Managing Discrepancies in Evaluation Methods for Interdisciplinary Research Programme: The Case of WPI in Japan

Alfonso Ávila-Robinson[1], Naohiro Shichijo[2], Shintaro Sengoku[1]
[1]Kyoto University, Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto, Japan
[2]National Institute of Science and Technology Policy (NISTEP), Tokyo, Japan

***Abstract*--Interdisciplinary research has recently been emphasised in science and technology policies throughout the world. Numerous organisational approaches are directed at accelerating interdisciplinary research; consequently, new research institutes have been established at the university level or as public institutions. However, principles of management regarding evaluation of interdisciplinary research have not been developed sufficiently in comparison to ordinary mono-disciplinary research. A practical approach is the use of the peer review by a small set of experts; however, the selection bias of peer reviewers, the lack of expertise in emerging disciplines, and a burden for evaluation tasks are currently pointed out as developmental needs. Alternatively, another approach that has been proposed is a more efficient and robust routine for applying scientometric intelligence with established bibliometric indicators and comprehensive publication databases. In this study, an empirical observation was conducted to examine six interdisciplinary research institutes operated for the last five years under an initiative of the Japanese government; results were expected to provide evidence regarding similarities of the two aforementioned approaches. Additionally, a discussion regarding the underlying reasons for possible discrepancies was initiated.**

I. INTRODUCTION

A. Current Conditions of Interdisciplinary Research (IDR)

Contemporary science and technology policies recognise the contributions of IDR to the creation of new academic fields, exchanges of knowledge and personnel across industries, and innovations. The importance of IDR has been stressed in Japan as well, and numerous policies and programmes have been implemented [1–3]. A considerable amount of public funding has been authorised for the establishment of new research and development (R&D) centres, which are characterised by a high degree of freedom in business management policies and sustained effects following the completion of funded programmes [4–8].

In Japan, Special Coordination Funds for the Promotion of Science and Technology, launched under the leadership of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in fiscal year 1980, marked the beginning of such funding. [1] This publicly funded grant programme was implemented to regulate and promote necessary and important matters pertaining to the advancement of science and technology according to standards of the Council for Science and Technology Policy. Research centres at 12 sites were operating in fiscal year 2013 as a part of the 'Creation of Innovation Centers for Advanced Interdisciplinary Research Areas Program'.

The World Premier International Research Center Initiative (WPI) was launched in 2007 under MEXT's leadership. [2] Consequently, four key objectives were identified for achieving the 'world's highest level of research standards', 'creation of fused domains', an 'international research environment', and 'reform of research organisations Additionally, the importance of IDR was emphasised. In fiscal year 2013, nine research centres were operating with a subsidy from this programme.

The Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) was launched in 2009 under the leadership of the Cabinet Office and the Council for Science and Technology Policy.[3] This comprehensive programme applies to various R&D phases ranging from basic research to application developments in various science and technology fields. It also promotes attainment of top-level status for advanced research within three to five years, with the corresponding objective of enhancing international competitiveness and yielding results that benefit society. Although IDR is not an eligibility requirement, most R&D projects have been conducted within a framework that includes multiple institutions and fields. Centres at 30 sites have been adopted and operated under this programme.

B. Issues regarding Evaluations of IDR

Because the evaluation of IDR includes the element of managing diversity, it differs from the traditional assessment for enhancing research that is rooted in the mono-disciplinary approach. Thus, establishment of an evaluation system tailored to IDR is required [8].

An evaluation system based on peer review has been an achievable method for appraising IDR, as it has been applied extensively and flexibly for research. Unfortunately, its limitations associated with the personal attributes of evaluators have been exposed recently [9]. First, few evaluators specialise in two or more fields; thus, bias is quite possible. The second issue is the aptitude of evaluators connected with implemented policies, programmes, and R&D projects. Currently, a researcher who specialises in the targeted field is appointed temporarily to perform evaluations. The standards for appointment in such instances, however, are based on the individual's experience and knowledge; hence, policies and evaluation debates are not necessarily considered. Third, there are restrictions in terms of time and

[1] Website: http://www.mext.go.jp/a_menu/kagaku/chousei/index.htm

[2] Website: http://www.mext.go.jp/a_menu/kagaku/toplevel/

[3] Website: http://www8.cao.go.jp/cstp/sentan/about.html

tasks. Many evaluators have regular responsibilities, and the task of evaluation is conducted as additional work. As a result, evaluators are often unable to allocate sufficient effort to the task. Finally, it is difficult for evaluation committee members to perform valuations when the unconventional activities that are often associated with innovative inventions and discoveries exceed their experiences or understanding.

Recently, attention has shifted to the scientometric evaluation approach that utilises document information such as academic papers and patent documents [5, 7, 8, 10–13]. One advantageous result has been a normative evaluation system for appraising R&D projects in a diverse range of fields. The system facilitates management by objectives and greater understanding, leading to a consensus among diverse stakeholders. Currently, evaluations for IDR and R&D deal only with the mono-disciplinary field. Although evaluation indices for IDR and fusion research are advocated widely, verification from evidence based on actual examples of this type of evaluation is lacking.

Burdens on resources associated with the evaluation process should be considered more seriously [14–15]. Otherwise, essential research performance can deteriorate as more evaluations are conducted. For example, top management at a research centre, which is subject to evaluation, mobilises considerable resources because of the significance of the assessment. Evaluators indiscriminately respond to requests from the government, its ministries, and agencies of jurisdiction; therefore, they may be conducting excessive or unnecessary evaluations. In the case of the WPI Program, interim and secondary evaluation sub-processes, such as site visits and responsive actions associated with audits, exist. Once integrated, the evaluation process is implemented roughly once each year. If 5% of the resources available to top management were directed toward evaluation, the burden would be enormous.

C. Purpose of this Research

An aim of this research was the examination of strategies contributing to the improvement of R&D performance; therefore, consideration was given to the evaluation process associated with IDR in light of the research background and issues described above.

