Research Evaluation, 2019, 1–11
doi: 10.1093/reseval/rvz022
Article

The balancing role of evaluation mechanisms in organizational governance—The case of publicly funded research institutions

Junwen Luo [1,2,*], **Gonzalo Ordóñez-Matamoros**[1,3] **and Stefan Kuhlmann**[1]

[1]Department of Science, Technology and Policy Studies (STəPS), Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede 7500 AE, The Netherlands, [2]School of Information and Communication Studies, University College Dublin, Dublin D04 V1W8, Ireland and [3]Faculty of Finance, Government and International Relations at the Universidad Externado de Colombia, Bogotá, Colombia

*Corresponding author. Email: luojunwen320@outlook.com

Abstract

Evaluation taking place within publicly funded research institutions (PRIs) has been practiced as a useful instrument to justify PRIs' public funding and to provide evidence for their internal decision-making. The role of evaluation in organizational governance is well-acknowledged as being important in PRIs' management practices. However, it has not attracted much attention from research evaluation scholars. In this article, we propose that evaluation mechanisms perform a balancing role in organizational governance of PRIs with respect to three main aspects: strategy, funding, and operation, where governance tensions often occur between different stakeholders. This research attempts to contribute to a better understanding of why and how evaluation helps to deal with such governance tensions by looking at three case studies, namely the Max Planck Society (MPG), the Helmholtz Association (HGF), both in Germany, and the Chinese Academy of Sciences (CAS). We illustrate the circumstances and conditions in which evaluation mechanisms, where evaluation procedures and culture are institutionalized and stakeholders' interactions are facilitated, help indeed to mitigate the governance tensions.

Key words: publicly funded research institutions; organizational governance; tensions; evaluation mechanism; the balancing role.

Introduction

As per Jansen (2007) 'the term "governance" is a very versatile one'. Across a range of disciplinary discourses among political scientists, lawyers, economists, managers, and sociologists, governance is understood as a highly complex phenomenon and a challenge. Organizational governance can also be defined in several ways depending on the organizations' missions, characteristics and the sectors in which they operate.

Publicly funded research institutions (PRIs) are non-university research organizations that have their own institutions for executing and managing research and development (R&D) activities. They are no exceptions where such governance challenges constantly emerge, especially as a result of public demands for research excellence, societal impact and efficient use of taxpayers' money. For large PRIs, organizational governance consists of various management aspects, stakeholders from hierarchical layers, affiliated institutes at multiple locations working in various fields, and a range of diverse organizational cultures.

On the one hand, organizational governance can be specified by organizational process in particular aspects such as resource allocation, essentially equivalent to a form of government. On the other hand, organizational governance can be understood as a mechanism to capture the effective authority in a collective stakeholder action, especially when no single stakeholder governs the organization. Our definition of organizational governance in the context of PRIs would include both perspectives; one, the processes whereby organizational objectives are set and pursued, and two, ways in which the activities and rules for managing research and researchers within PRIs are agreed, sustained, and regulated among stakeholders.

There are various tools of organizational governance for PRIs. The role of evaluation mechanisms in organizational governance of large PRIs is a question that current literature understudies and this article attempts to address. The rationales and the usefulness of some distinct research evaluation approaches have been extensively studied in literature. However, complex evaluation mechanisms involving various approaches within the same organization which consists of highly subjective and context-specific characteristics and roles have not been studied so far. In this article we study related experiences of world leading PRIs that operate in strong economies and embrace complex organizational characteristics.

We explore three international, large-scale and scientifically productive PRIs and focus on the co-evolution of their organizational governance and evaluation practices in the past 15–20 years. This study aims to fill the gap in literature where the role of evaluation in governance is well-acknowledged in research institution management practices but has not attracted much attention from research evaluation scholars.

To address this gap, a basic analytical framework was designed as a *priori* set of categories to define the vocabulary that would capture the complexity of organizational governance and evaluation practices (Luo 2016). This abductive framework was enriched by empirical research and further developed to a more refined and applicable framework beyond the cases.

