムの評価をしていきます。2006年の予算に絡めて評価を続けていきます。このときにも、このPARTのツールを使っていきます。ただ、もっと努力をいたしまして、よいパフォーマンスというのはどういうものか、我々の期待をより明白にする必要があると思います。また、リサーチプログラムの効果という意味でも、何が期待されているのか、より明白にしなければなりません。さらにこの努力を続けまして、一方で立法府、すなわち議会とも協議をしまして、そもそも何をしようとしているのか、アセスメントすると何がわかるのかクリアにしたいと思います。それによって議会側、立法府側でも、我々が行っていることを受け入れてもらう必要があると考えております。

用意した資料は以上ですので、何か質問がありましたら喜んでお答えしたいと思います。ご清聴ありがとうございました。

【大山会長】

トリンクルさん、ありがとうございました。

それでは、ただいまから30分ほどディスカッションの時間をとりたいと思いますので、よろしく願いいたします。ただいまの発表に対しましてご質問、あるいはご意見等、何でも結構です。よろしく願いいたします。

国武委員、どうぞ。

【国武委員】

I have 2 questions. One is, on page 16, I'm working at university, so

I'm concerned about your operations for the basic research. And you mentioned that you debate on the criteria for basic research which requires long-term capital, and has to take into account long-term character and a number of risks. More than other types of operations, the actual item of judgment for basic research is rather abstract, long-term character is not really well characterized in such a way suitable for judgment. How did you find the item for judgment for basic research?

2つ伺いたい点があります。1つは16ページでありまして、私、大学にいますが、基礎研究におけるオペレーションについて関心を持っております。先ほどのプレゼンテーションでは、基礎研究の基準に関しては長期的な要素を念頭に入れなくてはならないということ指摘されました。また、リスクが多いということも念頭に入れなくてはならないとおっしゃいましたが、実際に基礎研究について判断を下すときに、具体的にどういった要素を念頭に入れるのですか。長期的な特徴、長期的な要素というのは非常にあいまいで、具体的な判断基準にはならないと思うのですが、具体的な定義、あるいは判断基準について、基礎研究についてどのように定義されているのですか。

【トリンクル氏】

I think, we did point out here, the purpose what we don't want to do, what we can't seem to do today, because with the National Science Foundation, for example, we, they do the research sometimes in curiosity, it's, and we don't know what will come of it, we don't know when something will come of it, and we may not even recognize that what comes of it is run through the program. So we can't say each year, have you done what you are supposed to do based on performance, relative to expectations. What we can do is, say first of all, does the program have a process, or maintaining the quality of the research, was it funded based on merit, does the proven process over years that the National Science Foundation is that also going on in this program. So that's the quality component of the R&D criteria. There is also, ultimately we rely heavily on basic research, where we might not find this acceptable for other types of programs. We rely heavily on independent review, where either through the National Academy of Sciences, or through committees of visitors, or other advisory bodies for the National Science Foundation, to basically say, every year or three, depending on which kind of program this is, whether the program is focused in meaningful areas, whether it's going in a good

direction. And these things are somewhat qualitative. It's hard to say, yes or no, or, this gets a 10, or anything like that. But for basic research at least, we have to have some way of assessment that doesn't rely on, you know, did you do what you said you would this year, because it just won't work for basic research. Basic research, I have to add, is really the one area where the administration is saying, you know, we don't quite have the same expectations for performance measures and for annual results.

最も簡単な方法で説明しますと、まず我々としてやりたくないこと、つまり、我々としてこうやっていると見られてはならない方法というのは、全米科学財団の場合、例えばNSFが好奇心からだけで研究をやってしまうことがあるのです。そこから何が成果として出てくるかわからないし、いつ成果物が生まれるかもわからないし、また、成果として出てきたとしても、そもそもその発端となったプログラムと関連性があるのかもわからない。ですから、やるべきことをやったのか、パフォーマンスベースでやったのか、期待したことをやったのかということは毎年はかることはできないのです。でも、まず言えることは、果たしてプログラムの中にプロセスが定義されていて、研究の質を維持するためのプロセスが確保されているか、能力ベース、業績ベースで資金が配分されたか。NSFの歴史的な成果についても我々のプログラムにおいて評価しているのですが、それがR&Dクライテリアのクオリティーの部分の評価であります。

また、加えまして、基礎研究にある分野では非常に依存するが、ある分野では依存しないということがあると思います。独立レビューを、例えばナショナル・アカデミー・オブ・サイエンスを通して行う、あるいは委員会を設置する、あるいは訪問評価員、あるいはNSFの諮問委員会を設立して、毎年、あるいは3年に1遍、そのプログラムの種類によってサイクルは違うのですが、果たしてそのプログラムが十分に意味ある分野にフォーカスを絞っているか、よい方向に進んでいるかということの評価してもらいます。これはかなり定性的でありまして、イエス、ノーで白黒はっきりつける、あるいは10点法で10点つけるというのは非常に難しいところあります。ただ、少なくとも基礎研究におきましては、何らかの評価方法で、今年やると言ったことをきちんとやったかということではないような評価方法が必要です。つまり、つけ加えますと、基礎研究という分野は、政権としましては業績測定評価に関しては、毎年の業績としてはそんなに期待が高くないと言っている分野です。

【国武委員】

You mentioned quality criteria. Are they based on the expert opinions? What are they?

定性的な基準とおっしゃいました。これは専門家の意見に基づくものでしょうか。どういう根拠があるのでしょうか。

【トリンクル氏】

Yes, this is for that. Based on expert assessment that the program is on track, and is focused in productive areas. Now I should add that the, you may be familiar with these types of arguments, but the system of independent experts reviewing things, and independent merit review, those are good to find, those are good ways to assess research in conventional fields, in well-established fields. Innovative research can be that much harder to do, because the experts are not as well defined, the math oftentimes this research areas fall into non-conventional disciplines, or interdisciplinary research. So this, we don't have all the answers, but we are certainly trying to factor these things in, and I think, speaking, you know, for my familiarity at the National Science Foundation, we rely heavily on their processing, they have been struggling with these sorts of questions for years. And we try to base what we are doing on existing practices at agencies including Department of Energy and National Science Foundation and NASA that have proven results in these difficult to measure areas.

そうです。専門家の評価に基づいてプログラムが予定どおりいっているのか、生産性が上がる分野にフォーカスが絞られているのかどうかです。ただ、ご承知かもしれませんが。こういう議論をお聞きになったことがあるかもしれませんが、独立専門家による評価、あるいは独立機関による中身の質の評価、こういうことを実施すると、確かに従来分野、つまり伝統的な分野での評価はできるかもしれないのですが、新しい分野の評価は難しいかもしれません。専門家といっても、その定義、あるいは領域がはっきりしていません。しばしば新しい研究分野というのは、非従来型の分野にまたがっているといいますか、一つ一つの分野に分けられないのです。答えはすべて整っているわけではないのですが、ありとあらゆるこういう方面も配慮しようとは努力しています。

私の場合は、NSFを知っているということがあるので、こういう疑問については何年間も悩み続けているのが現状で、我々がしていることは、既存の省庁活動、例えばエネルギー省であるとかNSFとかNASAがしていることも配慮しています。こちらでは実証された測定がしにくい分野での業績もあるので、それは参考にしようとは努力しているところです。

【江崎委員】

Thank you very much for interesting presentation. My question is

how much cost of this kind of evaluation and I mean that you are about 400 programs one trillion every year you examine something like 20 percent, I understand this one. So what, you know, suppose you invest one trillion, and you use all this your part of system. Supposed you do everything, how much cost of those kind of. Very difficult to evaluate those things, but I just know approximate.

興味深いプレゼンテーション、ありがとうございました。コストについての質問です。どのくらい評価にはコストがかかるのですか。400ぐらいのプログラムで1兆ドルというようにおっしゃいました。毎年、その約2割を評価すると理解しているのですが。例えば1兆ドルとしてこういったシステムを活用する。それで全部やるとどのぐらいのコストがかかるのですか。大ざっぱな数字で構いませんが。

【トリンクル氏】

That's an excellent question. We don't, and I think it's a valid concern, that with more oversight, and the more reporting requirements, that there are the fewer resources, and the less the program managers time there is to focus on what he or she supposed to be doing in managing the program. The first answer to your question is, that the PART Tool formalizes some assessment efforts that OMB does anyway. Most of the questions that are on there are questions that many of us ask of our programs and have four years. What is new is that the PART basically makes each examiner ask each program each of these questions, and that is more effort than is traditional. Also, while the tool gives a good, is a good tool for dialogue with the agency of how well, how did consider this question or these sort of performance that the agency is trying to justify. What in order to, since this is somewhat of a negotiation, these are not assessments that OMB does in isolation. We work with the agencies and try to understand their point of view, and try to understand our points of view. It takes a lot of time. And we are just now, for example, starting the process for this year. The guidance went out this last week. And we will be done, mostly done over the next, I actually don't know what the due date is, but over the next two months, but we won't be complete, truly complete until November. And so, it takes a lot of time. As to what the real costs are, this is something that agencies and other groups have asked us. We don't know, and we have not done the cost and benefit analysis of our own efforts. And it's a valid

question. It's a lot of my time, it's a lot of agency time. And I think, this is part of we work in progress, and we ourselves are learning how much of this is worthwhile. And it's hard to say what the costs are, because people still need to do all the work they already have to, so some of the costs are hidden just in later hours, I think.

すばらしい質問です。当然、より監視する、監督する、そしてより多くの報告要件を課すとすると、プログラスマネージャーの時間もそれだけかかってしまい、コストもかかるわけですので、端的に答えますと、PARTにおきましてはツールとして、どっちにしてもOMBがやっているような評価を公式にPARTの中に組み込んでおります。つまり、質問のうちのかなりの部分は、我々がもう以前からプログラムに関して自問自答しているような問題です。ただ、それぞれの評価員がそれぞれの質問に関して具体的に口に出して質問をするというように形式化しているというところでありまして、これは従来のやり方よりも形式化しているというところでもあります。より努力が必要になってくるというところでもあります。

PARTは当局との対話を促進するということで、非常に有用なツールでありまして、これらの提起された問題についてどのように考えるか、我々が正当化しようとしているような評価についてどう考えるかということの対話を促進するよいツールなのですが、いわば交渉なのです。ですから、OMBが単独で評価をやっているということではなく、我々も先方の見解を理解しようとする、先方も我々の見解を理解しようとする時間がかかるのです。例えば今期のプロセスを始めたばかりで、1週間前にガイダンスが発表されたところでもあります。そして、この作業が終了するのが、期日は忘れましたが、大体2カ月間が設定されております。でも、本当にすべて終了するのは11月までかかりますので、かなりの時間がかかる場所でもあります。

実質コストがどのぐらいかということですが、これもやはりいろいろな省庁に聞かれておりますがわかりません。我々、コスト効果分析を自らの努力に関しては実は行っておりません。ですから、非常によい質問ではあります。私の時間もかなり費やしている、OMBの時間もかなり費やしている。ですから、これから我々が分析しなくてはならないことの一つではないでしょうか。我々自身がどれだけ価値のあるかということ自ら評価しようとしているところでもあります。ですから、コストは幾らかと聞かれても、具体的なお答えをするのは難しいです。と申しますのは、PARTのかなりの作業というのは、いずれにしてもやっていた作業ですので。

【江崎委員】

When you started the PART program?

PARTはいつから始めていらっしゃるのですか。

【トリンクル氏】

Last year was our second year, so we do it now been for 2 whole years, as a PART.

昨年が2年目だったので、丸2年経ったことになります。

【江崎委員】

Something like 40% of this program?

大体40%ができた。

【トリンクル氏】

That's correct.

そうです。

【江崎委員】

Thank you very much.

ありがとうございました。

【大山会長】

他の委員の方、如何でしょうか。

末松委員、どうぞ。

【末松委員】

I am also belonging to the university. I am also interested in basic research, and also in oriented research. Do you have any idea, how do you decide what part of the money is going to the basic research and what part is the oriented research?

私も基礎研究に興味があります。どのようにして資金のどの部分を基礎に振り向けて、どのぐらいをオリエンテッド・リサーチに振り向けるべきか決めていらっしゃるのでしょうか。

【トリンクル氏】

That's an interesting question. That gets to how do you define the research portfolio that is within an agency or across the government. The program assessment rating tool is very much focused at the program level, and so if the unit of analysis is essentially already defined there. In a way the investment, or the R&D investment criteria are little more flexible in that you could find them up, you can

use them at the agency level of the program, or at the project level. And to understand whether the whole portfolio is helpful, healthy, and useful. And that said, that not typically how we make those decisions. And I have to say, in my familiarity of the budget process across the research agencies, often a lot of those decisions are made independently. And we don't know how they mix with until we done. Working with the Office of Science and Technology Policy though, they sit in on our budget dialogues with the director of OMB, and it's often the case that the science advisor will try to defend a certain program, will say, you know, either a particular agency, you know, I think is being overlooked here, or a particular set of fields, such as the physical sciences, you know, may be in jeopardy if we don't focus more on those. While, frankly, nobody ever says, well, instead of 24% basic research we should be doing 26% basic research, what we do, is, you know, a particular agency which for various reasons may be a steward of a particular field, or in the case of NSF, a basic research wobbly, we may try to say, you know, this year let's continue a push forward of greater another funding, or decrease it. Now, what that doesn't do is, I mean by far our largest R&D agency is the Department of Defense. And the small swings along the margin of DOD are entire National Science Foundations at a time, you know. And so, by percentage we'll never do that. But we can at least track how the funding profile of the National Science Foundation looks, how the funding profile of the National Institutes of Health looks, of the Department of Energies' Office of Science, et cetera. And even if those are in isolation sometimes we are trying to either defend or strengthen, or sometimes even cap off the growth, like in the case of the National Institutes of Health. We have 5 years where they doubled the budget by the end of the fifth year, and that's just hard to sustain. So, that's a decision, let's level that off.