The target of the case study is WPI, which is considered to be a suitable system for this research for the following reasons: (i) the government ensures a suitable budget frame and implementation period as a national project; (ii) the government secures a considerable budget and implementation periods for national projects—the initial budget for the WPI Program was set at 7.4 billion yen and implemented progressively over 10 years; (iii) universities and public institutions, which are host research institutions, are committed to the operation of WPI Program centres (iv) interdisciplinary research and fusion-scientific research are included in the programme initiatives (missions); and (v) diverse natural science fields are encompassed in policies and programmes. In addition, (vi) the WPI Program is oriented toward fundamental research, and it is suited for adoption of the bibliometric approach, which applies a database of academic readings. Finally, (vii) a certain amount of time has elapsed since the commencement of the initiative, so there has been sufficient accumulation of information. An interim evaluation was conducted in 2012, and it is possible to verify and examine results of this research against those from the interim research.

The potential to implement the bibliometric method in the policy programmes of future centres committed to IDR will be determined. Related discussions will include the possible adoption of existing bibliometric indices and new indices, including synthesised indices based on existing ones. Therefore, we are presenting a hypothetical evaluation system for IDR.

II. METHODS

A. Bibliometric Approach

The bibliometric approach was applied to the organisation for collecting publications and corresponding references and citations. Regarding Table 1, the names of the indicators were taken from previous studies [8, 16].

For a given organisation and year range, three sets of publications were collected: (i) an original paper and direct publication of a WPI research institute, (ii) a *citing* paper that cites an original paper, and (iii) a *cited* paper that is cited by an original paper. The variety, balance, disparity, and Shannon entropy indicators were calculated separately for these three sets of linked publications. Further, the coherence indicators were computed according to the organisational perspective.

Regarding the applicable time periods, the values have been fixed for individual year and cumulative ranges. An individual year range is based on a start year and end year. Start year refers to the earliest year of publication of an original publication used for a given indicator, whereas end year refers to the latest year. The citing and cited publications are not restricted by year but their ranges are constrained logically by the year range specified for original publications.

B. Descriptive Observatory Approach

Following the completion of this study focused on the WPI Program, nine centres were adopted, with six centres adopted as of the end of fiscal year 2011 as targets of this research: Advanced Institute for Materials Research (AIMR) of Tohoku University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) of the University of Tokyo, Institute for Integrated Cell-Material Sciences (iCeMS) of Kyoto University, Immunology Frontier Research Center (IFReC) of Osaka University, International Center for Materials Nanoarchitectonics (MANA) of the National Institute for Materials Science, and International Institute for Carbon-Neutral Energy Research (I2CNER) of Kyushu University.

TABLE 1. DEFINITION OF INDICATORS USED IN THE PRESENT RESEARCH

Indicator	Definition	Formula	Reference
Rao-Stirling diversity	Degree of diversity with given two categories using the average distance and Harfindahl-Harfindahl index	$\sum_{i,j} p_i p_j d_{ij}$	[5]
Variety	Net number of categories of a set of publications in a given year range	n	[5]
Balance	Distribution across categories of publications, normalised according to the number of categories	$-\frac{1}{\ln n}\sum_i p_i \ln p_i$	[17]
Disparity	Degree of distinctiveness between categories	$\frac{1}{n(n-1)}\sum_{i,j} d_{ij}$	[18]
Shannon entropy	Distribution across categories (i.e., output publications, references in, or citations)	$-\sum_i p_i \ln p_i$	[19, 20]
Coherence	p_{ij} as the proportion of publications in a cross-reference relationship between categories i and j	$\sum_{i,j} p_{ij} d_{ij} \Big/ \sum_{i,j} p_i p_j d_{ij}$	[8]
Integration Specialisation Diffusion	Degree of diversity with categories, defined hereto as [1 – the Rao-Stirling diversity] for each set of citing, original and cited publications, respectively	$1-\sum_{i,j} p_i p_j d_{ij}$	[5]
Depth	Average number of times a research institute repeatedly has used the citations in its publications	$\sum_{y=t-n}^{t-2} r_{iy} \Big/ q_{i(t-1)}$	[16]
Scope	Proportion of previously unused citations (or new citations) in a research institute's list of citations for the focal year	$\frac{s_{i(t-1)}}{q_{i(t-1)}}$	[16]

Definitions of the denoted variables or parameters are as follows: n, the number of Subject Categories; p_i, the proportion of publications in category i; d_{ij}, the average distance defined as [$d_{ij} = 1 -$ (the cosine similarity between Subject Categories i and j)]; p_{ij}, the proportion of publications in a cross-reference relationship between Subject Categories i and j; q_{iy}, the total number of citation counts in a given Subject Category i in a given year y; r_{iy}, the number of repetitive citation counts in a given Subject Category i and year y; s_{iy}, the number of new citations in a given Subject Category i and year y.

III. RESULTS

A. Descriptive Evaluation

This section relies on the results of the interim evaluations conducted by the WPI Program Committee for each of the pioneering WPI centres. [4] These interim evaluations, conducted on a yearly basis, consist mainly of the review of a self-evaluation report submitted by each centre and the reports from two-day site visits conducted by a WPI working group, including the programme director, programme officers, an international working group, and MEXT and JSPS officials. The main outcome of this evaluation process is a brief report assessing each WPI centre for scientific achievements, implementation as a WPI centre, required actions, and recommendations. At the end of the report, a final score is given to each centre according to their level of achievement. This score is based on a five-level scale: *S*, exceeding goals; *A*, probably achieving goals; *B*, more efforts needed to achieve goals; *C*, having difficulties achieving goals; and *D*, recommending project termination. For the purposes of this section, items related to IDR were extracted from these reports. Additionally, these results were complemented with interviews with WPI programme directors and reviews of annual reports of each WPI centre. Table 2 shows the results of this analysis.