In the subsequent sections, we first review the literature on typical organizational governance tensions in the context of national science systems and the embedded research institutions, as well as the role of evaluation as a governance instrument in broader contexts. As a next step, the concepts of 'organizational governance' and 'evaluation mechanism' in PRI context are defined. Thereafter, the refined analytical framework is discussed which involves descriptions of patterns of tensions among the stakeholders, and the circumstances and conditions where evaluation mechanisms perform a balancing role. The method section outlines the overall research design and case studies. Then case study findings are discussed to outline the balancing processes that take place in the three PRIs and the ideal results of the balanced governance expected by the stakeholders. Finally, the conclusion and implications provide a critical discussion and outlook.

Literature review

Governance tensions in PRIs

This section focuses on the essence of 'organizational governance' in PRIs and their related challenges and tensions. We start by reviewing some definitions of 'organizational governance' in literature. Jansen (2007) considers governance in large research organizations as a central factor affecting research behaviour and decisions of the research organizations itself, any subsidiary groups and also the individual researchers. Hermanson and Rittenberg (2003: 27) describe organizational governance as 'a process dealing with the procedures utilized by the representatives of the organization's stakeholders to provide oversight of risk and control processes administered by management'. Heinze and Kuhlmann (2008) claim that organizational governance of research institutions involves distinguishable forms of institutional coordination of autonomous but interdependent units (like research centres or institutes) and stakeholders subject to different types of rules: hierarchy, competition, network, association, and community.

All the above definitions include a recognition of the various groups of stakeholders (affiliated institutes, groups, and researchers) and their behaviours, decisions, rules in specific governance arrangement of PRIs. These different elements are reflected in the two key components within our definition of organizational governance as 'forms of government' and 'collective authority'.

The stakeholders of a PRI include the internal institutes and researchers under the same organizational umbrella, the external funding bodies, the wider scientific communities, the collaborating universities and enterprises, and the public as taxpayers. According to stakeholder theory, which has gained wide acceptance since the 1980s in the domain of organization theory and design (Freeman 1984; Eden and Ackermann 1998), each stakeholder group has their own interests and power which affects the organizational governance arrangement in different ways. For instance, the pursuing of scientific excellence and economic impact of R&D are often controversial factors when planning and developing the research portfolio and strategy for PRIs. These are affected by controversial voices and actions of researchers, both inside and outside the PRIs, as well as the decision makers at different levels. In this way, the governance tensions at large PRIs can be attributed to the diverging interests and powers of stakeholders.

In several countries, like the UK and Netherlands, national governments funders seek justification through organized, regular evaluation of research excellence, and societal impact of all public universities and national PRIs, using both standardized metrics and peer reviews (de Jong et al. 2014; Wilsdon et al. 2015). This can conflict with the organizational development of the PRIs due to their differences in fundamental, strategic, and applied research orientations (Arnold, 2004; Hage, Jordan and Mote 2007; Whitley and Gläser 2007). The lack of convincing evaluation standards is a challenge for allocating public research funding between organizations with different research orientations. Curiosity-driven and open-ended fundamental research may ultimately bring about the most valuable breakthrough, but it is difficult to predict and evaluate their societal benefits early on.

A vigorous evaluation corresponds with the public request for more transparency of investments in PRIs. Public awareness, engagement, and acceptability of science push political actors to rely on vigorous evaluation standards. At the same time, however, this can dilute trust and scientific freedom. Both researchers and managers of PRIs strive for greater autonomy and scientific freedom beyond standardization, and they aim to develop unique profiles and strategies (OECD 2013). There is an ongoing discussion about the nature of public investment and if it should move from pure faith in science, because of its complex and non-transparent nature, to trust based on the evaluation of science and scientists that can also be understood outside science (Cozzens 2007).

The above discussions on governance tensions focus on science systems which embed various PRIs, but still do not unpack the 'black box' of internal governance of individual PRIs. The question remains how one single large PRI can cope with various scientific and societal challenges while adapting to the overarching organizational strategies. There is also the question concerning the problematic nature of the funding competition and research collaboration between the heterogeneous and (semi-) autonomous research units and researchers under the same organizational umbrella.

These questions, rarely discussed in the current literature, will be explored in our case studies.