興味深い質問をいただきました。どのようにしてこの研究のポートフォリオを決めるのか、省庁間であれ、全政府ベースであれ、どう配分をするかというご質問です。

このPARTのツールはプログラムレベルに焦点を当てています。そこで、その分析単位は既にもう決まっているのです。ある意味では研究開発への投資基準は少し弾力性がありまして、省庁レベル、あるいはプログラムレベル、プロジェクトレベルでも使用可能です。ポートフォリオ全体が役に立つのか、健全で有用性があるのかどうか、その判断が必要なのですが、それで典型的に決

定を下しているわけではありません。それが普通の決め方ではありません。私自身の知っている研究機関に絡む予算プロセスを振り返ってみますと、大体その決定は独立した形で行われているので、構成のミックスは終わるまでわからないのです。OSTPとも協力していますので、予算の対話は科学技術局、それからOMBの局長も加わってしております。

よくあるのは、科学顧問がある特定のプログラムを擁護しようとしします。例えば、ある特定の省庁、ここで忘れられているのではないか、ある特定の分野、例えば自然科学の分野はリスクですよと、もっと注目しないといけませんよとか、コメントがアドバイザーから出ます。正直に言いまして、例えば基礎研究を24ではなくて26%に引き上げるべきであるとか、そういう言い方は言われないのですが、そうではなくて、ある特定の省庁が何らかの理由があって、ある特定の分野を主導しているとか、全米科学財団であれば広く基礎科学を主導する重要な役割を果たしている、そういう例がある場合には、例えば、今年をもっと基礎分野にお金を入れていこう、もしくはこういう理由があるから減らそうという話はします。だからといいまして、これまでの最も大きな研究開発といえば国防総省なのですが、その国防総省の中でどれだけアップダウンがあるかどうか。これは全米科学財団がどれだけアップダウンするかとも多少の連動はあるかもわかりませんが、NSFの資金配分がどういうプロフィールになっているか、あるいはNIHのプロフィールがどうなのか、エネルギー省、あるいはOSTPの方の配分がどうなっているか、それぞれ連動していないこともあり、一部は弁護しようとしたり擁護したり、一部は強化しようとしたり、あるいは上限を設けることさえあります。例えばNIHに関しましては、5年経ったときには予算倍増という話が出ているのですが、これは持続可能ではないのではないか。ここでキャップにしようという話が出ているので、様々です。

【大山会長】

他の委員の方、如何ですか。

黒川議員、どうぞ。

【黒川議員】

I think that I understand it's your responsibility in the Office of Management and Budget to oversee the government managed funding independent agency and contract and other aspect, not only R&D. Now, that mean, I think this is very important, but your society have already built upon fairly open, and transparent peer-review system NSF and NIH, and as you said, DOE budget is very huge for the R&D. Therefore, I think you are focusing more on that kind of big swing, and the NIH, I think you are more making executive decision not

necessarily going into the detail of the evaluation, because you guide not an expert to evaluate the R&D. Is that my understanding?

恐らくOMBにおきまして、いろいろな官庁の予算配分を、R & D分野ではなくていろいろな分野において統括するのが責任だと思います。それは非常に重要だと思うのですが、また加えて、非常にオープンで透明性の高いピアレビューシステムを政府、NIHで確保した。OMBのR & D予算はかなり高いので、そういった変動を見えています。NIHがエグゼクティブな決定をしており、詳細の評価には入っていないというのは、皆様はR & Dの評価のエキスパートではないから、それはやっていないという理解で正しいですか。

【トリンクル氏】

That's essentially correct. We are to the extent that we do evaluate research, I mean; a number of the examiners within the office of the budget , health science are their member's background. But obviously we are not there to second chance the more qualified independent assessments that were made.

本質的にはそうです。我々OMBのエバリュエーターの何人もが科学、工学の分野の専門家ですが、ただ、もっと独立した人たちに評価をしてもらった後に、その人たちがセカンドオピニオンを出すという方たちです。

【黒川議員】

Can I have another question that exactly gets back to Dr. ESAKI's question? How much the cost incurred in this one billion budget spending is obviously inherent to your budget, your personnel, and also, I think, with understanding inherent for NIH and NSF, evaluating systems. Is that correct?

江崎先生の質問に関連するのですが、10億ドルの予算はOMBの人材費予算も含みでありますか。NIH、NSFの中では別個のシステムが内在しているということよろしいですか。

【トリンクル氏】

That's correct, although some of the agencies feel that this, you know, the term we sometimes use is an unfunded mandate, where we have more requirements but the same level of budget to do them.

それはそうですが、幾つかの省庁は、我々が使う言葉はアンファンデント・マנדートということで、より大きな要件が課せられているのに、その分の予算が追加的につけられていないということです。

【黒川議員】

So I think you evaluate each other justifying their own budget.

では、お互いに自分たちの予算を正当化させるということですね。

【トリンクル氏】

That is correct.

そうです。

【大山会長】

他の委員の方、如何でしょうか。

秋元委員、どうぞ。

【秋元委員】

In page 18, you have mentioned that classified result of a PART assessment, and you evaluated Oil Technology assessment was ineffective. And Geothermal Energy was moderately effective. And what is the main reason such a derivation of the evaluation was occurred in the same energy area. And also, second question, if you find it ineffective, what will you do? You give it, or you try it again with using a different system of the assessment or not, please?

18ページの中でPART結果を分類しています。石油の探査、あるいは地熱エネルギー、それに関しまして、最初の方は効果が低い。一方で地熱の方はそこそこ効果的で、例えば同じエネルギー分野であるのに、化石燃料と、あるいは地熱で、なぜこういう評価の違いが出てきたのでしょうか。また、効果が低いということになった場合にはどうなさるのですか。そのままにしておくのか、あるいは異なるシステムを使って評価をし直すとか、どうなさるのでしょうか。

【トリンクル氏】

Very good questions. First of all, in the question how do you select a program, how many programs per agency, or anything like that. We did not bother with discussion of how to define a consistent unit of analysis. We weren't able to do that in any that would make sense, so to some degree we left that to the judgment of the individual OMB staff and agency to work out together. In some areas where, for example the National Science Foundation is put in many different fields, but the processes to do so are essentially uniform across the agency. We don't really need that many programs necessarily in order

to do that, and other agencies for whatever reason are not so homogenous, they have many different types of activities in many different areas. And that may be a reason that you define a program in a smaller way, that's the easiest way to do this, the most effective way in order to understand the specific issues with that particular effort. Within DOE, Department of Energy, they are, I think, they did define programs smaller than average. Department of Defense defined programs larger than average, and that, it does effect how the assessments work, because it, at larger scales programs do tend to be a little more strategic in their outlook. But nonetheless it's, you know, we had about the easiest way to do these, the most practical way. As to the question, what do you do when a program is ineffective? It depends on the program. Within the Department of Energy, many of these programs are defined in law, or are even like, for example, in Clean Coal Technologies Program, was a presidential priority, but yet we find that the program isn't demonstrating results of its efforts. So we are given a choice there, and based on the assessments of the individual questions that make up the PART, we in the oil technology I happen to know, their budget was reduced somewhat, at least in the '05 request. And there are new management expectations communicated with them, and I confess I'm not terribly familiar with that program, other than for example, I always use of an ineffective program. But I do know that its funding was reduced somewhat. Now, whether that's, that has 2 effects. In one case for spending less now on a program that isn't producing, so right there we're perhaps more efficient. But since I believe this program has been specified by the congress, we are also putting the program on notice that we expect more from it, and that we hope to see the program do better in the future.

いい質問をいただきました。まず、どのようにしてプログラムを選択するのか、各省庁で幾つのプログラムを選定するのかなどですが、それに関しては、どのように一貫性がある数进行评估すべきかを決定していません。それは意味がある形ではできないと思っていたので、それについては個別のOMBの職員、あるいは各省庁に任せました。一部の分野、例えば全米科学財団においては多くの分野をサポートしているのですが、そのためのプロセスはいわば一本化されていまして、余り複数のプログラムを見なくても評価は可能です。他の理由で、他の省庁においてはそこまで均質ではない場合があります。複数の活動が複数の分野で行われていることもあるので、場合によっては、あるプログラム

の定義づけがより小規模な形で行われているという場合もあります。エネルギー省に関しては、プログラムの定義づけについても小さ目になっているかと思えます。国防総省の場合は大き目だったと理解しています。その結果、アセスメントにも影響は出ます。より規模が大きなプログラムであれば、見直しにもより戦略性が出てきます。

とはいいながら、実行可能な、そして効果的な方法ということであれば、現状こうなっております。もし効果がない、低いと思えた場合にどうするかは、プログラム次第です。エネルギー省においては、多くのプログラムは法律で定義づけされていたりします。あるいはクリーンコール技術、これは大統領の中で優先順位は高かったのですが、プログラムを評価したところ、その成果が十分に見えていないということがわかりました。そこで幾つかのチョイスが出てきます。幾つかの質問がPARTに含まれていまして、その質問への答えを見ていきますと、石油プログラムに関しましては、たまたま私は知っているところがあるのですが、予算がある程度減らされました。少なくとも2005年の予算要求という意味では減らされまして、どういう期待が持たれているのか再確認がされました。それ以上詳しく知っているわけではないのですが、余り効果が高くはなかったということで例に出してしまっているのですが、ただ、ある程度予算が減らされたということまでは承知しています。

2つの影響があります。一方でお金は減らしたわけですが。生産性が余りないということであれば、お金は抑えている。ただ、このプログラムに関しましては、議会でも必要だと確認がされていますので、もっと頑張ってもらいたいと期待はしているということは述べてありまして、将来への期待についてのコメントは付されています。

【大山会長】

黒田議員、どうぞ。

【黒田議員】

I'm interested in "criteria for all kinds of R&D" explained on page 7 onwards. As far as I understand, the criteria for quality and performance are not that different in this country. Here, the quality sometimes is close to outcome, however quality is quite separate from outcome or performance in this presentation. You said quality is a merit-based process, and performance is actually the outcome. Could you clarify more in detail, what's the difference between "quality" and "performance"? Because I think, many people think that, quality must be the outcome.

7ページ以降に説明されているあらゆるR&Dに対する基準に関心を持って

おります。私が理解する限りにおいては、クオリティーとパフォーマンスの基準はそれほど我が国では違いません。我が国ではクオリティーは成果に近い。でも、このプレゼンテーションではクオリティーはメリットに基づいたプロセスであって、パフォーマンスが成果であるというようにおっしゃいました。もう少し、このパフォーマンス基準とクオリティー基準の違いについて教えてくださいいただけますか。多くの人たちは、クオリティーは成果のことではないかと思うと思うのです。

【トリンクル氏】

That's a good question. If it helps, there is the same confusion in the United States. How are these different? I think, the way we are thinking about it is, quality is a characteristic of the program, the performance is what results from that characteristic. So, quality means that, is essentially process, the focus is process. And it doesn't have to mean that the program is, you know, a competitive merit-based program, because in some cases, particularly in the Department of Defense, it may not make sense to compete things wobbly. You may know who your contractor is, and et cetera. But at least they have to have some way of maintaining quality and justify that that's appropriate. So, we have adopted language that isn't necessarily clear on its own, but through guidance we try to make it clear that with quality you are talking about a process itself, and performance we are talking about the documented results of the program. But that is a good question.

いい質問です。ありがとうございます。

ちなみに、アメリカでも同じような混乱があります。どう違うかということですが、我々の考え方としては、クオリティーはプログラムの特徴であり、パフォーマンスはプログラムの結果、その特徴の結果生み出されるものは何かということです。つまり、クオリティーは本質的にプロセスにフォーカスを置いています。必ずしもプログラムが競争上に有能であるということでもありません。例えば国防総省のプログラムであれば、もう下請業者はどこかということがわかっているかもしれないということがあると思います。でも、クオリティーを確保するために、それを正当化する何らかのデータが必要かもしれないということがあります。ですので、何とか我々は文言を使ってはっきりとさせようとしているのですが、クオリティーというのはプロセスそのものであって、パフォーマンスというのは文書化されたプログラムの結果、成果であるというようにしております。

【黒田議員】

Can I just ask further? Can you give me one example, in which the performance is good and the process is not that good, except for on defense issues?