[4] This report is available at: http://www.jsps.go.jp/english/e-toplevel/data/08_followup/FY2010result_e.pdf. (retrieved 1/31/2014)

TABLE 2. SUMMARY OF DESCRIPTIVE EVALUATIONS BY THE WPI PROGRAM OFFICE

Aspect	AIMR	IPMU	• IFReC	MANA	iCeMS
Rationale for interdisciplinarity	• New phenomena and creation of original ideas coming from the fusion of fields • Materials research essentially an integrative field (BioChemPhys materials)	• Research fronts tightly tied to the integration of the fields of theoretical physics, astrophysics, experimental physics, mathematics, instrumentation, and applied mathematics	• Focus on the advancement of the field of immunology through the integration of three I's: Immunology, Imaging, and Informatics	• Convergence of five technologies of nanoarchitectonics for the promotion of fundamental studies and applications in the fields of nanomaterials and nanosystems	• Integration of cell and material sciences (chemistry, physics, and cell biology) as a way to achieve innovations in medicine, pharmaceuticals, the environment, and industry
Interdisciplinarity strategy	• Implementation of joint-research projects across research groups • Fusion as the basic strategy	• Establishment of activities encouraging interdisciplinarity, such as daily coffee breaks, weekly seminars, yearly retreats, and semi-annual workshops, among others • Interaction of young researchers with different backgrounds by top-down initiatives	• Establishment of programs oriented toward 'fusion', ex. *'Research Support Program for the Fusion of different Fields'* that funds research proposals involving a variety of disciplines • Research facilities hosting researchers from multiple fields	• Implementation of the top-down approach 'MANA Fusion Research Fund' encouraging interdisciplinarity • Other more informal approaches of interaction, such as 'Grand Challenge Meetings' and 'Melting Pot Activities' have been formalized	• Implementation of initiatives promoting cell-material integration (start-up grants, annual retreats, seminars, etc.)
Interdisciplinarity-related items from interim evaluations	• Increasing interdisciplinarity reflected in the number of joint publications • More systematic and strategic approaches are necessary for encouraging fusion	• Truly interdisciplinary research center, particularly strong efforts bridging mathematics and physics, and cosmology and particle physics	• Strong efforts to promote 'fusion' among the fields of immunology, imaging, and informatics	• Successful efforts, mentioned above, have been implemented for fostering fusion research	• Interesting interdisciplinary and collaborative projects, such as those combining stem cells and chemistry, materials science & biology, etc. • Insufficient integration among disciplines
Final score from interim evaluation	• B	• S	• A	• A	• A-

As can be seen from Table 2, each WPI centre appears to have well-defined rationale behind their interdisciplinarity efforts. Further, informal and formal activities have been implemented for facilitating and fostering interdisciplinarity in each centre. As shown in Table 2, differences were observed in the ability of WPI centres to crystallise their interdisciplinarity efforts. Here, two research centres stand out for their extraordinary efforts in this regard, namely IPMU and IFReC. As no special comments were made for MANA, it appears that the WPI working group was satisfied with the efforts of this research centre. In contrast, AIMR and iCeMS received critical feedback from their WPI working groups. Specifically, AIMR was requested to implement more strategic and systematic approaches for encouraging interdisciplinarity, and iCeMS was criticised for an insufficient integration among disciplines. Interestingly, the overall score obtained by each WPI centre tended to correlate with the degree of interdisciplinarity observed by evaluating groups. This finding, however, should be interpreted with caution, as the relative influence of interdisciplinarity for the evaluators was unknown.

B. Bibliometric Evaluation

Academic papers prepared at the six centres described above were collected and subjected to a bibliometric analysis during a six-year period between 2007 and 2012.

1) Evaluation Using Diversity Indicators

Measurements for coherence are shown in Fig. 1. The sequence was I2CNER, IFReC, AIMR, iCeMS, MANA, and IPMU, with values of 1.15, 1.11, 0.98, 0.97, 0.96, and 0.95, respectively. Furthermore, the WPI Program average was 1.02. Regarding time series variation, the sequence was IFReC, AIMR, MANA, IPMU, and iCeMS (excluding I2CNER from 2008 to 2012) with the compound annual growth rate (CAGR) of 0.67%, 0.66%, -1.27%, -2.25%, and -10.09%, respectively.

Balance, disparity, Shannon entropy, and variety were four indices used in this research. An additional index—diversity—was used as an index.

Measurements for balance are shown in Fig. 2. The sequence was I2CNER, iCeMS, IFReC, AIMR, MANA, and IPMU, with values of 0.79, 0.78, 0.74, 0.69, 0.64, and 0.49, respectively. Furthermore, the WPI Program average was 0.69. In terms of time series variations, the sequence was IFReC, AIMR, MANA, iCeMS, and IPMU (excluding I2CNER from 2008 to 2012) with a CAGR of 1.73%, 0.25%, -0.36%, -1.59%, and -4.99%, respectively.

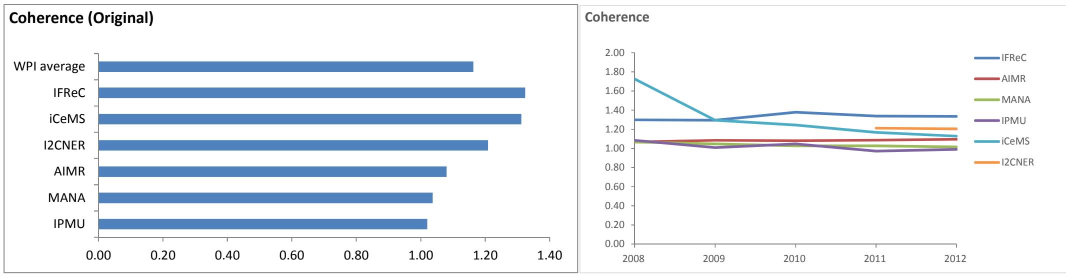

Figure 1. Coherence of WPI Program centres

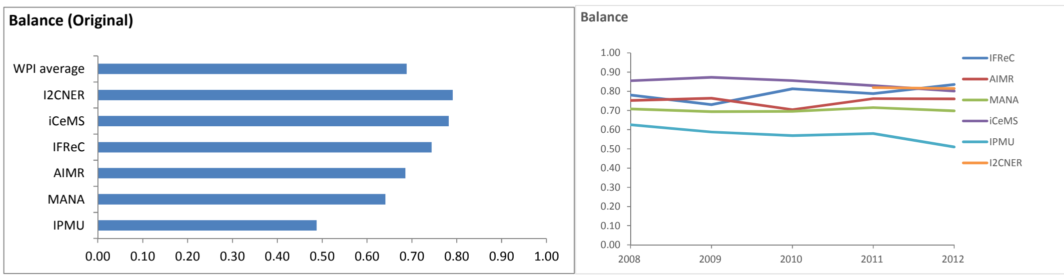

Figure 2. Balance of WPI Program Centres

Measurements for disparity are shown in Fig. 3. The sequence was IFReC, iCeMS, IPMU, MANA, AIMR, and I2CNER, with values of 0.83, 0.83, 0.81, 0.80, 0.80, and 0.74, respectively. Furthermore, the WPI Program average was 0.80. Regarding time series variations, the sequence was IPMU, AIMR, iCeMS, IFReC, and MANA (excluding I2CNER from 2008 to 2012) with the CAGR of 2.11%, 1.66%, 0.41%, -0.36%, and -0.89%, respectively.