In practice, a wide range of tools and procedures can address and help manage stakeholders' tensions in organizations. These include but are not limited to negotiation, mediation, and creative peace building (Golden–Biddle and Rao 1997). A common standard of these tools is that they do not intend to avoid any tensions but instead aim to develop the skills of people and enable them to share their experiences and engage in the resolution process (Moura and Teixeira 2010). It is important for researchers to get used to critical peer review for their paper publication, grant application and career promotion, as these are the areas where conflict often occurs. Onsite evaluation at department or institute level has developed to a large scale and carefully organized event for many PRIs where rich interaction between stakeholders takes place.

These evaluations can provide *ad hoc* and fragmented evidence for the decision-making process in the PRIs. However, an overall evaluation mechanism involving all the individual activities within one large PRI can be a powerful instrument giving the PRIs increased internal control in accumulating and synthesizing the evidence for dealing with governance tensions. This can help in justifying and further diversifying the purposes of individual evaluations. In summary, responding to and helping to mitigate complex governance tensions can serve as the rationale behind why PRIs evaluations are designed and applied as an organizational instrument in an overarching mechanism. The next section addresses the concept of evaluation mechanism and how it works as an organizational instrument.

Evaluation mechanisms as an organizational governance instrument

We agree with Jacob, Speer and Furubo (2015: 7) on the 'very few normative claims in literature regarding how evaluation should be embedded in the architecture of governance'. Evaluation was originally conceived as an instrument to guide and improve projects, programmes and policies in the early 1960s. In addition, changes to the functional conditions for research have led to a growing interest in research evaluation since the 1990s (Kuhlmann 2003). In fact, many types of evaluations are used by research organizations as governance instruments to fulfil multiple purposes (Schiene and Schimank 2007; Simon and Knie 2013), such as administrative routine (Hellstern 1986), public accountability (Hansson 2006), justification and improvement of research excellence and impact (Whitley and Gläser 2007), and to strengthen R&D management practices (Edler et al. 2010). However, the question of how evaluations can respond to and help mitigate governance tensions has been overlooked so far.

There is an increasing acknowledgement of a lack of systemic consideration of institutional contexts in evaluation studies (Raina 2003; Edler et al. 2012; Højlund 2014). Hansson (2006) emphasizes the need to understand research activities and their evaluation in the context of research organizations and to make the governance question visible via evaluation. While explaining the instrument of evaluation, the focus needs to be on the evaluating organization and its conditioning factors, rather than on the evaluation itself (Højlund 2014). As Peters and van Nispen (1998) claim, the understanding of institutional instrument could be a part of a cognitive paradigm in a framework of ideas, routines, and values shared by stakeholders. These previous studies support the proposed hypothesis that evaluations have the potential to contribute to the two components of organizational governance, namely specific governance aspects and stakeholders' collective authority.

Hence, we define the concept of evaluation mechanism in the context of PRIs as 'the ways an evaluation system within one single PRI, composed by individual evaluation activities, works and operates vis-à-vis the complex institutional environment and stakeholders'. No individual evaluation activity can, of course, answer all governance questions (Dahler–Larsen 2011). The systemic nature of evaluation mechanism would be the primary facilitator in responding to and in helping to mitigate governance tensions.

Arnold (2004: 3) posits, 'a systems world needs systems evaluations', where the scope and practice of evaluation needs to move beyond individual projects and programmes towards a systems perspective for the whole organizations. It is a general trend that evaluation scope increases with more strategic impact and that evaluation approach becomes more aggregated (Mark, Henry and Julnes 1999; Arnold 2004). The increasing integration of evaluations at organizational level is an effective demonstration of how evaluations become an integrated part of organizational governance (Hansson 2006). Hansen (2013) justifies the impact of evaluation on governance with a condition that evaluative information has governance initiative and is anchored in systematic assessment rather than scattered subjects. For instance, pieces of conclusion and recommendation from multiple evaluations targeting individual subjects (projects, fields, institutes, scientists, etc.) can provide integrated evidence for strategic decision-making and resource allocation across various subjects under the umbrella of one PRI.

