

「科研費データの特性と活用可能性」

畠田 敏行（茨城大学）

H26.6.6  
 科研費データを活用した  
 研究力把握に関する勉強会  
 茨：神戸大学

## 「科研費データの特性と 活用可能性」

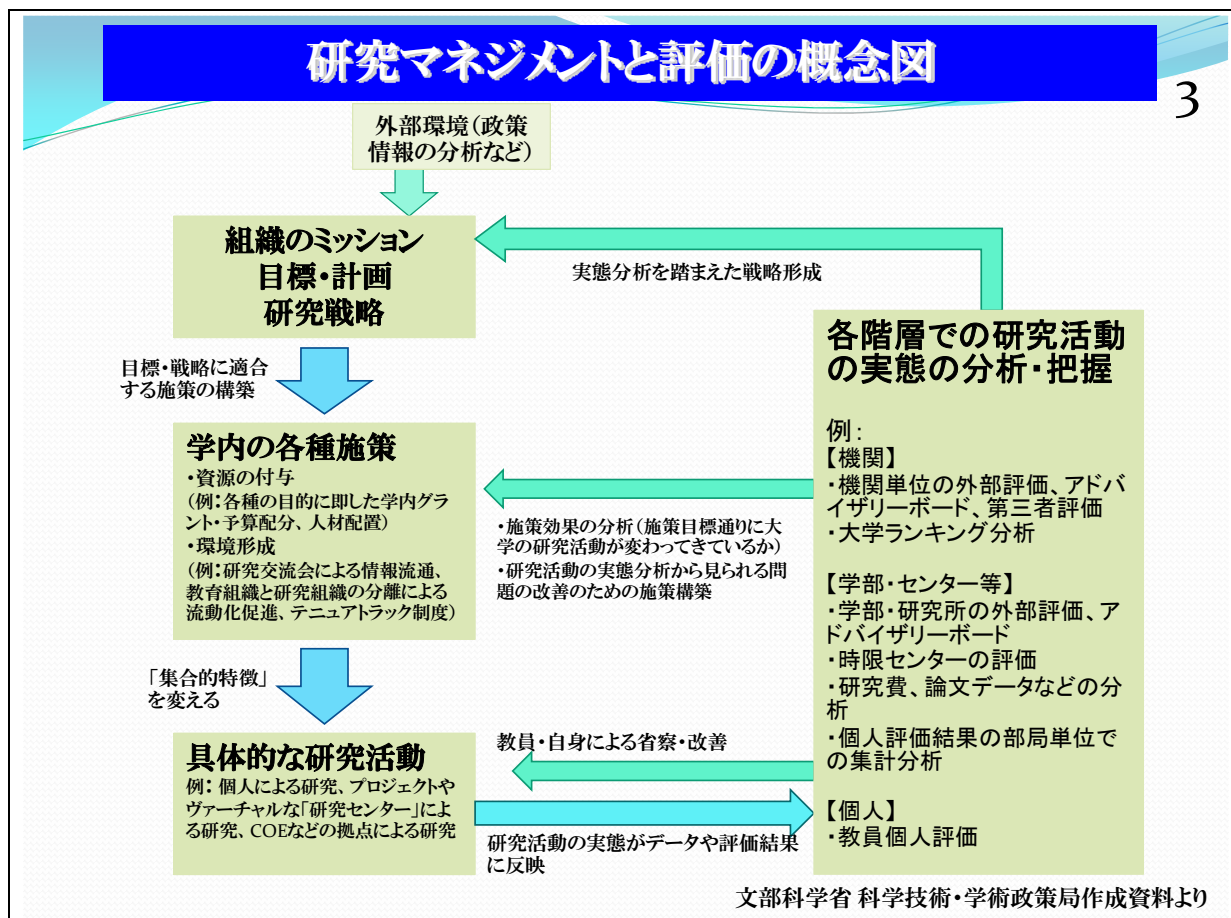
茨城大学 評価室  
 畠田 敏行

2

### コンテンツ

- (参考情報) 研究開発評価とはなにか。
- さまざまな研究に関するデータ
- 今回のデータの集計方法
- データをどのように情報に変換し、どのように役立てるか。