ちょっと追加的に伺いたいのですが、例を挙げていただけますか。つまり、パフォーマンスはいいけれどもプロセスがよくないのは、国防分野以外で何かありますか。

【トリンクル氏】

Ya, that may be difficult to come up with. I think, the reason, and I probably could if I thought about it. But just to explain that the reason that we focus on quality is primarily that in an ideal world we could find we could expect performance everywhere and show it after the fact. But particularly at basic research again, since we can't do that we lean a little more on process type issues, on structural issues. And I can come up with examples of programs that have good qualities that haven't yet really demonstrated performance, as any young researcher here is probably in that area. I think, that probably all programs in the government where perhaps we have, like within the Department of Defense, or within the Department of Energy, where a national laboratory is doing research, we never computed out, and yet it's, you know, they have a Nobel Prize to show for it, and, you know, we don't want to say that's a bad thing. But if, you know, a merit based competitive process is essentially a market mechanism for making sure that the dollars are flowing in an effective way. And we don't want, you know, we don't want to restrict national labs from doing, you know, things that have been successful, have shown themselves to be successful in the past. But we just want to make sure that we are encouraging, you know, best quality types of techniques and processes.

よく考えれば思いつくのかもかもしれませんが、ちょっと思いつきません。ちょっと説明しますと、なぜクオリティーにフォーカスするか、その理由は主に、理想的な世界であれば、恐らくあらゆるところで成果を期待して、事実をもって証明できるのだと思います。ただ、基礎研究に戻りますが、基礎研究ではそれができないので、プロセスにより多く依存しがち、そして枠組み、構造的な要素に依存しがちです。よいクオリティーがあるが、まだパフォーマンスを証明していないプログラムというのはたくさんあります。歴史のないプログラム分野はすべてその範疇に入るのではないかと思うのですが、恐らく中には政府

のプログラムとして、例えば国防総省内、エネルギー省内で国立研究所が研究をやっている、競争入札が行われたことはないがノーベル賞を受賞したということもあるでしょう。それは悪いということはいいたくないのですが、メリットベースの競争プロセスは市場メカニズムでありまして、十分に投資された額が活用されているということを担保するための方法であります。国立研究所が過去において成功例として上げた業績を阻害するようなことはしたくないのですが、こういったテクニックやプロセスを奨励しているということは明確に打ち出していきたいと思っています。

【黒川議員】

I think, from reviewing your performance, or with your department review by the last two years in Bush administration, in your feeling, you contracted out many research funds, or forwarded funds to corporate sectors, DOD and DOE, particularly. But compared to NSF and NIH, those are more competitive, open, transparent process, you wonder more in details on this contracting research process, and also national research agencies that do not necessarily go through a peer-review system. Because there the duplication may be and where waste of money you may find. And in your feeling, in general, let's say contracted research and some contracted work, through the corporate then always social scandals, I also find, in your general feeling, you tend to trust the NIH and NSF more, than the national research lab, and than corporate research with the big money.

パフォーマンスのレビュー、あるいは皆さんの省庁がブッシュ政権のもとで行ったこれまでのアセスメントをした結果ですが、多くの研究資金、あるいは公的資金を民間部門、例えば特に国防総省、エネルギー省からは委託していると思うのです。NSF、NIHといった、もっと透明でオープンな競争的なプロセスと比較いたしますと、受託研究、そして国ベースでのものに関しましても、もっと詳細なものが必要で、必ずしもピアレビューを行わないというものも必要かと思えます。そこでは、例えばむだ遣い、あるいは二重投資が発生するかもしれません。個人的な感触をお聞きしたいのですが、委託研究、特に民間へ委託をした場合、よくスキャンダルが出てくるかと思うのです。そこで感触をお伺いしたいのですが、これまでの経験に基づきまして、NIH、NSFをより信頼し、次に国立研究機関、民間への委託、こういう順番になるのでしょうか。

【トリンクル氏】

Ya, I think that's, I would say that's accurate. We absolutely, you

know, our large government contractors are very important to our system, and I don't mean to suggest, you know, that there isn't a trust level there. But any time that the, you know, when money leaves the government for a deliverable, in the case of a contract, that's for each contract and subcontract that follows, that's a, the control and the oversight, get watered down, get reduced. So, it's just harder to crack those. And as far as; I do agree that our relationship with industries, probably different from here, for example, and the, what we ask ourselves is, you know, does, is there something that the government should fund, that the industry wouldn't do anyway. Often, it's hard to say, but it's a question that we want to continue to ask.

大体そうだと思います。委託機関はとても重要な存在です。我が国にとっても重要な存在ですので、信頼関係がないとは言いませんが、資金が支出された場合、国としてお金を使う。例えば委託研究があっってお金が出ていきます。委託があっってサブの委託先もあるわけですが、下請もありますが、離れていけばいくほど制御、監視、あるいは監督の力は確かに弱まります。そういう意味で、国としてトラッキングをするのは一段と難しくなっていきます。

おっしゃるように、政府と民間との関係は、例えば日本の産学とはまた違うと思います。我々自身も自問しているのですが、これは国としてお金を出すべきであって、民間だけでは行われまいだろうかと考えてみてわからないことが多いのです。ただ、これは考えることは必ず続けていきます。

【畚野委員】

You have mentioned you made 58 R&D out of 400 ; is this a normal ratio? And also, how did you choose which one should assess or not?

400のうち研究開発テーマは58を評価したとおっしゃいましたが、比率としてはこれが平均的な比率ですか。また、どれを評価するかということの選択はどのようになさるのですか。

【トリンクル氏】

Yes, I am not sure what the actual ratio is. I do know that the, let's see, of the discretionary budget in the United States R&D is about one seventh. I do know that, because these R&D assessments are a greater challenge for us and for the agencies, many of the agencies have thought to not do those first necessarily to get out of the way some of these other assessments, to get to understand how to use this PART tool, and plus to have a little more chance to develop their measures, you know, before the assessment happens. Fair enough.

As far as how, these programs, these specific programs for selected, there is no strict guidelines other than the agencies and the OMB examiners negotiate which program should be assessed. In the case of the National Science Foundation, I think our biggest dialogue was, what's a program, how do you find what a program is. And then, once they showed that they had a, kind of a representative set of activities across the agency each year, that's all I need to see. With other examiners, they may have a particular issue, a particular program, and that might be a reason to assess it sooner, with other agencies the examiner may feel, well this isn't really ready to be assessed, let's focus on something else. So, it's different for each agency.

実際の比率がどうだったかということは把握しておりません。ただ、その裁量予算のうち、研究開発は7分の1ほどであります。また、これらの研究開発の評価は、我々にとって、また省庁にとってより大きなチャレンジですので、多くの官庁は必ずしもそれを優先するのではなくて、PARTツールを理解するのにもう少し時間をかけて、独自の評価を開発してアセスメントをするという順番にしております。どのようにその評価対象プログラムを選んだかということについては、厳格なガイドラインは特にございません。ただ、省庁とOMBのエグザミネーターが交渉して、どのプログラムを評価するかということを決めます。NSFの場合、最も大きな議論が、プログラムの定義をどうするか、何をもちいてプログラムとするかということだったと思います。NSFは、毎年行っているということで代表的な活動のリストを提示いたしました。他の評価員は、特定のプログラムを問題視しているから、早く先に評価したいという意見を持っている人もいます。他の官庁では、まだ評価の段階に至っていないから、これは先送りして別のものを先に評価しようということもありません。省庁によって違います。

【江崎委員】

I'd like to continue on to Mr. Akimoto's questions of the assessment, the PART assessment. One I was rather surprised the results were not demonstrated the Dept. of Agriculture and the Food Safety Research. Is that you said in adequate of this one? Well certainly you export many products oversea. This may connected with the mad-cow, we stopped to import the beef of this one. This is the one question. You have any comment on this one. Is somebody evaluate you, I mean effectiveness of PART process?

秋元委員がおっしゃった、PARTによるアセスメントの質問の続きです。結果が示されていないということで、農業、食品部門、これが該当していると

いうことを意外に思ったのですが、つまり、効果が不十分である、パフォーマンスが十分であるということですよ。多くの食品を輸出しているかと思うのですが、狂牛病で今、牛肉の輸入が差しとめになっています。こういうことも鑑みて、何かコメントはないでしょうか。それが第1の質問です。

第2の質問ですが、皆さんを評価している方はあるのですか。評価方法の評価をしている機関はあるのでしょうか。

【トリンクル氏】

First of all, I did not explain the results-not-demonstrated assessment. We talked four assessment outcomes are effective, moderately effective, adequate and ineffective. Those are based on whether or not the programs have results. The results-not-demonstrated category, we have shouldn't put at the bottom of this chart, I should have put it on the side of the chart perhaps, and that is to say that, we could not make an assessment. Primarily, because the performance measures weren't well defined. I think, the program received a score, but because new measures were needed we said, OK, let's hold off and coming up with one of these term for another year until the program can develop its measures better, there is a, programs are expected to not be in this category for more than a year or two before it can move into one of the other categories. And in the food safety research example, I don't know the details, but I do know that in a lot of the programs at the Department of Agriculture there was a lot of debate between that part of OMB and that part of or in that agency, whether the agency should be held accountable for the outputs of the research, which are, you know, techniques and research conclusions, or whether the agency should be held accountable for incidence of food borne illness. You know things that are actually beyond the control of the research group within the agency. I don't know where that conversation currently is, but that particular issue, I think, is true of agriculture, of health, of environment, you know, how much do you owe the agency or the program accountable for the open net outcomes that you want to see. So this is a measurement challenge, and I don't know the status of this one. I appreciate your concern for the US food at a point there. As far as whether OMB has been assessed, I can tell you, we did an informal assessment inside the building, which is not what we would have accepted for other agencies. We did not, we found ourselves to be lacking those. So I think, this

process is part of our trying to improve our ability to do what it is that we feel that we are supposed to be doing. A number of agencies have asked that very question, though, so you are not alone in this.

この一番下に書いてある実績が出ていないという部分、この説明はいたしませんでした。上から4つの項目の分類、エフェクティブ、中等度に効果的、十分、効果がないと、これに関しましては、このプログラムの結果実績が出ていないか出ていないかです。結果が示されているかという分類については、これは下に書くというよりはわきに置くべきだったかもしれません。なぜかといいますと、つまりアセスメントができなかったのです。そのできなかったという主な原因は、そもそもパフォーマンスを測定する指標が明確ではなかったのです。あるスコアがついたはずですが、新しい測定方法も必要なので、ちょっとわきに置きましょうと。あと、この同じ条件で1年やってみて、よりよい測定方法を考えていこうというような状況です。1年以上このまま、このカテゴリに入ったままではいけないということにはなっています。他の4つに入っていかなければいけないのです。

食品安全の研究分野に関しまして、私は詳細はわかりません。個人的には把握していないのですが、農務省が行っているプログラムの多くのものに関して、OMB、あるいは省との間で議論がありまして、そもそも農務省、あるいは省庁が研究結果に関して責任を負うべきであるのかどうか。例えば技術であるとか、その結論に関して省庁自身が責任を問われるものであるのか。あるいは、食品を媒介して食中毒が起こった場合に、それに関する責任は問われるものであるのか。これは、その省庁の中にあるリサーチグループの責任を超えるものではないかという分野もあります。この協議が今どうなっているのかまで、私個人としては詳細はわかりません。ただ、農業とか健康、保健、環境、こういう分野に関しましては、そのプログラムもしくは省庁がどこまでその成果の責任を問われるべきであるのか、負うべきであるのか、ここに問題が残っていて、どう測定をするかという課題で残っていると思います。ただ、米国の食品輸出に関して心配してくださって感謝します。

一方で、OMBに対する評価があるのかどうか。我々は、インフォーマルな評価は内部ではしています。これは、他の省庁がこうしたといっても私たちは認めないのですが、我々としては不足している部分があると自省する部分が出てきました。この自己評価のプロセスも重要で、我々としても、我々がすべきことをしているかどうか、その評価は必要だと思っていますし、他の省庁から自分たちはどうなんだと聞かれていますので、先生だけがそういう疑問をお持ちなのわけではありません。

【黒川議員】

I think, this relates to Dr. Esaki's question, because OMB, when

government, or the president's office may change, a determined position will change, that's clearly a responsibility of many political policies appointees. Is this correct?

これも江崎先生の質問と関連しているのですが、OMBは政府、あるいは少なくとも大統領がかわるときに、そういったところでかわるのですよね。トップにおいては政治任免のスタッフもあるのですよね。

【トリンクル氏】

No, that is correct. The office management and budget is unique in a sense that of the 400, 500 people that work there are 98% are career staff, like me. Many of us were in the last administration as well, and we are the corporate memory. We remember things from previous efforts. Whether the PART tool would still exist in the next administration, we'll have to wait and see. But I think, a lot of the practices that we have developed are essentially, you know, in our minds good government practices that, you know, maybe less formal in a definite administration, but we would expect to continue.