In order to be used to mitigate governance tensions, evaluation mechanisms of PRIs should ideally provide systemic and reliable evidence to respond to 'the forefront of the whole organizational thinking and behaviour' (Sanders 2003: 318). Almost two decades ago, Kuhlmann (1999: 136) proposed that evaluation procedures could be used as a medium for the 'moderation' of struggles, controversies and negotiations in the science and technology policy arena. At an organizational level, a conscious inclusion of the perspectives of various stakeholders would also strengthen the systemic nature and strategic function of evaluation mechanisms. We, therefore, raise the question and investigate if a similar mode of 'moderation', that we refer to as 'balancing' exist for PRIs.

Any evaluation that fails to attend to key stakeholders would be inaccurate and insensitive, and thereby insufficient for making the required improvements. This would then lead to the affected leadership groups avoiding evaluation in the future (Bryson, Patton and Bowman 2011). To prevent this and to create productive interaction (de Jong et al. 2014) within PRIs, the voices of the key stakeholders are aggregated in order to ensure the tensions are not hidden but instead brought to the table for them to be mitigated. Moderation of decision-making processes becomes possible in the negotiating arena of stakeholders only if the rules and different stakeholders' perspectives are known and influenced by moderators (Kuhlmann 1999). Therefore, it becomes important that the evaluation mechanisms in the PRIs are able to engage stakeholders in evaluation activities and make them negotiate productively regarding governance tensions.

Analytical framework

The analytical framework, shown in Figure 1, aims to guide us to open the 'black box' of PRIs' organizational governance and to

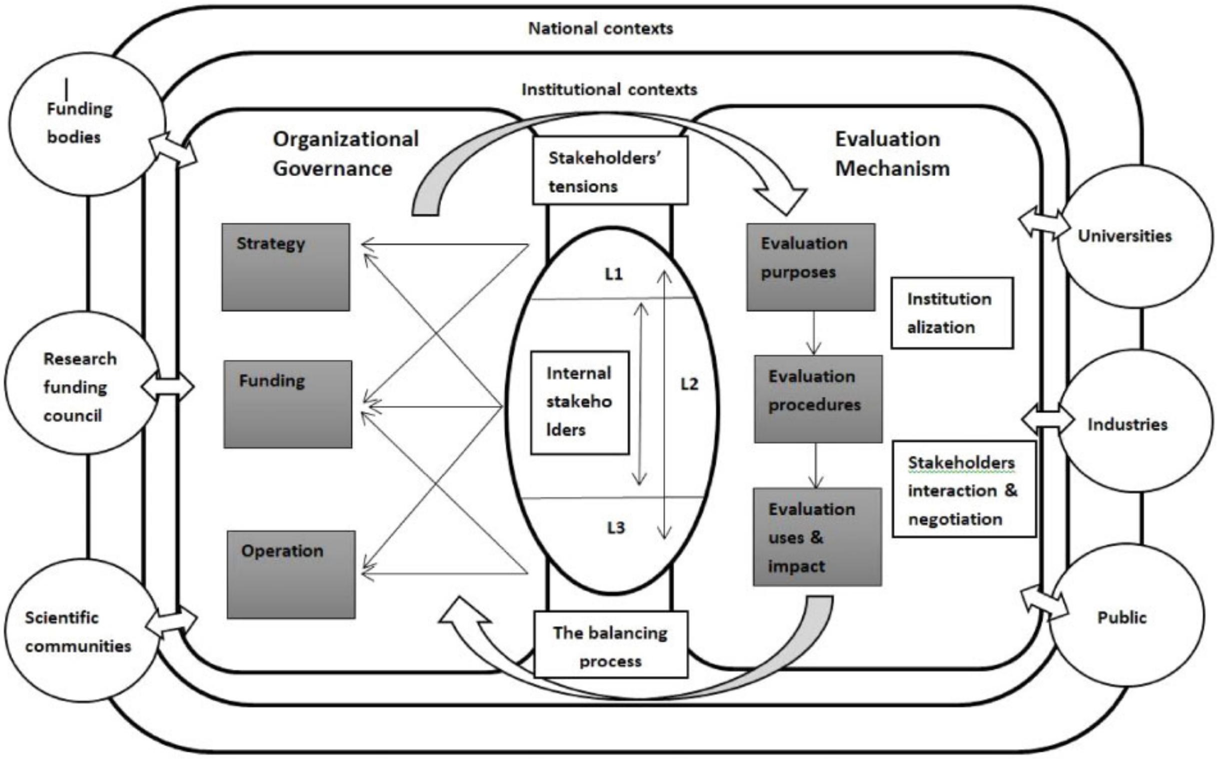

Figure 1. An analytical framework.

understand why and how the balancing role of evaluation mechanism works.