研究マネジメントと研究評価については、評価担当者や研究支援の方が担当すると思います。研究マネジメントを行う以上、全体的に組織のミッションとかあるわけですよね。



学内でいろいろ研究をやるために各種施策を打つ。その下に、具体的な研究活動というのをやっていくわけで、いろいろな階層があるわけです。じゃあ我々—評価、IR、URA、研究支援担当—は何をしなくちゃいけないかというと、例えば機関単位での外部評価に対応したり、例えば研究力把握のために論文データの分析をしたり、大学のミッションに照らして施策がちゃんとうまくいってるのかどうかなどをキチッと分析して対応していかなきゃならない。つまり、我々がやらなくちゃいけないのは、研究マネジメントを回すためのネタ、要するに、必要なときに必要な場所に必要なデータを提供して、研究マネジメントがしっかり回るような、そういう支援をしていかなきゃいけないわけで、今回の科研費のデータなんかもそういうのに使えるんじゃないかなと思っているわけです。

## さまざまな研究関連データ

- 研究情報を取り扱う様々な有料/無料のデータベース。
  - 1) 論文書誌情報データベース  
→ Web of Science、Scopus、CiNii、JDream2、PubMedなど
  - 2) 論文誌引用状況データベース  
→ JCR、JournalRank
  - 3) 学内のデータ  
→ 研究者情報管理システム、担当部署が把握しているデータ
  - 4) 科研費データベース  
→ 国立情報学研究所、文部科学省、日本学術振興会が無料で提供

4

研究関連のデータっていくつかあるわけですが、例えば書誌データベースとして Web of Science とか、Scopus とかあるわけです。あと論文の引用状況としては JCR とか Journal Rank とかあるわけです。学内のデータとしては研究者情報管理システムとか学内のデータベースがあるわけです。あと科研費データベースとかあるわけですが、要は書誌データベースとか Web of Science などの国際雑誌の書誌デ

ータベースなどは多分ちょっと高いです。大きな大学さんですと、結構、余裕で買っちゃうんですけど、うちの大学でも、導入に際しては結構、慎重に検討しました。うちの大学よりも研究規模が小さくなると、Web of Science などの論文書誌データや Citation のデータは有料のものが多いいので、どうしたらいいのっていう話になるのかな、と思っています。

5

## それぞれのデータの特徴

	導入費用	データの鮮度	個人特定	機関特定	データ特性
1) 書誌DB	やや高額?	ほぼリアルタイム	難しい	○	自然科学が多い
2) 論文誌DB	有料	1年ごと	—	—	自然科学が多い
3) 学内データ	DBはやや高額	学内の運用に依存	学内ID	—	全分野
4) 科研費	無料	ほぼリアルタイム	科研費番号	○	全分野

※話者の使用経験にもとづく個人的感想を含む

学内にデータベースを作って、研究者に論文データを入れてください、とお願いしても、なかなか、みなさん入れてくれるわけじゃないですから、そういう発生源入力データでは、実際分析をやりようと思ってもあんまり使えない。そうすると地方大以下のところで、割としっかりと充実した情報で、かつ無料で使えて、しかもリアルタイムで、しかも名寄せができて、かつデータの特性として全分野カバー

してるのは何ですかといったら科研費のデータしかないわけです。ですからせっかく国が無料公開情報として提供してるんだったら使わないのはもったいないでしょうというのが、今回の勉強会のそもそもの発想みたいなどころがあります。例えば、Web of Scienceなどを導入しても、どうしても自然科学の論文に偏重してるので、社会科学とか人文科学はどうなのかっていうところが把握しにくいってところがあります。