Measurements for Shannon entropy are shown in Fig. 4. The sequence was iCeMS, IFReC, AIMR, I2CNER, MANA, and IPMU, with values of 3.43, 3.25, 2.82, 2.77, 2.66, and 1.51, respectively. Furthermore, the WPI Program average was 2.74. In terms of time series variations, the sequence was AIMR, MANA, IFReC, iCeMS, and IPMU (excluding I2CNER from 2008 to 2012) with the CAGR of 2.50%, 2.49%, 2.45%, 1.37%, and -0.87%, respectively.

Finally, the measurements for variety are shown in Fig. 5. The sequence was iCeMS, IFReC, MANA, AIMR, I2CNER, and IPMU, with summary values for five years between 2008 and 2012 of 80, 79, 63, 61, 33, and 22, respectively. Furthermore, the WPI Program average was 56. Time series variations for the summary values were ordered according to IPMU, iCeMS, MANA, AIMR, and IFReC (excluding I2CNER from 2008 to 2012), with the CAGR of 12.2%, 11.7%, 10.67%, 3.23%, and 0.00%, respectively.

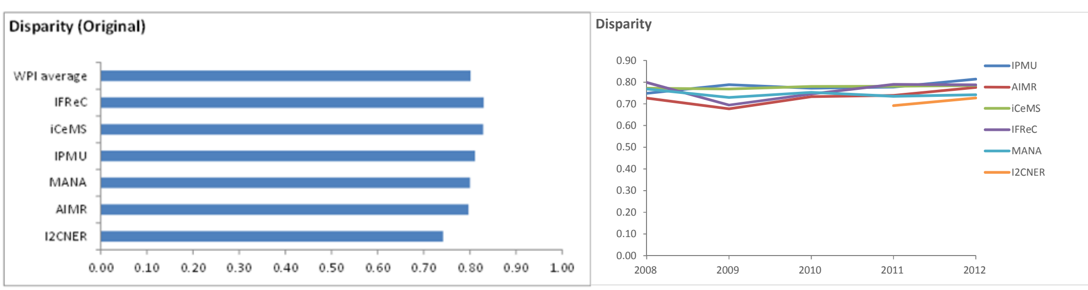

Figure 3. Disparity of WPI Program centres

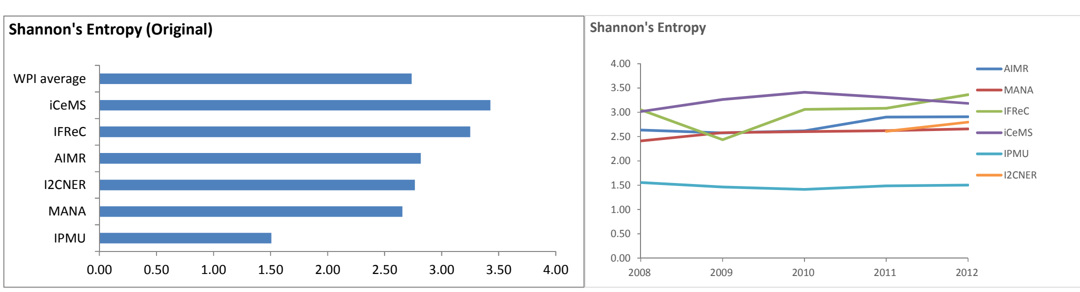

Figure 4. Shannon entropy of WPI Program centres

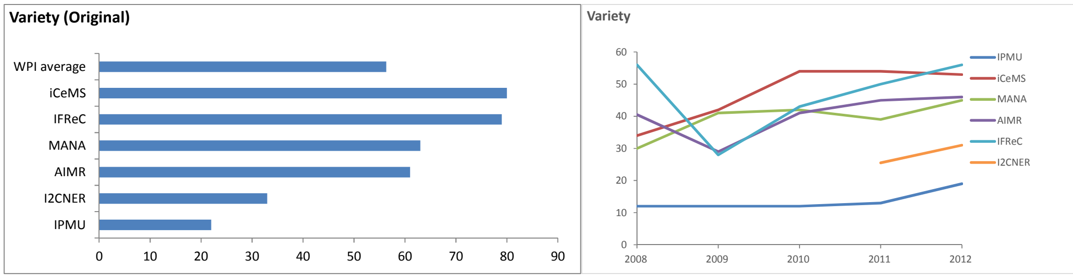

Figure 5. Variety of WPI Program centres

2) Evaluation Using the Diversity-Coherence Matrix

The Diversity-Coherence matrix shown in Fig. 6 summarises the results described in Fig. 5.[5] Rao-Stirling Matrix [5] was adopted as an index of diversity for the purpose of convenience. The origin for both axes was the average value of all six centres targeted in the analysis (diversity = 0.66, coherence = 1.02). As a result of plotting these six centres, IFReC and I2CNER were positioned in the first quadrant (upper right); IPMU was positioned in the third quadrant (lower left); and AIMR, iCeMS, and MANA were positioned in the fourth quadrant (lower right). Based on the relative classification of a precedent research [8], the results suggest that IFReC and I2CNER are *interdisciplinary* centres; IPMU is *mono-disciplinary*; AIMR, iCeMS, and MANA are *multidisciplinary*.

3) Evaluation Based on the Inflow/Outflow of Knowledge

Next, we attempted to evaluate the IDR projects based on Rao-Stirling diversity [5]. In addition to the evaluation for groups of papers published at targeted centres (*specialisation*), evaluations of paper groups referenced by a centre's papers (*integration*) and evaluations of paper groups referencing a centre's papers (*diffusion*) were conducted for each WPI centre. To compare the three evaluation results according to identical standards, indices were unified between 0 and 1, and greater diversity was reflected by smaller values.