そうです。ただ、OMBはちょっと独特な政府機関でありまして、400とか500人ぐらい職員がいるのですが、そのうち98%は私のようにキャリアスタッフです。ですから、前民主党政権におきましても私どもは職員を務めておりましたので、以前どうやっていたかということを経験している人間が多勢を占めております。PARTツールが次期政権においても踏襲されるか、まだわかりませんが、我々が開発してきたやり方は、政府の慣行としては評価すべきやり方だと思います。次期政権におきましては、これだけ形式化されたものにはならないかもしれませんが、基本は踏襲されることになりましょう。

【國井委員】

I'm from industry. Industry is mission oriented programs. In an industry, technology transfer is very important. I would like to know how you assess the effectiveness on the transfer, or how you improve such practice your assessment.

民間から参加しているので、ミッションオリエンテッドなプログラムに関心があります。技術移転が産業界にとっては極めて重要なので、皆さんのところでは技術移転の効果、あるいはその評価をどうなさっているのでしょうか。そのアセスメントの中でどのように促進をしようとなさっているのでしょうか。

【トリンクル氏】

I would say that technology transfer to the extent that a program's

goal includes that, or is technology transfer, that that should be reflected in its measures. I have to say it's not an area that, for example with the National Science Foundation, that doesn't tend to be their focus. But for NASA or Department of Energy, or Department of Commerce in the US, they do have a mission to, and the Department of Defense of course, they do have a mission to transfer technologies to that the private sector and get, you know, it's thought that if some of the technology research doesn't get out into industry it might be going to waste, you know. And so, I have to say I'm not very familiar with the measures that have been used for that, but I know that if that's the goal of the program, that's how those measures should be defined. So, I'm sorry, that's not a good answer.

技術移転に関しては、あるプログラムのゴールにそれが入っているのか、あるいは、ゴールそのものが技術を移転することなのであれば、それは評価の中にも含めるべきだと思います。その領域は、例えば全米科学財団に関してはそれが狙いであるということは余りありません。しかし、NASA、エネルギー省、商務省の場合は、その使命の一部として、国防総省もそうですが、もし技術研究の一部が民間に移転されないのであれば、無駄になるという考え方は存在しています。私、個人的には、これに使われた指標は詳しくは知らないのですが、プログラムのコードにそれが入っているのであれば、当然その測定指標についても定義がなされているべきだと思います。余りお答えになっていないと思うのですが、申しわけないです。

【大山会長】

他は、如何でしょうか。

【大山会長】

One question please. You indicated here the several plans of criteria, like investment criteria, or R&D criteria. What kind of effort to do make to get the mutual agreement among the OMB and all agencies?

一つ質問します。投資基準とかR&D基準とか、いろいろな基準を使うとおっしゃいましたが、OMBとあらゆる官庁での合意を取り付けるのにどういったプロセスを経るのですか。

【トリンクル氏】

That's an interesting issue we have to deal with. Ultimately, as I say, many of these sorts of things are what OMB examiners will do all the time, in a way through normal processes. The OMB examiner have

the last word, and there may not be negotiation in many cases on what the examiner feels to be the case for the program. The PART, even though its results are more public and out there more, it actually gets more of an avenue for this discussion, and for all the back and forth between the agency and the examiner where it actually does, I've heard a number of examples where the examiner didn't know what some of the activities, or some of the assessments of the program, and this brings them to light. Each year we have an appeals process. First of all, there's a negotiation where the part might start at the agency, or might start with the examiner, and then have moves back and forth as, as. No, no. I don't think you demonstrated that you can answer yes to this question, please provide more information. And the agency will do that. And, you know, they still have to convince the examiner, but if they can't do that, and you end up with a score, but then there is an appeals process where the agency can say that, we don't feel that this assessment was consistent with the guidance, we don't feel that this assessment was consistent with how that agency over there was assessed. So they, and I think this appeals process has shifted a little each year, and I don't know what's in the guidance this year. The first year was an external process, last year it was combination of a special part of OMB at some external entity that would look at all these parts to say whether they were consistently and fairly assessed. That is definitely a component of this whole effort that, I think, is a learning process and we are going to continue to figure out how to come to agreement. I can tell you that there are some agencies that still disagree with last year's assessments, and that will be true, I am sure.

これは、我々も対処しなくてはならなかった一つの関心深い課題ではあったのですが、結局はOMBの評価員は、PARTがあってもなくても、いずれにしても行う作業なのです。通常のプロセスにおきまして、OMBの評価員が最終決定をする、最終の発言権を持つ。そのプログラムのOMBエグザミネーターの評価に関しては議論の余地がないということがしばしばあります。ただ、PARTはもっと公のプロセスでありまして、官庁とエグザミネーターの間での交渉が可能なプロセスであります。多くの例で実際に評価員がプログラムの活動とかアセスメントを、以前は知らなかったが、PARTがあってやりとりすることによって知ったという事例があります。毎年、まず交渉が行われまして、PARTは評価員が始める、あるいはエージェンシー側で始めるということがありますが、そのやりとりがあります。そちらは、この質問に対してイエスと答

えているが、それを証明できていないからもっと情報をくださいと言うと、向こうの担当官庁からそれを出してくる。もちろん評価員を説得しなくてはならないが、説得できなくて不満なスコアがついたとすると、それに対して異議を唱えることができるのです。官庁は、この評価はガイダンスにのっとった評価ではないと思うという異議を申し立てることができます。この異議申し立てプロセスは毎年若干変わっておりまして、今年のガイダンスはどうなっているかは忘れましたが、最初の年は外部で判断します。2年目は若干外部の意見を聞くが、OMB内で異議申し立てを処理するというプロセスでした。つまり、内部で基本的に公平な判断がなされたということ調べるというやり方でした。これは学習プロセスで、いろいろこれから試行錯誤していってみるということだと思います。官庁の中には、昨年のアセスにいまだに不満を唱えているところもあります。毎年そういったところが出てくると思われます。

【大山会長】

Thank you.

ありがとうございました。

時間が参りましたので、これでディスカッションの時間を締めさせていただきます。

ここで5分間休憩させていただきますして、10時40分から後半の審議を始めたいと思いますので、よろしくお願いいたします。

[休 憩]

【大山会長】

時間が参りましたので再開いたします。

それでは、ルーク・ジョージューさんから、主にヨーロッパの状況についてご説明をいただきます。30分ほどでご説明をお願いいたします。

【ジョージュー氏】

Thank you very much, chairman, and good morning everyone. I will sit down to speak. Today, I will be speaking not about the evaluation of programs or institutes, but rather about the evaluation of policies for research and innovation. I will be taking some specific examples which are looking at whole system evaluations, and my examples will come from two countries, from Finland and from the UK. On slide number 1, I am discussing the purpose of evaluating at this high level. Many countries now use the so-called systems of innovation perspective to design the research systems. The systems perspective

which came from OECD, and from academic writing those attention to the institutions that support and carry out research, and the connections between them. When we look at these, we can then ask questions about what are the gaps, what are the bottlenecks, and what are the opportunities for learning. And one of the main benefits of this approach we have got, it shows us that research and innovation policies are interactive, and they either reinforce or even block each other. And just to give you an example, on page 2 there is a chart which I prepared for a panel, which I was chairing on behalf of the European Commission. This panel was looking at ways in which we could use public policy to increase R&D spending in Europe to 3% of gross domestic product. I am not going to go into the detail of the diagram, only to show that we have many different types of policy measures that we can use. Some of them work on the supply side, on the left hand side is more basic research funding, we work through the support for industry in various ways. We also have demand side policies, which increase the demand for innovation, and therefore for research. And along the bottom, very importantly, framework conditions, where government policies more generally influence research performance. The production of human resources, the intellectual property framework, competition in market regulations, and so on. And if these policies are not in some way coordinated, we sometimes cannot make progress. For example, if you wish to stimulate strongly an increase in R&D without a human resource policy to increase the supply of researchers, either through training, migration, or something else, you could end up not being able to implement your policy. Moving them to the direct case studies of evaluation, I'm picking two evaluations of the Finnish system that I've been involved in the last few years. The first of these begins to be described on page 4. This is called "Assessment of the additional appropriation for research". The purpose of this evaluation came from a policy decision made by the government of Finland in 1996, to increase its R&D spending by 25% over the next two year period. This was quite a brave decision by that government. It came at a time when the country's deepest recession following the loss of its markets in the former Soviet Union. When they made this decision they decided to appoint a panel to evaluate the use of these funds, and the effect on the Finnish economy. The panel had two foreign members, one from

Germany and myself, and it is from this experience I am speaking. On page 5 you will see the main problem that the panel faced. It's a box which shows on the top row research over time and on the bottom row the economy over time. The question we have set is represented by the arrow with dotted lines. What is the effect of the new research spending on the economy currently? Well, clearly this question could not be answered directly because of the time lags involved in research reaching its effect in the economy. So, what we had to do in this evaluation was take a series of proxy or substitute measurements. We had to look at the effect of research carried out in the early 1990s, on the economy of that time, the diagonal arrow, and previously as well. And we also had to see whether the new funding had caused changes in the research system, or indeed whether the economy had changed in that time, the red arrows. So that was our conceptual framework. This was not only a panel exercise, we were given a substantial budget for supporting studies, we commissioned various teams of experts to gather evidence for us. The evidence included a bibliometrics, a whole series of economics and econometric studies, and a review of all the past institutional and program evaluations that had taken place in Finland. Those are listed on page 6. On page 7, we summarize what we did ourselves. Over a two-year period we carried out extensive hearings, we talked to heads of ministries, of the major companies, such as Nokia, universities and research institutes. I can leave behind a copy of the summary report which I have with me. I'm not really here to talk about the results of this evaluation, only about how we did it. But on page 8, there is this one example of one of our conclusions. We discovered, well, our opinion was that the resources were being excessively spent on technologies and science targeted towards high-tech industries, the so-called new economy. Analysis shows that these sectors were not sufficient for Finland to maintain its competitive success. The country would remain dependent for many years to come on its traditional industries, on wood, on forestry, and so on. And therefore it is very important that research and technological innovation were also targeted toward areas of science that were a concern to the old economy as well, to modernization. There are many other conclusions and much evidence, but I won't go into those. Just briefly, on the impact of evaluation, on slide number 9, this report was presented to a high level audience, two cabinet

Ministers, a copy given directly to the Prime Minister as well, at least in Finland, had significant media coverage. And the government decided to continue its policy of additional funding for a further two years. The agency, the research funding agency, also changed their policies in the light of these recommendations. I move to my second case from Finland, on page 10. This is an exercise which has only recently been completed. It began in June, 2002 when the Ministry of Trade and Industry decided that it needed a new evaluation, looking this time more at the Finnish innovation system, and particularly at the balance between the agencies and networks within that system. This time the panel was only four people, two foreigners and two from Finland. The key stakeholders, instead of being in the panel formed more of a support team, a consulting panel that we could speak to. On page 11, there are some of the basic theoretical questions, which underpinned this evaluation looking at what the reasons were for the government to be intervening to support innovation, whether there was an underlying justification, a market or system failure, whether, if there was one, the policy maker could do something about it, and then, what specific measures could be taken. Again, perhaps I can illustrate the kind of work we were doing by showing some extract from the results. On page 12, you can see a map of the science and innovation support systems Finland in the 1980s. Along the bottom axis we move down the chain of basic research, applied research, product developments and commercialization, and on the vertical axis the type of support that is given, the ground loans, equity, and services. You will see, at that time it was arranged pretty much on a linear model basis, and the weight of support was quite heavily down the stream towards support for product development and commercialization. When we move to the current situation, shown on page 13, we find a very large change. Almost all of the support agencies have broadened their mandates, most of them have moved upstream in what they do. So for example, Tekes, the agency which supports industrial research and collaborative research between universities and industry has shifted its center of gravity towards the research end, it's now overlapping strongly with the Academy of Finland, the basic research funding agency. The venture capital supports agencies, also are moving strongly towards so-called funding, the so-called valley of death, and away from market type investment.