In order to be able to clarify and explain the two primary components of our definition of organizational governance, the three aspects (strategy, funding, and operation) are used to specify the 'forms of government' within the PRIs' organizational governance where tensions occur. First, strategic planning and management of PRIs are highlighted in term of their governance arrangements (Porter 1996). Second, a financial system with principles of resource accumulation and allocation is considered as an underlying requirement for governance advantage in successful organizational development (Guerrieri and Tylecote 1997). Third, an operational process is considered that makes a system work towards its strategies and involves organizational structures, arrangements, and stakeholders within a national and institutional context (Arundel et al. 2007). We propose that the organizational governance of one PRI can be characterized in terms of these three aspects, where each aspect involves the organizational objectives, activities and management rules vis-à-vis the environment and stakeholders.

The second component, 'collective authority', is specified by the three organizational levels, shown in the centre of Figure 1, which enable us to locate the position of the internal stakeholders within PRIs and describe their tensions. These are decision-making and supervisory bodies at macro level (L1); research institutes at meso level (L2) under an organizational umbrella; and individual researchers at micro level (L3). Within the national and institutional context, various stakeholders from inside and outside of PRIs interact with each other with their diverging interests in different governance aspects, as well as holding powers at different levels. Therefore, the patterns of organizational governance tensions of PRIs can be described as both aspect-wise (strategy, funding, operation) and level-wise (L1, L2, and L3), which reflect the two components in the definition of organizational governance.

This article aims to understand and outline the relationship between 'organizational governance' and 'evaluation mechanism' as a substantive contribution to existing literature. Governance tensions, as the first bridge (shown by the arrow at the top in Figure 1), justify the purposes of constructing evaluation mechanisms within PRIs. The second bridge (shown by the lower arrow) denotes the balancing process of evaluation mechanisms responding to and mitigating governance tensions. The balancing role of evaluation mechanisms is defined to involve both these bridges and perform dynamically as co-evolution between organizational governance and evaluation mechanisms. Such co-evolving process can be explored by addressing two questions: one, what types of governance tensions need to be balanced that justify evaluation purposes (why to balance); and two, how evaluation mechanisms are designed and used to respond to and mitigate the governance tensions (how to balance).

It is important to explore the tensions both aspect-wise (strategy, funding, and operation) and level-wise (L1, L2, and L3) to answer the first of these two questions. The second question addresses the complementarity of the purposes, procedures, uses and impact of individual evaluations composing the overarching evaluation mechanism. We identify two elements, institutionalization of evaluations and interaction patterns of stakeholder, as necessary conditions of the balancing processes. To explore the balancing process, we need to review the institutionalization dynamics of organizational governance and evaluation mechanisms from the previous years. Furthermore, we need to analyze how various stakeholders are engaged in evaluation activities, and how individual evaluation evidence is aggregated in a mechanism that allows it to be systemically deployed in organizational decision-making.

Method

This study consists of two rounds of desk research, prior to and after the field work. In the first round, we did an exhaustive collection, review and critical analysis of a broad range of existing literature. We selected three case studies because of their similarity in scale, funding source, three-level structure, international reputation and scientific success since CAS ranked first, MPG third, and HGF eighth on the Nature Worldwide Research Institutions Index 2014 (Nature Publishing Index 2014). These three cases also differ in their organizational missions, research orientations and funding allocation principles. It is expected these similarities and diverging characteristics would provide a comparable framework to test our hypothesis.

The primary analytical framework with a *priori* set of categories was designed to unfold the complexity of organizational governance and evaluation practices. The overall research questions were designed to guide the empirical research in an abductive way. Empirical data was collected via visits to the three PRI headquarters and 18 affiliated institutes. Along with this, 57 semi-structured interviews were conducted across the three organizational levels. To ensure comparability, we selected the institutes working in similar fields. The software tool ATLAS.ti was used to code, categorize, and integrate the qualitative data from documents and interview transcripts.