The six targeted WPI Program centres were plotted with Diversity in terms of Rao-Stirling diversity measure on the horizontal axis and Coherence on the vertical axis. Both axes are indicated by a divergence rate (%) from average values for the six subject centres.

Two synthesised indices—the ratio of specialisation with respect to integration (hereinafter *inflow diversity*) and the ratio of diffusion with respect to specialisation (*outflow diversity*)—were defined for evaluating the extent of change in diversity along the knowledge flow. Inflow diversity indicates how the large extent of change within scientific fields has been utilised by a centre through citing-cited relationships to precedent research, whereas outflow diversity indicates how the large extent of change in scientific fields has been influential to the centre's publications. In other words, the former term indicates how a centre acquires diversity, and the latter indicates how a centre creates impact.

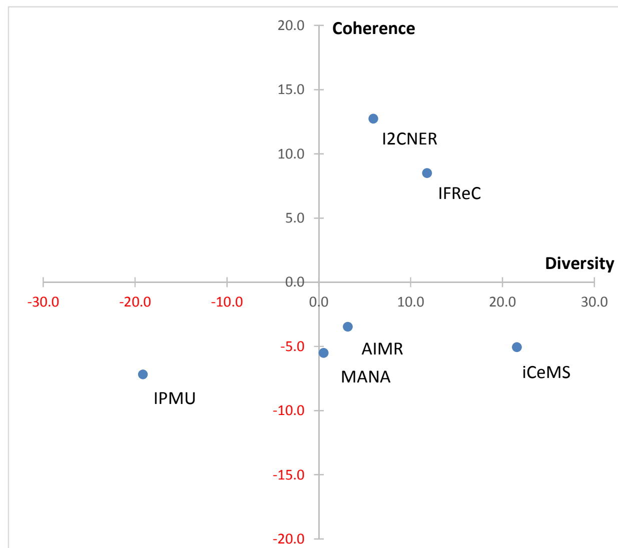

Figure 6. Disparity–coherence matrix of WPI Program centres

Results of the analysis are shown in Fig. 7. Significant differences were found between the original values of the respective centres. The sequence was iCeMS, IFReC, I2CNER, AIMR, MANA, and IPMU, with values of 0.20, 0.26, 0.30, 0.32, 0.34, and 0.47, respectively. The relationship between respective values for citing, original, and cited were such that the original was smaller than citing and cited in all case studies, which suggests that the diversity of research conducted at a centre tends to be greater than the diversity of research used by that centre and the diversity of sites influenced by that centre. In short, broader diversity is required.

In terms of the relationship between knowledge flow and diversity, the inflow of knowledge (original/citing) order was iCeMS, IFReC, I2CNER, IPMU, AIMR, and MANA, with values of 0.49, 0.62, 0.73, 0.73, 0.74, and 0.86, respectively (average of 0.73). Regarding the relationship between knowledge flow and diversity, the inflow of knowledge (original/citing) order was iCeMS, IFReC, I2CNER, IPMU, AIMR, and MANA, with values of 1.42, 1.57, 1.73, 1.73, 1.98, and 2.40, respectively (average of 1.63).

[5] Note that while I2CNER was not subject to an interim WPI evaluation, it was still listed for the purpose of this analysis.

4) Evaluation Based on the Depth and Scope

Finally, we attempted to evaluate the IDR projects based on a set of measures, Depth and Scope [16]. As defined in Table 1, the Depth indicator is defined as the average number of times a research institute repeatedly has used the citations in its publications, whereas the Scope indicator is defined as the proportion of previously unused citations (or new citations) in a research institute's list of citations for the focal year. In other words, the former term indicates how much a centre has been specialising its activities to a targeted area, and the latter indicates how much a centre has been diversifying to broader areas.

The inflow diversity is defined as the ratio of specialisation with respect to integration, indicating how the large extent of change within scientific fields has been utilised by a centre through citing-cited relationships to precedent research; the outflow diversity is defined as the ratio of diffusion with respect to specialisation, indicating how the large extent of change in scientific fields has been influential to the centre's publications. Please note that a smaller value presents a more diverse situation in both indicators.

Results of the analysis are shown in Fig. 8. Significant increase over time was found in the cases of Depth index, particularly greater in IPMU (0.813 in 2010 to 2.309 in 2013) and MANA (0.515 in 2010 to 1.375 in 2013) than other three centres. On the other hand, little change or decrease was observed in the cases of Scope where both the values of IPMU and MANA showed an inverse trend to those in Depth.

The Depth indicator is defined as the average number of times a research institute repeatedly has used the citations in its publications, whereas the Scope indicator is defined as the proportion of previously unused citations (or new citations) in a research institute's list of citations for the focal year. Each centre has relevant dataset in a three or four year-timeframe.