6

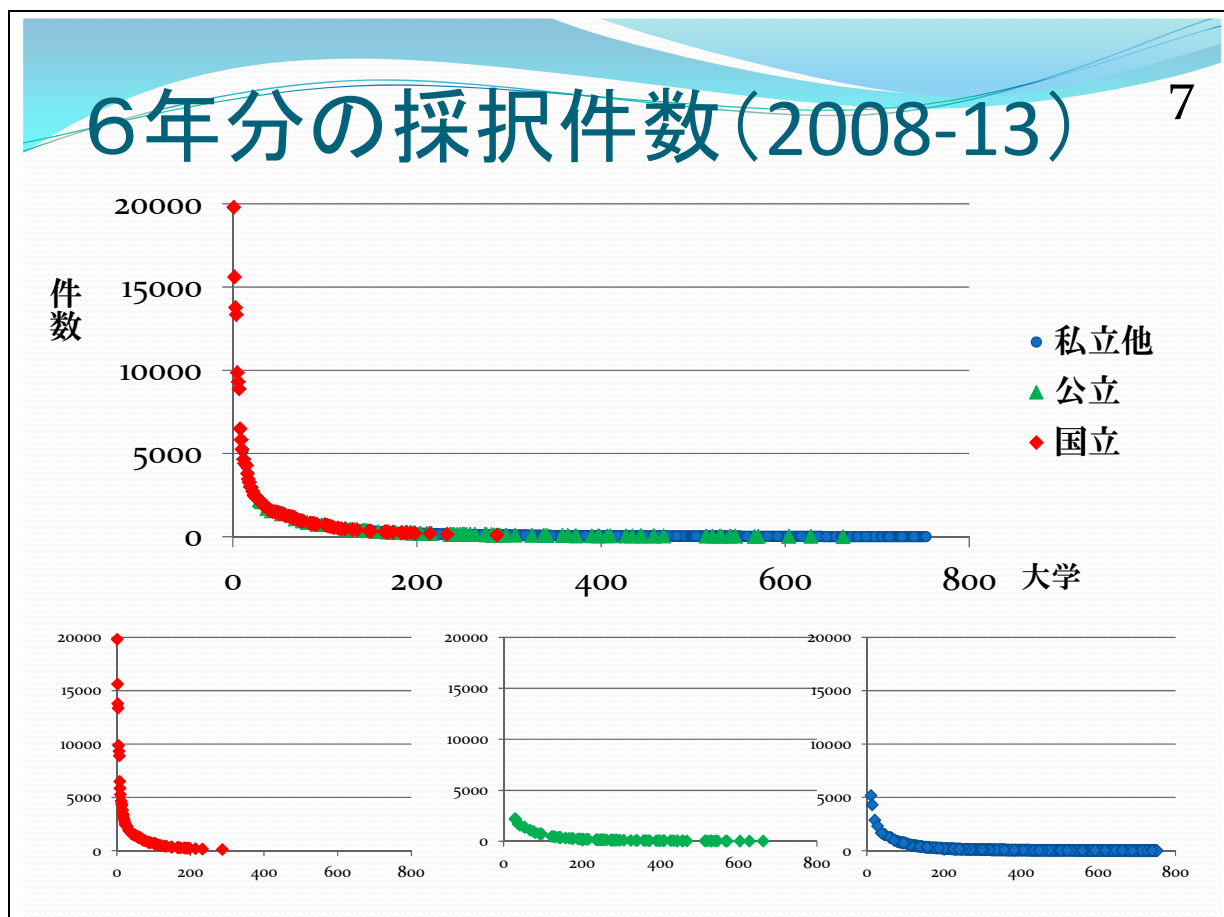
## 科研費データ

- 科研費は単に研究提案の善し悪しで決まるわけでは無い。
- これまでの実績も重視されるピアレビューによる評価で採択が決まる。単なる提案力の評価では無い。
- 文系の分野などもカバーしている。
- 偏りがあるので、そこをどうするか。(シェアや教員数による規格化が必要か。)