The second round of desk research includes analyses of the data to find out why and how evaluation mechanisms in PRIs contribute to the mitigation of governance tensions. The above iterative exercises helped to enrich the research and resulted in the proposed analytical framework (Figure 1), which can be used to understand the phenomena more broadly. The next section will present the case findings in detail.

Case study findings and discussions

Characteristics of organizational governance at MPG, HGF, and CAS

Germany and China have both embraced significant growth through extensive R&D investment with complex research funding structures and similar categories of powerful stakeholder groups, particularly national funders. Germany has a more comprehensive R&D landscape and clearer categories of PRIs missions with the four key PRIs including MPG and HGF (EFI[1] reports 2012; 2015; Kupferschmidt and Vogel 2014). MPG aims at conducting fundamental research at the highest level worldwide. HGF contributes to solving grand and complex challenges of society, science and industry. The German Council of Science and Humanities (Wissenschaftsrat[2]) conducted the first (and only one so far) systematic evaluation of all the four PRIs in 2001 (Wissenschaftsrat Evaluation Report 2001). This can be seen as a shift towards output assessment and an important tool for strategic change in the German public R&D landscape.

The stated missions of the three PRI cases determine their research orientations and governance arrangements. The core strategy of MPG to conduct the best fundamental research in the world is supported by its high level of public funding (84%) and a person-centred governance framework called the Harnack Principle.[3] The 82 MPG institutes enjoy a high level of freedom to conduct curiosity-driven research in an open framework and without ties to specific applications (Schruff 2012; Max Planck Society 2015). For HGF, also with a high level (70%) of public budget, the 18 centres conduct strategic research in six fields[4] with approximately 30 highly collaborative programmes (Gazlig 2009, 2012; HGF 2014). They call it Programme-oriented Funding (PoF).

As the unique and dominant PRI in China, CAS often gets labelled as 'no organization in the world with so many functions' as well as 'being too big and unwieldy and lacking of inappropriate evaluation system' (Cyranoski 2014: 468). CAS is undergoing massive transformation with a prominent focus on R&D spending and an evaluation framework aiming at clearer categorized R&D landscape (Bai 2012; Sun and Cao 2014). CAS has complex research orientations fragmented among its 104 institutes with a funding framework of 50% block funding and 50% competitive grants. Its governance reforms aim at aggregating individual institutes to pursue the major outcomes of the whole institution (Bai 2012; Luo, Ordóñez–Matamoros and Kuhlmann 2015).

Governance tensions found at MPG, HGF, and CAS

The three-by-three matrix in Table 1 summarizes the patterns of governance tensions found in the three PRI case studies. Overall, the three cases share typical elements of organizational governance and similar categories of internal and external stakeholders. In the interviews, the word 'tension' sounded negative and led to many interviewees being reluctant to talk about it explicitly. We observed that rather than the existing tensions, many interviewees talked about the actions and efforts that had allowed them to successfully prevent certain negative consequences in the past, or problems predicted for the near future. It was a process of probing, interpreting, and refining rather than explicitly identifying these tension patterns.

Strategy

Two tension patterns have been identified between the stakeholders at the highest organizational levels L1 related to strategy. The first is between the external justification of public funding and internal governance of heterogeneous research activities across diverse disciplines and fields. While the former asks for public understanding, support, and transparency of organizational governance, the latter is always complex, and, to some degree confidential and difficult for external stakeholders, like politicians, to understand. The complexity of the internal governance of heterogeneous research activities is the fundamental reason to deploy certain evaluations, and a shared reason by some other tensions across the three levels.

This observation is seen more explicitly in MPG because its fundamental research has fewer immediate societal applications and responds less actively to national policies compared to HGF and CAS. Moreover, the scientists that were interviewed at MPG were seen to prioritize research quality and reputation in scientific communities over the standardized expectations of the funding bodies and the public.