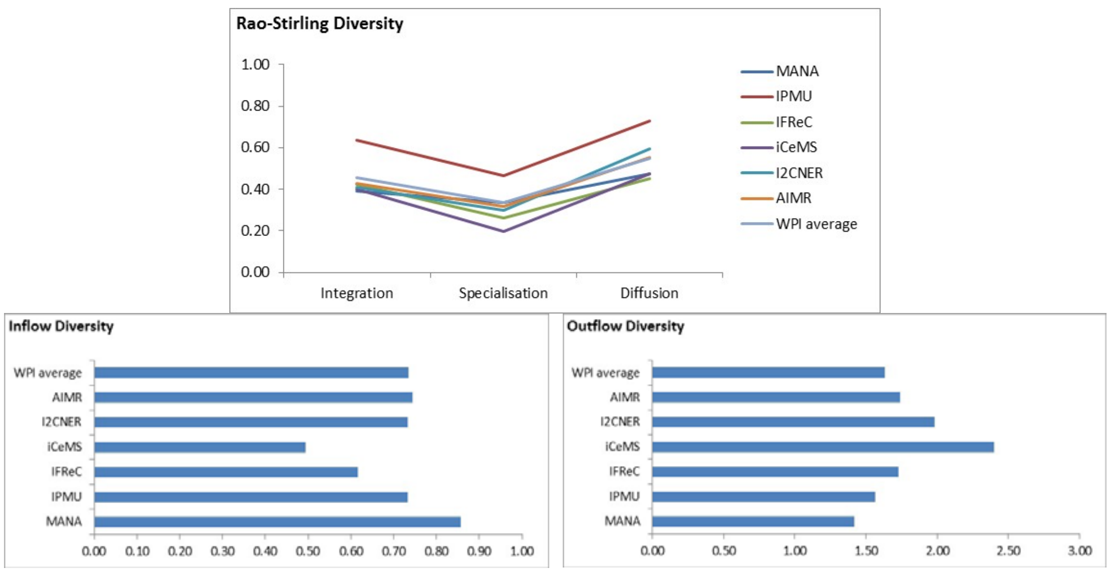

Figure 7. Changes in diversity along the knowledge flow at WPI Program centres

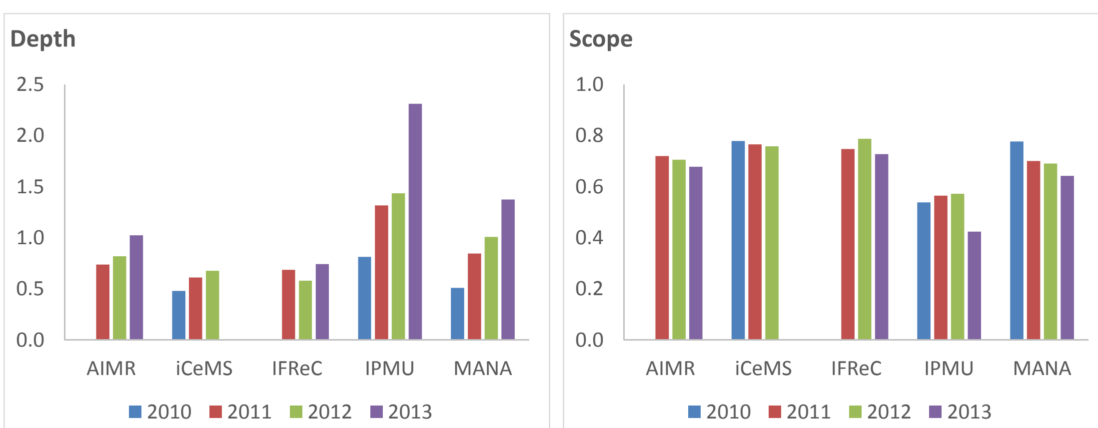

Figure 8. Changes in Depth and Scope at WPI Program centres

IV. DISCUSSION

A. Interpretation of Evaluations and Indices

Differences across the array of diversity-related indicators were observed among WPI centres; amongst the indicators we have tested, the Variety indicator gave the most simplistic comprehension and also significant differences in the present case, however, there is a risk of overestimation associated with the usual correlation of the Variety with the sampling size. Alternatively, we think that the Rao-Stirling diversity index holds a superiority to others because of two reasons: (i) the indicator considers how far a given set of fields are located using the average distance (d_{ij} in Table 1); (ii) the indicator accounts for relative contribution levels in terms of the share of publications in given fields (p_i and p_i in Table 1). On the other hand, it should be noted that this methodology has some technical issues as discussed below (see Section 4.4.)

With most centres, increasing trends in terms of time sequence were also observed for some diversity indexes particularly Variety, and Scope, implying that the targeted interdisciplinary fields increased in association with the development of IDR. The increase in diversity is welcomed on the one hand as it increases technological opportunity; yet, from the perspective of resource management, it may lead to resource dispersal. The existence of a maximum value has been suggested for establishing diversity in IDR [21]. To what level diversity should be set when a centre is established and how diversity should be managed subsequently may be points of contention for management teams at IDR centres.

Differences were confirmed between centres with regard to Rao-Stirling diversity; thus, differences in the interdisciplinary scope were perceived. In particular, the relatively high Rao-Stirling diversity values for almost all research centres pointed to not only the wide array of disciplines being tackled therein, but also to the cognitively diverse nature of their disciplines. The high Rao-Stirling diversity values may reflect attempts of WPI centres to build bridges across typically unconnected scientific and technological disciplines. Nevertheless, as shown in their low levels of coherence, these bridge-building efforts are yet to be reflected in the bibliometric indicators. These findings were confirmed in papers published by the centre and in citing/cited publications, with the prior exhibiting higher interdisciplinarity to match the diversity trend.

Differences in the levels of coherence were observed across WPI centres; however, no differences over time could be confirmed (Fig. 1). In particular, IFReC and I2CNER showed greater levels of coherence compared to the rest of the WPI centres. This result may suggest the difficulty for research centres to interrelate or 'fuse' the array of scientific/technological disciplines they work on into a *coherent* knowledge base. One probable reason comes from precedent observations [13] from a considerable number of key researchers who had a collaborative relationship prior to the launch of a research institute and from researchers with a preference for interdisciplinary and fusion-scientific research. Furthermore, the declining levels of coherence observed for iCeMS should be highlighted as they may imply a gradual regression of IDR activities implemented by this WPI centre. Here, more detailed analyses should address the reasons behind the decline in coherence. Yet, the results should be regarded cautiously, given the uncertainty about the robustness and reliability of this indicator [8].

Summarising, we believe that the results presented in this section may provide useful evidence-based approaches for measuring the interdisciplinarity performance of research institutes.

B. Meta-Understanding of the Present Peer Review Evaluation

We attempted to interpret the results of peer review evaluations based on the evaluation results of this research and grounded in the examination results described in the previous section. As revealed in Table 2, the peer review evaluation results for IPMU, IFReC, MANA, iCeMS, and AIMR resulted in overall assessments of S, A, A, A-, and B, respectively (I2CNER was outside the target of a peer review evaluation).