	件数%	金額%
1-10位	33.75	51.24
11-100位	45.64	36.70
101位-	20.61	12.07

科研費のデータですが、先ほど藤井さんからも話がありましたけど、研究提案の良しあしだけの勝負ではないわけです。要するに科研費の事務を担当した方なら分かると思いますけど、研究業績をしっかり書く欄があって、あそこが白紙だったら通らないわけです。だからこれまでの研究をやってる方がしっかり次のステップに行くみたいなどころ、要するに事前評価なんですけど事後評価を兼ねているみたいなどころがあります。

萌芽研究というのがあって研究業績を付けずに出すんですけども、これもいろいろ JSPS の人に聞くと、研究をしっかり続けてきた方がちょっとこういうこともやってみようか、みたいな提案を想定しているような節があって、何もやっていない丸腰の方がこういうことをやりたいという研究提案ではないわけですね。つまり、科研費は、研究をしっかりやってないと通らないっていうのが基本的な話になってます。



このグラフが何かっていいいますと、縦軸が件数で、横軸が順位となります。6年分の科研費の採択件数をプロットして、赤いのが国立、緑が公立、青が私立です。このトップ、東大とか京大なんですけど、要は6年間2万件とか万単位で取ってるわけです。多分うちなどはこの辺ですから、全然住む世界が違うなみたいな感じです。ただこれを見ていただくと、国立が結構ほとんど持っています。私立大学さんは非常にロングテールになっちゃうんですけど、私立大学のトップクラスだと結構国立のそこそこいいところと結構勝負ができるんですけども、割と科研費は国立偏重になってるっていうのは確かにそうなんだろうと思います。また、金額ベースで見ると1位から10位の大学で半分持ってるわけです。だからすごく富が非常に偏っているなという感じもします。1位から30位のところはインターナショナル・ジャーナルもいっぱいあるわけですから、そういうところは国際的な書誌データベースを中心に正解との比較をやっていただいて、その次のグループあたりが割と科研費は感度がいいのかなと、と思っています。どういうことかと言うと、中小規模の大学がインターナショナル・ジャーナルで各分野別で世界の平均引用数に対してうちの引用数は何本かみたいなことを調べても、結構、寂しい数字しか出てこないわけです。ただ科研費だったらとなりのあの大学に勝ったとか、この分野ならあの有名大に肉薄してるぞ、とかそういう話になってくるんで、まだ話が面白いというか、あんまり寂しい思いをしなくて済むみたいなのところがあるわけです。

8

## データの作成方法




• <http://kaken.nii.ac.jp/> にアクセス

データの作成方法なんですけども、取りあえずこの科研費データベースにアクセスしてください。とりあえず全部の大学のデータを、2008年から2013年までひたすらダウンロードします。

9

## データの作成方法



年度には、希望年度(今回は2008~2013年の6年間)を設定し、研究機関には1万件を超えないように大学名を次々と入れていく。

1万件がダウンロードのリミットなので、そこに気を付けてください。

10

## データの調整

- csvで出力し、開いたファイルのセルの書式設定を行い、折り返して表示する設定を解除。
- 異動年次は分からないので、獲得大学がすべて研究費を獲得したと看做すために、所属機関が複数ある場合、最初の機関以外は削除する。
- 金額データを年度ごとに分解する。1行が1年度となるようにデータを調整(3年間もらっている科研費ならば3行になる)。
- Lookupテーブルなどを用いて、機関のJSPSコードや分野、分科コードなどを附与する。

この辺には、詳細な方法を示しておきます。複数の機関に所属する場合、機関数で割ってもよいのですが、処理がたいへんなので獲得大学が総取りという方式にしました。ポイントは、年度ごとに金額データを分割して処理するというところです。

11

## データについて

- 科研費の分野・分科・細目などは毎年見直しがある。
- 前は平成25年度に大きな変更があった。
- 分野、分科等の枠を越えて変更があったため、今回は、2008年度から2013年度で汎用的に使える分野・分科・細目を作成した(資料4)。
- その後、金額コード、種別コードなどを附与し6年分の国公立大学の年度ごとの科研費配分状況のマスターデータを作成(データ1)。
- マスターから種別(データ2)、金額(データ2)、分野・分科ごと(データ3)の件数、配分額データを構築。
- 併せて、2008～2013年度の機関別配分額を集計し、順位やシェア等の基本資料とした(データ4)。