On the contrary, HGF has been addressing complex societal challenges through its interdisciplinary R&D with strategic orientation. To react to the national changing policies and emerging challenges, HGF has drawn upon diverse teams working together across disciplines and fields. As discussed previously, there can be a potential tension between policy-based strategic or applied research and open-ended basic research. The explicit and high expectations placed on the societal impact of the HGF R&D leads to it having the most strictly defined guidelines on research themes and applications from the three cases. An interview with a high-ranking executive in the HGF headquarters shows that policy-driven strategic

Table 1. Patterns of organizational governance tensions at MPG, HGF, and CAS

Organizational levels	Governance aspects		
	Strategy	Funding	Operation
Macro level L1	External justification of public funding versus internal governance of heterogeneous R&D activities	Conflicting factors in allocation of internal rescourse to the research institutes	Overarching policies of the PRIs do not one-size-fits-all the research institutes
Meso level L2	PRI's overarching strategy versus institutes' autonomous development	Institutional block funding versus third-party competitive grants	Autonomous power of institutes varies
Micro level L3		Applications and evaluations of competitive grants distract scientists and become their burdens	Scientific freedom is threatened by management instructions and evaluations

research gives rise to many other external expectations, including political advice on education and internationalization, which is far less interesting for natural scientists.

This type of tension is also found in CAS. Over the past two decades, its overarching policies have focused more on national demands than institutional development. This is particularly driven by economic impact that is expected from R&D in line with the innovation drive in China since 1990s. After the economic crisis in 2008, many research programmes in CAS have focused on the capability of technology transfer of all the institutes, including those doing basic research which are vulnerable in funding applications.

The second type of strategic tension exists between the headquarters' overarching strategy and the institutes' autonomous development. This tension occurs primarily in HGF and CAS because they have large numbers of legally independent units at L2 that are required to strongly align with the overall organizational strategy. However, most of the HGF centres, with well-equipped facilities and world-renown scientists, have been powerful for decades, even before the HGF umbrella organization emerged. The autonomous HGF centres are capable of developing beyond national policy guidelines. Even more difficult is the case of CAS where the large number of heterogenous units at L2—research institutes, universities, laboratories, technology support centres, companies, etc.—and their diverse development, make the unified organizational governance nearly impossible. The complex and diverse research orientations of CAS have caused low efficiency of diffusion of strategies from L1 to L2.

Funding

A typical tension identified at L1 is the allocation of internal resource to the research institutes at L2, the key R&D producers and evaluated subjects. The two common, but often conflicting factors, for resource allocation include the short-term performance and long-term potentiality of the research to be funded. Standardizing scientific performance and potentiality across diverse research fields is challenging as confirmed by all the interviewees. Funding tensions at L2 exist in all the three PRIs, in varying degrees, in terms of funding priorities between the institutes and different programmes, and also between ongoing excellent research and strategic potential research that have not performed satisfactorily.

Another tension between research collaboration and funding competition of the institutes at L2 is more obvious for HGF and CAS both of which have a programme funding framework as opposed to MPG with a person-centred funding framework. The HGF centres collaborate actively for complex programmes, meanwhile competing in the same PoF period, which is controversial between large- and relatively small-scale centres, and between senior and newly established centres. Larger centres working on multiple programmes are found with more capability and flexibility, as compared to the smaller centres with only one programme, to adjust their competition and cooperation with other centres. CAS faces the lack of scientific collaboration, according to the interview with the CAS president by *Nature* who stated that 'scientists shy away from collaborations because co-authorship dilutes their achievements in the eyes of grant committees' (Cyranoski 2014: 469). This lack of collaboration results in duplication of research efforts and also missed opportunities to share knowledge between the CAS scientists and their embedded institutes.

Additionally, third-party funding focusing on certain research themes can facilitate institutes' R&D diversity and autonomous development but also entails the risk of diluting PRIs' overarching strategies. Because of this, third-party funding is only a low proportional supplement in the two German cases. Yet, the relatively high level of third-party funding in CAS—over 50% and in some institutes nearly 70% leads to the scientists being distracted from institutions' research and facing a heavy burden of numerous applications and evaluations.