First, we confirmed that the evaluation results from the Diversity-Coherence matrix did not necessarily conform to the peer review evaluation result (Fig. 6). For instance, IPMU, which was positioned as mono-disciplinary within the quadrant of the prior, was highly evaluated in the latter, and evidence indicated that it was evaluated as 'truly interdisciplinary'. In fact, the ratio of diversity and coherence indicated better matches with the peer review evaluation results.[6] These findings imply that policies for peer review evaluations did not take into account a higher diversity in terms of further average distances likewise that of iCeMS although it requires more effort for integration than a narrower case. Importance, instead, was placed on the internal consolidation that corresponded to the extent of diversity in cases where a certain degree of coherence had been secured. This evaluation trend signifies the importance of diversity in technological opportunities for and compatibility with the selection and concentration of research resources, leading us to believe that research management policies were appropriate.

This interpretation is reconfirmed with another set of analyses using Depth and Scope indexes (Fig. 8), showing that centres with a higher and/or increased level of Depth score tends to be given with better evaluation results, particularly the case of IPMU or MANA. This tendency seems to give a good fit to the key concept of promoting interdisciplinary research which usually requires the integration of heterogeneous specialities and resources into a single or a few one(s) to form a new research discipline. It also implies the importance of defining the centre's research

[6] The low evaluations of AIMR are considered to have been influenced by other factors including the replacement of the heads of centres.

scope in an early phase of the programme; in other words, a risk of changing the centre's directions in the course of programme, which may lead a limited Depth because of a probable dispersion of its academic interests.

In terms of evaluation results regarding Rao-Stirling diversity (Fig. 7), outflow diversity was a good fit with the peer review evaluation results. On the other hand, inflow diversity was not a good fit. These results imply that the diversity of a centre's output has been appraised sufficiently as a value of an IDR project, whereas that of input has been appraised less. We suggest that this trend is potentially fragile, given the concept that the WPI Program centre is oriented toward basic research. In the lead time required until the reference evaluation is established from essay references, fundamental research often requires more time than application research. There may be a danger, then, of promoting fundamental research excessively in the short term and undervaluing activities on the long-term axis. Therefore, we propose that it is time to reconsider the weights of evaluation, with a focus not only on outflow but also on inflow.

C. Proposals for Further Usages

Finally, responsive actions required in the future relative to IDR are identified in this paper, along with considerations for the examinations presented thus far.

Proactive Adoption of the Bibliometric Approach

The usability of the bibliometric approach was indicated as one of the outcomes of this research, signifying its potential to be an alternative methodology to support the existing evaluation approach. Therefore, we propose implementing the bibliometric approach to evaluate IDR projects under the administration of the relevant programmes.

For example, as shown previously, the overall results by peer review evaluation can be well interpreted with the present analytical results. If this trend can be shown to be reproducible, the model of the evaluation process can be planned by adopting this index hereafter; simultaneously, a reduction in evaluation tasks can be expected. Similarly, there is also a potential to use the Rao-Stirling diversity measure for management evaluation of diversity relating to knowledge flow.

Implementation of the bibliometric approach will progress in stages; therefore, there could be instances when it is dealt with merely as a reference index. Even if the peer review evaluations were considered vital, the bibliometric analysis results would inform evaluation committee members in advance; consequently, they would support objective evaluations.

As a precautionary note, there is potential for subjective actions taken by those evaluated in their efforts to receive positive evaluations, such as manipulating the makeup of authors or referenced papers. If such manipulations were conducted arbitrarily, it would be difficult to recognise them merely through an observation of the paper database. Thus, it would be necessary to establish a verification process, such as examining a sampling of published papers.

Meta-Evaluation Using the Bibliometric Approach

Through this research, we learned that the bibliometric approach is an 'evaluation of evaluations'. In other words, it can be utilised to perform meta-evaluations and optimisation of existing evaluation systems.

According to the examination conducted using the Rao-Stirling diversity measure, for instance, the trend to emphasise diversity of outflow over diversity of inflow was observed. If outflow is simplified as an outcome of evaluation and inflow as a process of evaluation, then this research will assist in incorporating perspectives regarding the process designated for future evaluations.

Discussions relating to the indices, other indicators such as the number of citation counts to a specific paper or on average have been widely accepted thus applied as one of standard quality measures, including the case of WPI Program. However, these indices are applied primarily to cases in which (i) the citing/cited paper trends can be predicted as an extension of track records; (ii) a sufficient number of published papers exist within the observed time frame; and (iii) the evaluation on research is limited to a single or small number of fields, in which the lead time for papers becoming referenced is sufficiently short. As a matter of fact, prior research has pointed out the danger of applying such indices to IDR [8].

An Integrative Bibliometric Framework

Making reference to patents, the citation index in patents is recently widely recognised as a method to evaluate the effect of scientific and technological innovation, which is nowadays called science linkage [22,23]. In the present research, the interdisciplinary index was defined and calculated based on the academic publication data. However, achievements of academic research are not necessarily limited to academic publications. In particular, focusing on economical values generated by the outcomes of scientific or technological activities such as patent application, licensing, material transfer, are essential evaluation indices in discussing innovation from universities and public research institutions.

In this sense, a project management style, or managerial framework for academic research institutes must be elaborated by acquiring the technicality of bibliometric approach, taking international and industrial perspective into consideration [13]. Specifically, extraction of requirements demanded of universities and public research institutions that are necessary for the promotion of collaboration between industry and academia activity should be the top priority for empirical demonstration of the present analytical approach.

D. Limitations of this Research

Regarding limitations of this research, there is, initially, the issue of classification sizes in the academic field. The classifications for academic fields adopted in this research were formed by using the first layer of the classification rule independently set by Thomson Reuters. This classification rule is hierarchical, and the n+1 layer is comprised of sub-classifications of respective fields in the n-th layer (and the same thereafter). The first layer was applied as the field classification for convenience; however, based on the results of this research, it would be necessary to apply field classifications based on more layers.

Second, there is the universality of academic field classifications. Many of the classification rules we used have been utilised in a number of prior research studies; therefore, a certain degree of reliability has been secured. Since it relies on particular business enterprises and varies depending on the database source, this application must be considered as one of several interpretations. The optimisation intended for the formation of a more universal academic classification must also be sought.

The third issue is the diversification of document information. This research has dealt with a case study of a policy programme focused on a fundamental research field. Therefore, it was possible to conduct evaluations based solely on document information. On the other hand, in cases of R&D projects that include application research or development, it is possible that the document information may not always reflect true results. In such cases, it would be necessary to expand the database to include patent applications, clinical developments, and so on. Fourth, we have the addition of case studies.