And we also found that startup firms were typically getting support from several of these agencies at the same time. They didn't see it in any way as a sequence. One more type of analysis on page 14, that's a map of where the funding is for different types of activities, and it's allowed us to look at overlaps between agencies. From an evaluation point of view, an overlap is an interesting question. It could be negative, it could mean there is wasteful diversification, or it could be positive, because it means there is a space in which agencies can work together. And in particular, overlaps between Tekes and the Academy of Finland were very productive, because they represented joint programs that were bringing industry and the universities together. On page 15, a few observations on the methodology. This was an unusual evaluation, because it went all the way through from looking at the basic rationale for government intervention to demonstrations of how it works in practice. And we found, it was necessary to this, in fact, close examination to really understand the rationale for these agencies. And simply asking one produced very routine textbook responses. Finland is a good place to do evaluation, because they have excellent statistics covering almost every site of every firm in the country doing R&D. It's possible to do work, but maybe more difficult, in other places. And on page 16, a couple more observations. Finland normally comes top with world competitiveness indicators, and yet there is still a strong desire on the part of its senior government officials to carry out evaluation and make it even better. One last point, even looking at a whole research system sometimes is not enough in terms of scope. We found ourselves to constantly struggling with what are called here the boundary conditions. In reality, it's very hard to separate research and innovation policy from other areas of public policies, such as education policy, fiscal policy, and public purchasing. Moving on to the examples from the UK, I'm going to talk about some rather different things here. First of all, I'm going to talk about the general framework for evaluation in the UK, and then give one example of a policy level evaluation. The general framework, I'm beginning on page 18. Since the election of the Labor government in 1997, there has been a system which is called Public Service Agreements, where public spending is all covered by a system of targets. Each public service agreement covers one government Ministry, and contains a statement of the aim of that Ministry or

department, its objectives, and then a series of performance targets, plus a more general value or for money or efficiency fact target for the Ministry, and a statement of who is responsible for delivery. This bears some similarity to the government performance and results act in the USA, except it is not a parliament based system; it is a government based system. So, the same government accepts the targets also receives them, and is expected to act upon that information. Across governments as a whole, there are 160 targets, which are then made more detailed in what are called Service Delivery Agreements. And I'll give you an example of how those effect science in a couple of minutes. Before that, I also wanted to give an overview of the many different types of evaluation, which go on in the UK at different levels. On page 19, I have a slide that's called the hierarchy of evaluation. It is a hierarchy in the sense that it occurs at higher or lower levels, but not in the sense that they are very well connected to each other. At the top of the system is a Ministry level evaluation. I've already mentioned public service agreements. In a few minutes I will talk about the spending review system, where the government sets its budgetary program every two years for the next three years, on a rolling basis, in an attempt to break out of the one year public spending cycle which inhibits planning. At a level of agencies, all public agencies are subjected to what is called a quinquennial review. The word "quinquennial" means, every five years, which tells you how often that takes place. This is a review of the rationale for existence of that agency and of its mission statements. Moving down the system to evaluation of institutions, we have the very well known research assessment exercise, looking at university research. I think, that was covered in the reports you received earlier, and it's also true that this of research council institutes, public laboratories doing basic or strategic research are evaluated in quite a strict way. We have a field level evaluation, they are becoming more common nowadays, where an international panel is supported to give a report on the health and quality of an entire field of science. We have recently had reviews of physics, of chemistry, of computer science, for examples. And QPIE is another internal system which I won't go into today. The next one is program evaluation is the area of greatest activity and action. It's not what I am talking about today, though I am happy to answer any questions you may have about it. And as well as that, we have various

thematic crosscutting evaluations going on. Now, coming to a case study, I've reduced it to just one slide on page 20, this is the Department of Trade and Industries' experience of public service agreement targets. The Department of Trade and Industry is the Ministry responsible for science spending in the UK. The budget of the research councils pass through that Ministry. So, it has six targets all together, two of which relate specifically to science, and they are listed on this page, PSA5 and PSA6. One of them is an international comparative target, to improve the overall ranking of the UK science and engineering base, as measured by international measures of quality, cost effectiveness and relevance. And the other is concerned with knowledge transfer, increasing the level of exploitation and technological knowledge derived from the science and engineering base, as demonstrated by a significant rise in the proportion of innovating businesses consulting such sources. These targets have changed over time between public spending reviews. There has been some learning also in how to write the target. The first time they had a target for science, it was to increase the number of spin-offs from universities by 50%. It was immediately realized that this was a trivial target, that could be met by manipulation, rather than by any real improvements in performance. And now, there is an attempt to make the target more substantial, and to derive from them a series of rather specific sub-targets. On the next page, should be 21, the second case is one of a policy-level evaluation. This is an evaluation which takes place in the context of the spending review I mentioned earlier. It's called the Cross-Cutting Review of Science. We have two of these so far. There is not a guarantee that it will take place in every spending review, but because science is very essential to government policy in the UK at the moment, essential to economic policy, it seems likely that it will continue for the time being. This is a high-level review carried out by Her Majesty's Treasury, our Finance Ministry, although it is under the leadership of the Minister for Science, who comes from the Department of Trade and Industry. This spending review is a system which is meant to set overall priorities for government spending. And so far, science has done well after this review. The method is not all that systematic. It consists of a literature review in collation with evidence. There are some supporting studies commissioned, although they tend to be different from one review to another, and the key

people involved consult quite extensively with the main stake holders, with leading research universities, with industry representatives, with the medical charities, and other such bodies. I believe you have an interest in benchmarking policies, it is an area of activities both in the UK and in the European Union. I've been involved also on panels doing that. And I've decided to just give you an example of a benchmarking study that was done as a part of one of these spending reviews. I know more about this one, 'cause it is done by my institute with a collaborator. And it was called The Review of Strength and Weaknesses of the UK Science. It's an example in many ways of evaluation in the real policy world, in the sense that the government only decided it wanted this information quite late in the review process, and they gave us only twelve weeks to produce information about the strength, weaknesses, and relative standing of all fields of the UK science. So, what you can do under pressure, I think, is the result here. I'll skip, I should mention maybe from page 24, that the scope of this work was scientific engineering, medical and social science research. It did not include arts and humanities. What did we do when faced with this challenge? The methods we used are summarized on page 25. There are actually five listed here, but maybe we talk three of the important ones. On the right hand side you will see that we used bibliometrics, we did various analyses of citation performance of the UK, compared with certain other leading countries, and in particular compared with the USA. It did not produce very encouraging results, I should say. When we looked at average figures, we found that our top 20% of institutions by citations were normally only just competitive with the average in the USA in over half of the disciplines concerned. And in only one discipline was our top 20% above the US's top 20%. If you went down to individual institutions, you could find more encouraging information for the UK and Europe in some fields. In life sciences, we found that in terms of citations of paper the top 4 institutes in the world came, number one from the UK, and the next 3 were Max-Planck-Institutes in Germany in the years we were looking at. However, all of the rest of the Top 20 came from the USA. And if you looked at the total number of papers from the Top 4 institutes, it was 500 papers per year collectively between the 4 of them. I think, number 4 was probably Harvard, if I remember correctly, they were producing 10000 papers per year at that level of

quality. So, the breadth of high quality in the US was far and away above what we could manage. We could only manage needle peaks of excellence, which could sometimes exceed the USA, but never in scales. So that was the bibliometrics. The top points of the pentagram Review of Reviews, this was a collation of all of these international reviews and other studies which have been done, literature based. And then on the left hand side, we did a self-evaluation. We took the five-star departments, the most highly rated university departments in the research assessments exercise, and we asked the heads of those departments to complete a questionnaire ranking their field against other leading nations. We were fortunate to get to response in excess of 70%, even in this short time, because they recognized the importance of this exercise. A typical graph from that is shown on page 26. You'll see, I forgot, I think this is probably a life science's field here, and this was done in more detail than, disciplines. And in this particular one you will see that we are ranked as generally weaker than the USA, stronger than the other comparator countries. Page 27 is perhaps more interesting. It shows the results of a question which asked them to look at strengths and weaknesses in more detail. And in which sub-field did we have a few outstanding groups? Strength and depth, general weakness, or weakness with islands of excellence. We were very worried about asking scientists which were the areas of national weakness. All our previous exercise had failed on this. National foresight programs had asked panels to identify not only priorities, but negative priorities, things to cut, and the panels just refused to do. In this case, we did get some results, which was quite helpful. Well, it's not very significant to go into which areas came in, it's interesting that when an evaluation does identify an area of weakness, the policy conclusion is not obvious. You can decide, because it's weak you should cut it, or you can decide, it's weak but important that we should maintain and improve it, and therefore you invest more in that area. The next slide on page 28 shows why that could be important. What we found was that, however much we want to have well selected priorities, we also have to recognize the interdependence between fields of science. Our respondents were asked which other fields were critical to the performance of their field. The list on the left shows the most frequently named on our discipline basis, on the right you can see the general dependency pattern.

Engineering departments tended to mention maths and physical science as going very important for them, physical sciences and biosciences both mentioned the importance of engineering, particularly in supporting them for their equipment needs. And in the social sciences, there very often citing each other. Leaving then that study, that study did have quite an influence on the cross-cutting review, instead I'll talk about the effects of the cross-cutting review itself, again to show that the high level evaluations often do shift policy quite strongly. In this particular case a decision was made to increase science spending quite significantly, by 5.4% per year in real terms for a further three years. A major fund was announced for renewing infrastructure, certain fields got specific support, genomic, e-science and basic technology, PhD researchers got increased stipends, that was an identified bottleneck, and the thing higher education innovation fund got the fund for improving industry liaison capabilities in the universities, that was also increased. That content, rather that process. If I can finish then, joining some general conclusions about this practice of high level policy evaluation in the UK. The first thing to say is that science is treated in a similar way to other areas of public spending in terms of target and institutional review. If the government does not differentiate or give it a special character at this top level, the evidence underneath may be a difference, but not the fact of their targets. Nonetheless it is an area where using indicators has proven to be difficult. Because science has become such a central focus of economic policy, and features in almost every treasury policy statement nowadays, it does receive special attention in the spending reviews. Evaluation evidence is used in assessing policy, but is not always used systematically. So from one review to another, they might be a doing different kinds of studies, taking different types of evidence. And as I mentioned, there is a hierarchy of evaluation in the UK, but the evidence is not systematically passed from one level to another. It is certainly known about, but this transfer is more ad hoc. Overall, the UK is a country strongly committed to evaluation. Evaluation is supported centrally from the treasury and from the cabinet office, and as a result of this, I believe that policy is more securely grounded and supported, but evaluation is not a panacea, it is not the only way to support policy making. It is almost always used in combination with other types of

forward-looking and policy analysis. And it is probably stronger being used in this combination than it would be on its own. Thank you for your attention.

ありがとうございます。皆様、おはようございます。着席のまま話させていただきます。

本日私は、プログラムの評価、あるいは機関の評価についてお話しするつもりではなく、政策の評価、研究イノベーション政策の評価についてお話ししたいと思います。具体的な例を取り上げていきまして、システムの評価方法について例を取り上げていきたいと思ひます。フィンランドとイギリスの2カ国を例に取り上げて説明いたします。

スライド1をご覧ください。こちらでは、国レベル、高レベルで評価する目的を掲げております。多くの国々が、現在ではイノベーション・システムの視点から研究システムのデザインをしております。これはそもそもOECDの発想、あるいは学会での発想なのですが、研究を実施する機関やその連携に焦点をおくものです。これを見ると、システムを見て、そうすると学習のギャップはどこにあるか、ボトルネックはどこにあるか、機会がどこにあるかということを見ることができます。このアプローチの長所の一つは、研究政策及びイノベーション政策は相互作用を及ぼして、お互いに奨励したりシナジーを図ったり、お互いに阻害要因になったりするということがあり得るということです。

2ページ目に図があります。これは、私が議長を務めておりましたEU委員会の諮問パネルで使ったチャートで、これは公共政策を活用して、欧州におけるR&D費用をGDPの3%まで増やす方法を提案した図であります。詳細は割愛いたしますが、非常に多くの政策上の手段が活用できるということです。左側にサプライサイドがあります。基礎研究が左側にありまして、業界へのサポートもあります。そして需要サイドが右側にありまして、イノベーションの需要が高まってリサーチの需要が高まる。そして、下の方ですが、フレームワーク・コンディションというのがあります。つまり、政府政策が一般的にインパクトを及ぼす人材、知財、そして競争、市場規制に影響を及ぼすということを示しております。ですから、必ずやコーディネーションが必要でありまして、政策間のコーディネーションがないと進歩を遂げられないということです。例えば、R&Dを増加させるためには、研究者を提供するために他の分野から連れてくる、あるいは研究開発者の訓練をするというような人材育成プログラムがない限り、R&Dを増やすことができないというような関連性です。

さて、次に具体的なケースを挙げたいと思ひます。フィンランドのシステム、ここ数年間で私は2件かかわりました。

まず1件目ですが、4ページから説明があります。これは緊急の追加予算を評価する方法でありまして、この評価の目的のそもそもの発端は、フィンランド政府が96年に行った決定に起因しております。その際に研究開発費をその

後2年間で25%増加させるという決定を下しました。かなり果敢な決断だったのです。と申しますのは、この時期、フィンランド経済は、元ソ連邦の市場を失って歴史上最悪の不況期にあったからです。その際にパネルを設置して、これらの資金活用の評価をしてもらうということにしました。そして、経済にどのような影響を及ぼすかということの評価してもらうことにしました。外人が2人名を連れまして、1人はドイツ人、そしてもう一人が私で、このように携わったことから皆様に説明しているところであります。