科研費コードは不変のものではなくて、見直しがあります。ですので、読み替え表を作って、現在の分野・文科・細目に対応させておく必要があります。

12

## 資料4: 分野・分科対応表

- 科研費の系・分野・分科・細目は、学術の進展に従い、改正される。
- 毎年、微少な見直しがあり、何年かに1度、大きな見直しがある。直近では平成25年度(2013)年度採択分から大きな改正があった。
- 細目番号も原則変わらないものの、大きな改正の際には付け替えなどが発生する。
- 今回は、そのために平成24年度以前と25年度以降を折衷した大学評価コンソーシアム独自の系・分野・分科・細目を設定した。
- 原則的に、現行の体系を利用している。
- (全体的にコードは確認中。)

資料の4というのがあって、科研費の分野とか分科っていうのは毎年変わるんです。要するに系というのがあって、その下に分野、その下に分科、最後に細目っていうふうに、要するに超大大分類、大分類、中分類、小分類となってるわけですけど、結局これが毎年のように変わるわけなんです。なぜかという、学問が進化するからです。

## (参考)考え方のヒント

- 15 研究情報を解釈する:誰の意思決定を支援するのか
- 15-1 研究担当理事
- 16 研究情報から見出したいこと:評価担当者
- 16-1外部資金は増加したか 16-2外部資金獲得支援の取組は成果があがっているか
- 17 研究情報から見出したいこと:研究推進、研究協力担当者
- 17-1外部資金獲得支援策。17-2学内共同研究・プロジェクトのマッチング支援。
- 17-3弱いところのテコ入れ/強みを伸ばす 17-4学部・研究科別/研究分野別
- 18解釈の視点
- 18-1金額、件数。18-2順位。順位パーセント。シェア。18-3経年変化。全国平均や他大学との比較。
- 19解釈時に留意しておくべき科研費事情(Cでよい意識)
- 19-1小額で研究可能な分野は基盤Cへの応募が多い 19-2 1件あたりの金額は研究分野によってちがう、19-3「エフォート」により一人の研究者は応募数の上限がある。19-4どの分野で申請するか問題(農学? 環境学?)

それでどれを何を見るのっていうのはこの辺に書いてあるので、それを見ていただければいいですし、考え方のヒントっていうのはちょっと藤井さんからもらったメモをもうちょっと僕なりにアレンジしてやろうかなと思ったんですけど、多分どうせ時間がないんだろうと思って、そのまま貼りだしちゃったんですけど、じゃあ何の

ために、よく IR の話が出てくると、データは集めたんだけどなんかやることがないんですよ。みたいな話がよく出てくるんですけど、要するに多分大学の執行部とか学術担当の理事とか各学部の研究マネジメントの担当の先生は、本当はなんか困ってるんですよ。次どうしたらいいかなと。どういうところにお金をまいたらどういうふうに研究が進むのかみたいなところを本当は知りたい。そうすると今どうなっているのか、隣のあの大学と比べてどうなっているのか、次にものになりそうな分野はどこなのかとか、そういうところを意外と、本人たちは強いなと思ってるけど本当に強いのかとか、その辺のところを数量的なデータできちんと示してあげることによって、相当先生がたの議論が深まります。実際こういう話を学内で大体年に2回くらいは、全学の研究企画委員会みたいなところで報告するんですけど、結構食いつきはいいです。だからやっぱりデータの定義をきちんと説明して、現状をきちんと説明すると、自然と結局先生がたは興味深いデータはこっちでネタを出してやれば、それで勝手に食いついて議論をしますんで、いかに研究マネジメントを担当する先生方が楽しいなと思えるようなデータをわれわれは提供できるのか、そこがポイントかなと思ひまして、そういうところにこういう科研費のデータが使えればいいなと思っております。ちょっと駆け足だったんですけども、説明は以上となります。