Finally, this research focused only on six centres in Japan originating from the WPI Program. It is essential, then, to increase observation case studies while taking into account each factor described above.

V. CONCLUSIONS

In this research, we examined the potential for implementing the bibliometric approach by focusing on how to conduct evaluations of academic and fusion-scientific research with regard to a current policy programme. By applying the academic paper database and Diversity-Coherence matrix as well as the Rao-Stirling diversity measure, differences were detected among the six observation case studies and the evaluation indices. Their synthesised indices adequately described the policies and trends of the current peer review evaluations, suggesting that the bibliometric evaluation indices can be used alternatively or as support for current peer review evaluations. In the future, it will be necessary to develop a new evaluation system and evaluation process for IDR by taking into consideration the characteristics of other relevant approaches and sustaining the quality of evaluations while reducing the cost and time required for them.

ACKNOWLEDGMENTS

This study is part of the WPI Program under the administration of MEXT in collaboration with the R&D Project of Science for RE-designing Science, Technology and Innovation Policy (SciREX), launched by Japan Science and Technology Agency and MEXT. This research is supported financially by the WPI Program and SciREX. We thank Professors Takanori Ida of Kyoto University and Kuroki Toshio of JSPS, and the SciREX committee members for providing advices on research outcomes; Professor Dr Jordi Molas-Gallart, Dr Ismael Rafols and other related members of Instituto de Gestión de la Innovación y del Conocimiento (INGENIO), Consejo Superior de Investigaciones Científicas (CSIC) - Universidad Politécnica de València (UPV) for their critical comments to an earlier version of this manuscript. We are grateful also to Thomson Reuters for providing analytical support. Our gratitude extends to each of the parties listed above.

REFERENCES

[1] Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, M. Trow, *The new production of knowledge: The dynamics of science and research in contemporary societies*, London: Sage, 1994.

[2] Hage, J. T., R. Hollingsworth, “A strategy for analysis of idea innovation networks and institutions,” *Organization Studies*, vol. 21, pp. 971–1004, 2000.

[3] Corley, E., C. Boardman, B. Bozeman, “Design and the management of multiinstitutional research collaborations: Theoretical implications from two case studies,” *Research Policy*, vol. 33, pp. 975–993, 2006.

[4] Nightingale, P., A. Scott, “Peer review and the relevance gap: Ten suggestions for policy-makers,” *Science and Public Policy*, vol. 34, no. 8, pp. 543–553, 2007.

[5] Stirling, A., “A general framework for analysing diversity in science, technology and society,” *Journal of the Royal Society Interface*, vol. 4, pp. 707–719, 2007.

[6] Shibata, N., Y. Kajikawa, Y. Takeda, K. Sakata, K. Matsushima, “Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications,” *Technological Forecasting and Social Change*, vol. 78, pp. 274–282, 2011.

[7] Rafols, I., M. Meyer, “Diversity and network coherence as indicators of inter-disciplinarity: Case studies in bio-nanoscience,” *Scientometrics*, vol. 82, no. 2, pp. 263–287, 2010.

[8] Rafols, I., M. Leydesdorff, A. O’Hare, P. Nightingale, A. Stirling, “How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management,” *Research Policy*, vol. 41, no. 7, pp. 1262–1282, 2012.

[9] Oppenheim, C, “The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments,” *Journal of Documentation*, vol. 51, no. 1, pp. 18–27, 1995.

[10] Porter, A., A. Cohen, D. Roessner, M. Perreault, “Measuring researcher interdisciplinarity,” *Scientometrics*, vol. 72, no. 1, pp. 117–147, 2007.

[11] Leydesdorff, L., I. Rafols, “Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations,” *Journal of Informetrics*, vol. 5, pp. 87–100, 2011.

[12] Wagner, C. S., J. D. Roessner, K. Bobb, J. T. Klein, K. W. Boyack, J. Keyton, I. Rafols, K. Börner, “Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature,” *Journal of Informetrics*, vol. 165, pp. 14–26, 2011.

[13] Anzai, T., R. Kusama, H. Kodama, S. Sengoku, “Holistic observation and monitoring of the impact of interdisciplinary academic research projects: An empirical assessment in Japan,” *Technovation*, vol. 32, pp.

345–357, 2012.
[14] Gibbons, M., L. Georghiou, *Evaluation of research: A selection of current practices,* Organisation for Economic Co-operation and Development (OECD), Chapter III, 1987.
[15] Kostoff, R. N., "Federal research impact assessment: State-of-the-art," *Journal of the American Society for Information Science*, vol. 45, no. 6, pp. 428–440, 1994.
[16] Katila R., G. Ahuja, "Something old, something new: A longitudinal study of search behavior and new product introduction," *The Academy of Management Journal*, vol. 45, no. 6, pp. 1183–1194, 2002.
[17] Larivière, V., Y. Gingras, "On the relationship between interdisciplinarity and scientific impact," *Journal of the American Society for Information Science and Technology,* vol. 61, no. 1, pp. 126–131, 2010.
[18] Leydesdorff, L., I. Rafols, "A global map of science based on the ISI subject categories," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 2, pp. 348–362, 2009.
[19] Carayol, N., T. U. N. Thi, "Why do academic scientists engage in interdisciplinary research?" *Research Evaluation*, vol. 14, no. 1, pp. 70–79, 2005.
[20] Adams, J, L. Jackson, S. Marshall, *Bibliometric analysis of interdisciplinary research*. Report to the Higher Education Funding Council for England. Evidence, Leeds, 2007.
[21] Kodama, H., K. Watatani, S. Sengoku, "Competency-based assessment of academic interdisciplinary research and implication to university management," *Research Evaluation*, vol. 22, pp. 93–104, 2013.
[22] Narin, F., K. Hamilton, D. Olivastro, "The increasing linkage between U.S. technology and public science," *Research Policy*, vol. 197, pp.101 –121, 1997.
[23] Harhoff, D, F. Narin, F. Scherer, K. Vopel, "Citation frequency and the value of patented inventions," *The Review of Economics and Statistics*, vol. 81, no. 3, pp.511–515, 1999.