5ページにおきまして、主たるパネルの課題を説明いたしました。箱になっておりまして、上は経時的なリサーチ、そして下の2つの箱は経済の経時的な変化です。我々の提起した問題は点線の矢印で示しておりまして、新たな研究開発費が、現在は経済にどのような影響を及ぼしているかということです。ただ、この質問は、リサーチの遅効効果があるために、即時その時点では経済へのインパクトを判断できないわけです。そこで我々がやらなくてはならなかったことは、代替的な尺度を活用するということでした。90年代前半に行われた研究が経済にどのような影響を及ぼしたかということ調べました。斜めの矢印が示しております。また、加えて、追加予算によって研究開発システムが変わったか、あるいは経済がその時期変わったかということを示しております。それが我々の概念的な枠組みでした。これはパネルでやるというだけでなく、スタディーをサポートするためのかなりの予算を配分されました。そして専門家のチームを設置しまして、我々が使うための証拠集めをしてもらいました。文献の統計的分析、一連の計量経済学の研究も集めてもらい、そして、フィンランドにおける過去の制度またはプログラムの評価のデータを集めてもらいました。そのリストが6ページにあります。

7ページは、我々自身が行ったパネルによる活動ですが、2年間でいろいろな面談を行いました。ノキアや大学研究機関のトップと面談しました。こちらに報告書の要綱があります。今日は結果については話さないで、そのやり方についてしかお話いたしません。8ページはその結果の一つであります。我々の意見としては、資源が過度に費やされている、科学技術の中でもハイテク業界指向の分野に偏っている。つまり、ニューエコノミーに偏っているということを見出しました。分析の結果、この分野は、フィンランドの競争力を維持するためにはニューエコノミーだけでは不十分であって、伝統業界（オールドエコノミー）、木材や森林業などにも今後中期にわたっては依存していかなくてはならないということだったので、オールドエコノミーの関連サイエンスに関する研究開発も行っていないといけないという結論が出ました。これ以外にもいろいろな結論が出たのですが、詳細は割愛いたします。

簡単に9ページに評価のインパクトをお示しして列挙しております。閣僚2名に提供しまして、また、首相にも最終答申のコピーを渡しました。かなり報道されましたし、政府はさらにこのプロジェクトを2年続けるための追加予算

を出しました。また、研究機関もこういった勧告にのっとなって、そのやり方を変えました。

そして、10ページからはフィンランドの第2事例であります。これはごく最近完了したばかりのケースでありまして、始まったのが2002年6月で、通産省が新たに評価が必要だと決定を下しました。今回はフィンランドのイノベーション・システムに焦点を置くというところで、各官庁、そして各ネットワークのバランスを見るということでした。今回はパネル員としては4名のみ、2名の外国人と2名のフィンランド人、ステークホルダーはパネルの委員として参加するのではなくて、コンサルティングパネル、あるいはサポートチームを構成しました。その人たちに我々は質問するという形をとったのです。

11ページに、基本的かつ理論的な大きな問題をリストアップしております。なぜ政府がイノベーションをサポートするために介入するべきか、その正当化できる要因はあるか、市場の失敗やシステムの失敗があったか。そして、それを避けるために政策決定者にでき得ることがあるか、どんな方針が必要か、政策が必要かということです。

幾つか、そのレポートの結果、結論の図式を入れ込んでおります。12ページに1980年代からの変化を示してありまして、下にあるとおり、基礎研究、応用研究、製品開発、商用化とあります。縦軸にどのようなサポートがあるか、無償、グラント、融資、エクイティー、サービス。当時はかなり直線的なモデルで、サポートは下流に偏ってございました。つまり、製品開発と商用化に偏ったサポートが提供されておりました。ところが、それからさらに時が進みまして、13ページに現在のシステムがあります。大きな変化が起きました。ほとんど全サポートエージェンシーがマנדートを拡大いたしまして、川上に集中するようになったのです。Tekesという官庁は、産業研究、また産学連携研究を奨励している官庁ですが、その重点項目を基礎研究側に移しております。今、フィンランドアカデミーという資金源と協力しておりますベンチャー・キャピタル系のエージェンシーも、シード・コア・ファンディング、いわゆる「死の谷」に移行してありまして、マーケット的な投資から離れています。また、スタートアップ企業も複数のエージェンシーから同時にサポートを受けているということがありまして、順番というのは特にはない。

14ページです。さらなる分析ですが、どういった活動に対してどういうところがファンディングを提供しているかということを示してありまして、官庁間の重複を見ております。重複があるということはマイナスがあるかもしれない。無駄な重複があるかもしれないということを示している可能性もありますし、さらに、エージェンシーが協力できる。Tekesとアカデミー・オブ・フィンランドの重複は非常に生産的でありまして、これがわかって共同プログラムを開催することによって、産学連携をさらに図ろうということにつながりました。

15 ページに方法論について意見を示しております。これは特異な評価でした。というのは、基本的な政府の介入の動機づけから実践上のことまで、理論から実践まで見るということをしたのであります。ルーチンな教科書上の質問だけでは不十分だということで、そういうように範囲を広げました。フィンランドは、こういった調査にとっては非常に適切な場所でありまして、あらゆる企業、あらゆる研究開発機関のデータが潤沢にあります。よって、他のところではできないような作業がフィンランドではできます。

もう何点かコメントが続きます。世界競争力市場ということになりますと、大体フィンランドは上の方に来ます。それにもかかわらず、政府首脳としては評価を行い、さらに立場を強くしたいという希望がありました。

最後の点ですが、全体の研究システムを見ただけでは十分ではないことがあります。よく苦労したのが、境界条件と私はここで書いているのですが、現実には、研究・イノベーション政策を他の分野と分けて考えることは難しいものです。例えば教育、財政政策、あるいは調達などと区分けをして評価することは難しいというのが現実の悩みでした。

ここからイギリスの例です。少し様子が違ってきます。まず、イギリスにおけます評価の一般的な枠組みで、政策レベルでの評価を後で紹介します。

まず18 ページです。労働党政権が97年に発足して以降、パブリック・サービス・アグリーメントという制度が導入されました。公共支出はすべてターゲットに基づいた制度で管理されることになります。各アグリーメントが各省庁をカバーしておりまして、その省庁としての目標、目的が定められており、ねらいが定められており、幾つかパフォーマンスに関する目標が述べられております。また、バリュー・フォー・マネー、すなわち効率をはかるような指標が含まれており、だれが成果物に関して責任を負っているのかが述べられております。これは米国のG P R Aと若干似ています。しかし、これは議会ベースではなく、政府が定めたシステムであるところが異なっています。政府自身が目標を設定し、政府がそれに基づいて行動をとることが期待されています。政府全体として160のターゲットが設定され、さらに詳細レベルでの内容がサービス・デリバリー・アグリーメントに述べられています。具体的にご説明をしていきます。

そこにいく前に、概観として幾つか異なる評価方法がイギリスであるので、ご紹介します。異なるレベルで評価があります。それでヒエラルキーと呼んでおります。19 ページです。なぜ階層的だということかといいますと、高いレベルから低いレベルまで評価があるためですが、上から下の間までうまくコネクトはできていないという問題があります。一番高いレベルにおきましては省庁レベルでの評価があります。先ほど触れましたパブリック・サービス・アグリーメントです。支出評価（スペンディング・レビュー）について後でご紹介します。政府が予算を2年毎に向こう3年に関して策定します。これはローリング

ベースで定めているのですが、単年度制度に問題がありますので、それを乗り越えるべく3年ごとに見ております。

エージェンシーレベル、庁レベルに関しましては、我々はクインケニアル、これは5年に1度という意味ですが、5年ごとのレビューを行っております。これは、その庁がなぜ存在しているのか、あるいはその使命にどれだけ根拠があるかを確認する意味があります。

その次にインスティテューション、それぞれの機関レベルがあるのですが、よく知られていますR A E（リサーチ・アセスメント・エクササイズ）があります。これは大学におけます研究評価で、よく知られているかと思えますし、皆さんが事前に受け取っていらっしゃるレポートにも入っています。また、例えばリサーチ・カウンシル・インスティテュート、公的な研究機関で基礎戦略研究が行われているところに関しまして、評価がかなり厳格に行われています。

フィールドレベルでの評価も行われています。これは今では広がりつつあります。国際委員会が設置され、ある全体の分野に関する健全性の評価が行われます。例えば物理、化学、コンピューター科学に関するこのような調査が行われました。Q P I Eはもう一つの内部のシステムなのですが、今日は申し上げます。

次のレベルがプログラムの評価です。これが一番様々な活動が行われている分野で、今日の私の話ではないのですが、何か質問があれば、どうぞ後で私にください。それに加えてテーマ毎、あるいはテーマ横断的な評価も現在進行中です。

ケーススタディーを1枚にまとめてあります。20ページです。これはD T I（貿易産業省）のP S Aターゲットです。このD T I（貿易産業省）が、イギリスにおけます科学費の責任を持つことになりまして、リサーチ・カウンシルの予算もD T Iが管轄しています。6つのターゲットがありまして、うち2つがサイエンス関連になっていまして、ここに書いてありますP S Aの5と6、これがサイエンス絡みのものです。1つは国際比較で、イギリス全体として科学・工学能力を高めるということです。国際的に、例えば質、コスト効果比、関連性を評価して決めます。もう一つ、これが技術移転を見ます。どのように技術的な知識を十分に活用しているか、科学・工学力をどのように生かしているかを見ているのですが、これは、革新的なビジネスがこのような能力を引用していることが増えているかで示されます。

学習のプロセスも経ておりまして、初めてターゲットを科学について設けた段階においては、これはスピノフの会社を増やせばいいという考えがありました。50%スピノフ会社を増やそうという指標が出たのですが、これは余り意味がなく、簡単に数字合わせができてしまうということがわかりました。実質的なパフォーマンス改善がなくても数字合わせができるということが見え

てきたので、もっと中身のある指標を使うようにしております。具体的なサブターゲットもつくっております。

次の21ページ、第2のケースで、これは政策レベルでの評価です。これは支出評価、スペンディング・レビューという言葉で最初に申し上げましたが、この中で行われておりまして、科学の横断的な評価です。2回やっておりますが、すべてのスペンディング・レビューで実施するという事は決まっています。しかし、政府政策の中でイギリスは科学をととても重視しております。経済政策の中心に科学が置かれていますので、当面は続くのではないかと考えられます。これは高いレベルでの評価で、大蔵省に当たるトレジャリーが管轄をとっています。しかし、これはDTIのもとにある科学革新省が主導をしています。この支出評価を行うことによりまして、国としてどういう優先順位を持ってお金を支出するかを決定しているのですが、この評価を行って科学方面は重要視されています。ただ、系統立った方法であるわけではありません。どういう方法で評価をするか、まず文献、あるいは証拠物を調べます。委託して調査委員会が設けられることもあります。ただ、時と場合によって異なります。どういう人がかかわるか、この人たちは主たるステークホルダー、例えば実際に研究をしている大学、あるいは産業界の代表者などと協議を行います。例えば医療の慈善団体、あるいはその他の関連機関との協議も広く行っています。

政策のベンチマークも、皆さん、ご関心がおありではないかと思えます。イギリス、EU域内でも広くありますし、私自身がそういうパネルにかかわったことがあります。ベンチマークのスタディーがこれらの支出評価の一環として行われていたのでご紹介します。私がいるところで、他の機関と協力をして行ったことをご紹介します。SWOT、イギリスの科学の強み、弱み分析です。これは実際の政策の現場での評価の例と言えますが、というのは政府は、評価かなり後の段階になってこのような情報が必要であると言い出して、12週間だけで強み、弱み、あるいは全体的なイギリスのあらゆる分野における位置づけを調べるよう我々に言いました。そのような状況で、このような成果を出しました。

24ページです。この調査の規模ですが、科学、工学、医学、社会科学をカバーしているもので、人文は含まれておりません。

何を具体的にしたかということですが、どういう方法を使ったか、その手法が25ページにまとめられています。ここに5項目書いてあります。上にある3つが重要だと言えるかもしれません。これは出版物に関する数字も見ていますし、例えばどれだけイギリスの文献が引用されているか、他の主要国と比較しています。とりわけアメリカと引用件数の比較をしています。この結果、我が国にとっては余りいい結果ではありませんでした。平均値で見ますと、引用数でトップ20%の研究機関がアメリカの平均的な機関と互角という状態が、半分以上の分野でみることができました。イギリスのトップ20がアメリカの

トップ20%を超えていました。

個別の機関を見ていった場合には、より高い水準のところもあります。イギリス、ヨーロッパで、特定の分野についてはヨーロッパ勢が強いというところがあります。生命科学では論文当たりの引用ですが、全世界のトップ4機関は、第1位がイギリスで、次にドイツのマックスプランク協会の3機関が続きました。対象期間においてはそういう順位でした。しかし、トップ20の残りの機関は、すべてアメリカでした。論文の数全体がこのトップ4機関からどれだけ出ていくかという、あわせて年間500の論文でした。第4位はもしかしたらハーバードではなかったかと思いますが、年間1万の論文をそれだけ高いレベルで産出していたんです。それだけの厚みがある。アメリカの厚みは、はるかにイギリスを超えていたのです。我々は、ある特定の分野でエクセレンスを達成して、そこだけはアメリカを超えていたかもしれないのですが、厚みとか広さは全然追随しておりませんでした。

それがその文献の統計で、一番上のところ、レビュー・オブ・レビューと書いてあります。国際的な評価、それ以外の調査をまとめたものです。文献ベース、レビューのレビュー、左側のところですが、これは自己評価です。一番レーティングが高かった大学部門、これをファイブスターとっておりますが、ここを調査いたしまして、この部門長の方にアンケート調査をお願いしまして、他の主要国と比較をしてみました。幸い70%を超える回答率が短期間にもかかわらず得られました。問われた側としてもこれが重要だとわかってくださったためです。

その結果のグラフが26ページに載っています。多分生命科学の分野だったと思います。詳細な調査も行われているのですが、ご覧いただけますように、アメリカに比べれば全般的に弱い。しかし、それ以外の国に比べれば少し上回っているというのがイギリスの位置づけです。

27ページの方がおもしろいかもしれません。これは質問をしまして、もっと詳しくどこが強い、弱いか確認しています。どういう分野で強い。まず一番強いところ、あるいはかなり強みがあるところ、あるいは全般的に弱いところ、あるいは一部については優良であるところ。科学者に対しまして、国全体として弱いのはどこかと聞いてみたのですが、これまではそういう質問はすることができませんでした。フォーサイトのパネルに対しまして優先順位を聞いただけではなく、優先順位が低くカットする分野のことも聞いてみたところ、それを答えていただくことはできなかったという経験が以前はあったのですが、今回はこういう聞きにくいことも聞いたところ、きちんと回答が出ました。だからといいまして、実際の分野を挙げていく意味はここでは余りないかと思うのですが、弱いところが出てきた場合に、政策上どうするかははっきりしないものなのです。弱いからやめるべきなのか、弱いから重要なので、もっとお金を注ぎ込むということをするのか。必ずしも問題がある、弱いからといって結

果が決まるわけではありません。

28ページ、何故こういう調査が重要なのかです。できるだけ優先順位を決めたいと思いつつも、各分野間にはつながりがあるということを忘れてはいけません。回答者に対して、他にどういう分野が自分たちの分野にとって重要であるかと思うかと聞いております。左側に書いてあるのが最もよく出てくる学問分野で、右側が一般的にどのような相互依存関係があるかです。例えば工学系であれば数学、自然科学が関連していて、自分たちの分野にとって重要であると答えていますし、物理学、バイオサイエンス、両方とも工学が関連性が高い。とりわけ機器装置のニーズ上、工学は大事だと。社会科学の分野においては、お互いにお互いの分野が重要だと言いつつパターンが多くありました。

その調査を行った結果、部門横断的なレビューにはそれなりの影響が出ました。その横断的なレビューを行った効果です。高いレベルで評価を行った場合に政策が変わることがよくあるということを示します。今回の場合には、決定が行われて、科学に対する支出をかなり増やそうということになりました。年間実質5.4%ずつ、向こう3年間増大させていこうということになりました。しかもインフラ更改の基金が設けられました。ゲノム、電子科学、あるいは基礎、こちらについてはお金を増やすことに決まりました。**PhD**研究者に関しての手当、これがボトルネックとなっているということがわかりましたので、もっとお金を出すということになりました。また、高等教育、イノベーションファンドとしては、これは産学を促進していくための基金ですが、こちらでも増額が決まりました。これはプロセスではなく中身ということになります。

結論をまとめます。イギリスでの高いレベルの政策評価についての結論です。まず科学の取り扱いですが、ターゲットを設け、制度的に評価をするという意味では他の公共政策分野の支出と全く同じ扱いを受けております。科学と他の部門を区別しているわけではなく、この高いレベルにおいては科学は特別扱いをしていません。下のレベルでの根拠資料については違うかもしれませんが、ターゲットがあるということについては同じです。とはいいながら、指標を用いるということは難しいということもわかっています。科学は今、経済政策でも中核的な位置を持っていて、ありとあらゆる財務関係の政策ともかかわりがありますので、この支出評価（スペンディング・レビュー）の中では特別注目されています。政策を評価する中で根拠証拠は用いられているのですが、系統立った形で用いられているわけではありません。異なる調査を行って、異なるタイプの証拠物を比較する場合があります。先ほど言いましたが、評価といっても、上から下まで異なるレベルで複数行われていますが、違うレベルから別のレベルにきちんと系統立って移転がされていません。この情報移転に関しましてはケース・バイ・ケース、単発的で系統立った形ではありません。

全体的に見ますと、イギリスとしては政策の評価をきちんとするという事にコミットしています。これは大蔵省、そして内閣府でも必要であると言って

おりまして、その結果、今、イギリスの政策立案は評価による結果を反映して、具体的な根拠を持つようになってはいますが、これですべて問題がなくなったわけではありません。評価の結果のみですべてが決まっているわけではありません。評価プラス、それ以外のタイプの先を見通した政策分析もあわせて最終決定を行っています。多分、組み合わせで評価をした方が効果的であると思っております。

ありがとうございました。

【大山会長】

ありがとうございました。

それでは、30分ほどディスカッションの時間をとりたいと思います。ただいまの発表に対しましてご質問、あるいはご意見等がございましたらお願いいたします。

黒川議員、どうぞ。

【黒川議員】

Let's say about the UK, UK has been very well known for excellencies in science over last 100 years or so, but as you recently seen, this is a very political decision, and Tony Blair decided your College Fees Policy, English higher education institutions have adapted many foreign students, which may pay in higher fees. But how do you see the sort of brain drain? Also more investment in terms of higher education would benefit, not necessary benefit science, but I think the overall science policy in the UK, which must be attached policy in education to?

イギリスは、サイエンスにおける優秀さに関しては、過去1世紀非常に高い評判を得てきたわけです。でも、ブレア首相のもとでカレッジ・フィー・ポリシーということで、外国人留学生を誘致する大学授業料についての方針を決定いたしました。政策転向であります。全般的な科学政策は、こういった教育分野における政策変更の影響を受けまじょうが、それについてどのようにお考えですか。

【ジョージュー氏】

Yes, I've been talking about funding for research, not funding for the rest of the university, but it's the same. The health of the universities is dependent upon funding of the whole, and as with many countries we've been asked to take more and more students, but without equivalent increasing in funding. The unit of funding per student has

gone down drastically over time. And it is essentially now uneconomic to teach undergraduates from our own country. Foreign students pay agent real-market fees, so we have an incentive to take more of them, and fewer students from our own country. And this is not healthy for either the universities, or for the country. It is not even beneficial for the foreign students, we have some courses where more than 50% of the students come from China, for example. And we are not giving them a proper UK experience in this kind of situation. So universities very strongly support the introduction of fees for which there will be effectively very low-interest or interest-free loans, only recoverable when they are earning above-average salaries in their future careers. The brain drain issue is not considered to be a great problem in the UK, it is probably a problem at the peak, rather than a problem at the average. I think, some of our very best people are attracted by the high resources in the USA. We, in turn though are probably draining the brain of other countries, and particularly of developing countries. As an English speaking country which is also multi-cultural and therefore easy for foreigners to live in it is quite easy for us to attract people in.

研究のファンディングについてお話しして、大学の他の予算についてお話ししていないのですが、大学の財政はファンディング全体の影響を受けます。多くの国と同様に、より多くの学生を受け入れて欲しいということが求められています。それによってコストが上がるようでは困るということで、1学生当たりのコストは経時的に大きく減ってきました。今、学士を我が国で教育するに当たっては、外国人留学生も実際の市場に即した授業料を払いますので、国内の学生よりも海外の留学生の方が収益が上がるわけです。でも、これは健全な状況ではなく、我が国にとってもプラスにならないし、外国人留学生にとってもプラスにならない。例えば、学校によっては5割以上の学生が中国の学生というところもありますので、実質上英国での留学経験を与えていないわけです。ですので、大学は授業料の導入をサポートしておりまして、低い金利か実質無金利の融資で、将来に職に就いたときに大学卒業のキャリアで、平均以上の給料で回収されるような学生融資制度が受け入れられております。

英国では頭脳流出は重要な問題とは考えられておりません。この問題は平均の問題ではなく、ピークの問題です。最も優れた人たちが資金の豊富なアメリカに出ていってしまうというのが最も大きな問題ではないでしょうか。ただ、我が国も、特に途上国から頭脳流出を受けているということで、英語圏だと、特に我々も母国語が英語ですので人も来やすいということで、他の国からの流出というのを受けています。

【黒川議員】

Also about the Finland, Finland has been sort or analyzed by the IMD and Finlanders parameters as one of the leaders as of strong economy, and with a potential of top five particularly strong in Science and Technology. But I think Finland is a very small country with only 5 million people, so they have very close ties within Nordic countries. And so people may be moving around. And what fraction of research and higher education students are from foreign countries, within Nordic country, and even outside of Nordic country?

フィンランドもIMD、あるいはそれ以外のところが評価をして、経済を刺激するために科学技術政策が活用されているという評価が出ていると思います。私も訪問したことがあります。フィンランドは人口500万のとても小さな国なので、国の中でのつながりも緊密かと思います。高等教育研究部門における北欧における留学生、それから北欧外からの留学生はどのような比率かご存じですか。

【ジョージュー氏】

Hm. I can't tell you the answer numerically, but it's very low. It may be, Finnish students go to other countries, but very few people come from other countries to Finland and back. This was one of the conclusions of our evaluation. We said, it is a weakness in the Finnish economy, understanding of foreign cultures, eventually effect their ability to innovate for global market.

数字は承知していませんが、少ないと思います。つまりフィンランド人の学生が外に行くということはあると思うのですが、留学生の受け入れは少ないという結論が私たちの評価からも出ています。それが弱点であると、フィンランド経済の弱点の一つとして、外国文化の理解が低いということであれば、グローバル・マーケットに向けてイノベーション力が足りないことにつながるという結論も出ています。

【大山会長】

他の委員の方、如何でしょうか。
末松委員、どうぞ。

【末松委員】

Regarding to the question debated by Dr. Kurokawa. You have mentioned about encouragement of PhD student in the research area

in the UK. I think, one more thing encouraging the PhD student is how the society welcomes graduated PhD. For example, increasing salaries of the industry or something like that. And do you have any particular idea, please let us know about that.

黒川先生の質問に関連する質問ですが、**PhD**の学生を研究分野で増やそうとしているということをおっしゃいましたが、もう一つ、**PhD**を取得する生徒を増やすためには、社会が**PhD**の人たちをどうやって受け入れるかということの策をとらなくてはならないと思います。例えば企業において**PhD**取得者の給料を高めるといったような策が必要だと思います。そういった策について何かアイデアがあったら教えてください。

【ジョージュー氏】

I wish I did, but I think I am like many people here an example of those whose lifetime earnings are depressed by having a PhD, rather than increased. So if you subtract the years when you are not earning while you are training for you PhD from your lifetime earning, you certainly end up on average earning less from having one. Some of our industries strongly value a PhD, including our strongest industries, our pharmaceuticals. In others, there is still a lot of education to be done, clearly in the engineering industry. Gradually, I think, we are moving to a system where our industry is expecting people with a higher level of research training.

わかっていたらいいなと思うのですが、皆様のご経験もそうだと思いますが、私などは**PhD**によって生涯賃金が上がるどころか下がった例の一人ではないかと思っています。一生の稼ぐ額から**PhD**を取得するための期間に稼げなかったのを引きますと、**PhD**のある人となない人だと、ないの方が稼げるというのがでております。ただ、業界によっては、製薬業界が**PhD**をかなり受け入れているという姿勢を示しておりますが、まだまだ業界に対する教育が必要です。エンジニアリングの分野でも必要ですが、少しずつ産業界が、より高い学位、研究のトレーニングを受けた人を求めるような兆候が、徐々に始まっていると思います。

【大山会長】

他の委員の方、如何でしょうか。

国武委員、どうぞ。

【国武委員】

Short questions on page nine, you mentioned the impact of

evaluation the Finnish case, and you mentioned it in relation it is this enlarging programs on social based innovation. What is social based innovation? What is social based innovation? Is not technologically based innovation?

9 ページ、評価をした場合のインパクトがフィンランドに関して述べられています。科学ベースのイノベーション、これは技術ベースのイノベーションと違うのでしょうか。どういうものなのでしょうか。

【ジョージュー氏】

It's not necessarily technology based, but in this case the problem we identified for Finland was linking in a innovation system to social problems. If I can give an example, one of the major problems that Finland has is depopulation of rural areas. The population is moving very quickly into the Helsinki region, and they wish to keep the rest, to maintain the population in the rest of the country. It links to the aging society, as mainly old people stays in the rural regions, and the cost of delivering health services becomes more and more expensive. What they are now trying to do is, turn research and innovation towards this problem. And that includes, for example, strong programs in developing health telematics, so that you can delivery health services more cheaply and efficiently to a rural region—that's a socially based innovation.

必ずしも技術ベースではありません。我々が見たところ、フィンランドにどのような問題があるかといえば、イノベーションに関するシステムと社会問題との間のリンクが十分にできていないということでした。どういう意味かといいますと、フィンランドとしての問題ですが、地方で人口が減っています。かなり急速に人口がヘルシンキに集中しつつあります。国としては人口分布を維持したいということで、高齢化とも問題が絡んでいます。高齢の人のみが地方に残っているので、社会保障のコストが上がりつつあります。そこで研究とイノベーションによって社会問題に対応しようとしているわけで、例えば強力なプログラムを使って電子診断のようなシステムを活用して、地方においても、より安価で、効果的に健康保険分野でのサービスを打ち出そうということです。

【国武委員】

That is including educational programs to the elder people?

教育プログラムで、高齢者向けの生涯教育などもかかわってくるのでしょうか。

【ジョージュー氏】

Yes, it would include that as well.

はい、それも一部となります。

【国武委員】

They have to become accustomed to the high-tech things.

そうすると、高齢者もハイテクに馴染まないといけませんね。

【ジョージュー氏】

Yes, certainly.

はい、もちろんです。

【国武委員】

And another related question is, do you have a similar innovation program in the UK?

関連して、同じようなイノベーションプログラムは英国でもあるのですか。

【ジョージュー氏】

We don't. I mean, no, there is work on, you mean on socially based innovation more generally. We do have some work on health telematics, but of population structure is very different from Finland. I think we need, it is a growing trend. So it's a growing trend in Europe generally, the idea that you should propose innovation, rather than technology, which means that you should support the combination of technology with services, rather than technology alone.

もし一般的に社会をベースにした革新ということであれば、健康方面での電子サービスということはあるのですが、人口構造がフィンランドとイギリスではかなり違います。傾向としてはヨーロッパ全般で同じように見られていまして、技術ではなくイノベーションが大事であるということは、技術プラスサービスの組み合わせが必要であると、テクノロジーだけでは不十分という考え方が強くなりつつあります。

【黒川議員】

I think each country has a unique history and social background. Now the transition to the so-called globalization and with the mobility of our people across national borders, this is creating problems. Now Finland obviously has a very unique also history, having been independent for maybe 200 years. But I think; now recent years,

Finland is very well known as the most transparent country in the world, very democratic and transparent. So the people and society have a high level of trust in the government, and also equality between the sexes has been very well promoted. And if we look at this panel here around this room, in Finland you'd expect maybe 50-60% would be women, isn't it? And Nokia has been very strong. But the UK has a different history again. Very small government in common law has, I think, promoted excellence in science with very well known institutions of higher education. Now, what do you see Japan's issue from the viewpoint of the UK, your experience in UK and Finland?

各国の歴史も社会的背景も違います。ただ、グローバル化が進んで、人の移動が国境を越えて起こることになりまして、問題も出てきております。明らかにフィンランドにおきましても独特の歴史がありまして、約200年間独立した国として歩んできたのですが、最近では世界で非常に透明性の高い民主主義国として知られるようになりました。ですから、フィンランド政府に対する国民の信頼は高い。そして女性の機会均等が奨励されております。このパネルを見回してみると、フィンランドであったとしたならば、50%以上が女性で占められていたでしょう。ノキアも非常に顕著。でも、イギリスは違います。コモンローで小さな政府で、サイエンスにおけるエクセレンスが促進されておまして、高等教育機関も素晴らしいものがあります。そういったイギリスとフィンランドの経験に照らし合わせて、今、日本の課題は何だとお考えですか。

【ジョージュー氏】

It's a very big question. Well, maybe I'll stick only to evaluations, since that's the purpose of our meeting today. And I'm not qualified to comment on society as a whole. I mentioned this already yesterday, I've been visiting Japan almost every year since 1982, so I've had some chance on short visits at least, to see some of the changes going on here. And maybe even 10 years ago, I would have said that evaluation in the way we are talking about it today could not very easily be implemented in Japan, because you did not have the right kind of critical culture. Now, what I mean by this, I think you had a critical culture within organizations, within a department, within a company or something, people would be very critical of each other, and try to make quality improvements and so on. There was a strong inhibition to make public criticisms, and criticize between organizations. And if you do not have that, you cannot have evaluation. Because evaluation has to be open, people have to be able to see it, to take criticism, and

to learn from it. Now, I think that has really changed, especially in the last 5 years. I now see much more debate and self-criticism in this country. And because that cultural condition has changed I think it is much more likely that you can institute evaluation. Even so, in, well, all policy transfer between countries is difficult, it often fails because the policy is not sufficiently adapted to local conditions. Somebody picks it up and tries to put somebody else's system in, without understanding why that system is there. And I think, that's also true for evaluation. In any country it must be designed to be compatible with the, well, we are talking about public policy with the government and agency system of that country.

非常に大きな質問です。私のこのミーティングの目的は評価にありまして、社会全体についてコメントする資格もありませんが、これは昨日申し上げましたが、私は1982年から日本には1年に1回は来ております。ですから、滞在は短期かもしれませんが、日本で起きている変化について少しは見る事ができております。恐らく10年前でしたら、我々がここで議論している形での評価は簡単に日本では実施できないであろうと申し添えておりました。なぜならば、批判の目が十分に日本にはなかったと思います。そういった風土、文化がなかったと思います。組織内におきましては批判をするという文化があったのだと思います。企業内、あるいは部署内でお互いに批判をして品質の改善を来そうとするような努力はしていたが、公の場で非難する、批判するという事を組織を超えてやるということはない文化だったと思います。ただ、それだと評価はできない。評価は、オープンになって批判を受け入れて、そこから学習しようという姿勢がなくてはならないのです。ところが、ここ5年で本当に変わってきました。昨今におきましては、自己批判やディベートが増えてきております。その文化的条件が変わってきましたので、可能性としては、日本におきましても評価を制度化できる可能性が高まってきたと思います。

ただ、国から国へ政策を移転するのはあらゆる面で難しく、しばしば失敗します。その原因は、現地の状況に順応する形で政策を導入しない、別の国の政策を十分に理解しないでそのまま入れようとするということがあります。評価に関しても同じことが言えます、どの国であっても、その国の状況に則した形で設計し直して導入しなくてはならない。政府の制度や官庁の制度に合わせなくてはならないということが言えます。

【大山会長】

畚野委員、どうぞ。

【畚野委員】

In the UK you have many levels of high range evaluation systems, and you mentioned they are not done systematically. I'm afraid they spend their time for responding to the evaluation not to have the time to do research?

イギリスにおいては評価が複数のレベルで行われる、しかも、系統だった形で情報が伝達されていない、とおっしゃったのですが、多過ぎませんか。研究者は十分に時間がとれなくなるのではないですか。こういったことに縛られて辟易したりしていないですか。

【ジョージュー氏】

So you mean the cost issue? Cost evaluation, it's a good point, that we are in some areas, we could be sensibly suffering from evaluation fatigue, too much. This is particularly true at the European Commission programs, where in my identity as a researcher I get research grants from the European Commission. And in the life of a research grants, and afterwards, you could expect to have maybe 5 different studies going on, where you expected to fill in a questionnaire about the impact of that work. And this is very important to systematize the collection of innovations, so the collection of information, so that researchers only have to give back data once, and then it is used for all subsequent evaluation studies. And the best evaluation system collect that basic data automatically, and leave the evaluation panels to concentrate on higher level issues.

評価のコストがかかり過ぎる、あるいは負担が大き過ぎるということですか。大事な点だと思います。一部の分野では、確かに評価疲労といいますか、その負担はあるかもしれません。とりわけ欧州委員会のプログラムでその問題があります。私自身は研究者として研究助成を欧州委員会から受け取っております。それを受け取った後なのですが、多分5回ほど異なる調査があり、アンケートを求められたりします。どういうインパクトがあったかなどと、かなりいろいろ問われたりします。イノベーション全体を系統立てて活用するというのは極めて重要であり、情報も系統立てるということが難しいものです。研究者としては、1回だけデータを出せば、それ以降必要な評価には共有ができるようにできればいいのだと思うのですが、その基礎データについては自動的に集めることができ、あとは評価パネルがより高いレベルの問題に集中することができればいいとは思いますが。

【大山委員】

他の委員の方、よろしいでしょうか。

OK. Doctor George, thank you very much for your presentation.

ルーク・ジョージさん、プレゼンテーションありがとうございました。

これで海外研究開発機関からのヒアリングを終了させていただきます。

続きまして、評価における問題点、課題等に関する自由討議に移らせていただきます。

これまでの専門調査会において提起された主要課題及び先般各委員からご提出いただいたコメントを踏まえまして、お手元の資料4のとおり主要課題等に関する整理をいたしましたので、事務局から簡単に説明をいただきます。

<事務局から、資料4「これまでに提起された主要課題」について説明が行われた。>

【大山会長】

この評価に関する問題点、課題等に関しまして、各委員の先生方にいただいたご意見がこういった形で、次回の中間取りまとめに向けて整理が進行してございます。本件につきましてご意見等がございましたらお願いいたします。

なお、本日は大分時間が経っておりまして、時間の制約がありますので、十分な議論ができないと思いますので、後ほどご案内申し上げますが、本日議論が尽くせなかった点につきましては、ご意見等を別途書面でご提示をいただきたいと思います。ご協力をよろしく申し上げます。ご意見を伺います。

畚野委員、どうぞ。

【畚野委員】

これは感想ですが、今まで特に研究の評価というと、我々日本の場合、個々のプロジェクトのパフォーマンスの評価に重点が行きがちだったのですが、今日の海外の方のヒアリングが象徴するように、段々と上の方のレベルのプログラムから政策、そういうところの評価というのが重要だというのが認識されてきたので、非常にいいことだと思います。

【大山会長】

他の委員の方、如何でしょうか。

末松委員、どうぞ。

【末松委員】

前回出席できなかったものですから、あるいは少し場違いな発言になるかもしれませんが、本日の議論を伺っております、あるいは本日の資料の海外の状況というようなことを拝見をいたしましても、やはり海外におきましては基礎的・基盤的、あるいは萌芽的な研究というものの重要性と、それから、政

策目標の遂行というものがかなりはっきり分けられて議論されているというように思うわけです。この2章以降というのは、かなり具体論に入りますが、そういうかなり性格の違った研究カテゴリーがあるということで、それぞれに対しては違ったスタンスで評価しようというようなことが、どこか1章の辺にはっきり書かれていた方がいいのではないかと。これは後ろの方の6ページ等にそういうことが少し書いてありますが、それよりもう少し前に書き込んでおいた方が、具体論に全部関係いたしますので整理しやすいというのか、わかりやすいのではないかとというような感想を持ちましたので、ご意見を言わせていただきます。

【大山会長】

ありがとうございます。

他にご意見等は、ありませんか。よろしいでしょうか。

16ページにわたる大変膨大な資料ですので、またお持ち帰りになりまして、もう一回精読していただいてご意見を賜りたいと思います。

なお、本議題につきましては、これまでの審議を踏まえまして、次回の専門調査会において一応の中間取りまとめを行いたいと思っております。そういう意味で、先ほどもご案内申し上げましたとおり、本日の議論等で足りない点、その他ご意見等は、誠に恐縮ですが、本日配付いたしました追加意見書等によりまして、4月2日、金曜日までにご提出いただければと思います。

なお、前回の本専門調査会後にも同じような資料についてご意見等を提出いただいておりますので、同じ意見を再提出いただく必要はありませんので、よろしく願いいたします。

議題2：評価専門調査会（第33回）の議事録について

平成16年3月4日開催の評価専門調査会（第33回）議事録について、確認が行われた。

【大山会長】

続きまして、議題2の評価専門調査会の議事録について確認をいたします。

前回の議事録（案）は、お手元の資料5のとおりであります。各委員のご発言の部分については、書面で事前に確認いただいておりますので、ご承諾をいただきたいと思っております。よろしいでしょうか。

何かお気づきの点がありましたら、事務局の方までご連絡いただければ修正可能です。

また、本日の配付資料は公表することにさせていただきます。

それでは、そろそろ閉会にしたいと思います。次回の日程について、事務局から報告をお願いします。

【鵜戸口参事官】

次回の日程ですが、前もってご連絡もしていると思いますが、4月21日水曜日の午後1時から4時ということで、3時間にわたりますが、この庁舎の11階の第1特別会議室で開催いたしたいと思っております。毎回大変長時間で恐縮ですが、今回は、先ほど大山会長からもご案内がありましたとおり、海外研究開発機関等の調査報告、これの2回目と、あわせて、これまでご審議いただきました評価における今後の課題と改善方向の検討についての中間取りまとめのご議論をしていただくように予定しております。よろしくお願ひいたしたいと思ひます。

【大山会長】

それでは、本日の会を閉会いたします。長時間ありがとうございました。

また、デービッド・トリンクルさん、ルーク・ジョージューさん、素晴らしいプレゼンテーションをありがとうございました。（拍手）

—了—