Operation

The operational tension at L1 is that the top-level policy is not one-size-fits-all the heterogeneous and (semi-) autonomous units at L2 and individual researchers at L3. The MPG institutes develop quite independently in scientific domains, but they are legally dependent on operational rules set by the headquarters, such as the restriction of the number and budget for staff positions. In contrast, although conducting research on strictly defined themes, the HGF centres are operated highly independently from the headquarters in terms of administration. Some of the 104 CAS institutes have legal independence while some do not, due to complicated reasons like history, geography, and size, and they embrace different degrees of flexibility when following the operational rules set by the headquarters.

An operational tension actively discussed at L3 is that the scientific freedom of individual researchers, such as their ability to pursue their own initiatives, becomes increasingly threatened by vigorous management instructions and evaluations. A consensus emerged from almost all the interviewees at L3 that research should be evaluated but researchers' freedom should not be hampered.

The interviewees appreciate independence, freedom and trust to explore curiosity-based research rather than being strictly monitored.

The balancing process of evaluation mechanisms

The three organizations, MPG, HGF, and CAS are not the subjects of nationally regulated science evaluation systems. They design, fund and use their own evaluations for their internal governance. The case studies confirm the systemic nature of our definition of 'evaluation mechanism' because the individual evaluation activities of the three PRIs are systemized to supplement each other in a number of dimensions. They include levels (L1, L2, and L3); subjects (institute/centre, project, scientist, etc.); timing (*ex-ante*, interim, *ex post*); scale (big, medium, and small); frequency (*ad hoc*, regular); participation (of which stakeholders); uses (for what purposes); and impact (in short or long term).

For instance, MPG's cluster evaluation of collaboration potential and organizational synergies of the several institutes working in similar disciplines (every 6 years) makes use of the previous evaluations of these individual institutes (every 2 years) as well as that of the individual scientists (annually) (Max Planck Society 2013). This is done in order to save resources and to recommend institutional change with longer term perspective across the individual units. Another example is the timing of the HGF's two key evaluations, both every 5 years and taking place in the mid-term of each other's timeframe, to make sure evaluation evidence stays consistent every two and half years. The CAS's ongoing reform on the 'One-Three-Five' (one position, three breakthroughs, and five directions) evaluation mechanism, aims at aggregating the evaluation evidence on the 'One-Three-Five' aspects from each of the 104 institutes for developing the 'One-Three-Five' strategy of the whole institution (Chinese Academy of Sciences 2012, 2014; Chinese Academy of Sciences Evaluation Centre 2013).

Institutionalization of evaluation mechanisms

This and the next section will present the two conditions in the balancing processes of the evaluation mechanisms in the organizational governance. The first is the institutionalization of evaluation mechanisms. In this study, this refers to the process of professionalizing and systemizing individual evaluation activities at the level of the whole research institution. This institutionalization is identified in all the three cases as an increasing trend in their organizational dynamics in the past 15–20 years. The steps and extent of institutionalization of each PRI under different contexts are very different, even for the two German PRIs in the same national context, so hardly comparable with each other.

By reviewing their dynamics, we found that the three PRIs have constantly institutionalized their evaluation activities by enriching a number of dimensions, including levels, subjects, timing, scale, frequency, participation, uses and impact, as discussed earlier. The overarching evaluation mechanisms, therefore, cover these dimensions to a large extent. More multi-functional evaluations have been in use since the 1990s and they have consistently contributed to the increasing capability of the evaluation mechanisms responding to the governance requirements. Their institutionalizations all have experienced decades of internal spread and diffusion of evaluation principles, procedures, and cultures.

Such institutionalization processes are confirmed by the interviews as well as the development of their evaluation protocols and guidelines that are more accessible and transparent than evaluation results, e.g. the Scientific Advisory Board Rules for MPG, Strategic Guidelines and Position Papers for HGF, and One-Three-Five Evaluation Policy Document for CAS. These protocols and guidelines specify evaluation procedures and have increasing emphasis on instructions of guaranteeing objectivity and consistence of evaluation evidence when serving governance decision-making. Most of those guidelines are kept updated along with the period or round of evaluations. For instance, the increasing involvement of international experts is considered by the three PRIs as an explicit contribution for the evaluation institutionalization because it addresses global scientific concerns and helps guarantee objectivity of the evaluations. Furthermore, each PRI studied has an internal unit responsible for organizing and coordinating evaluation activities and analyzing evaluation results. The development of these units in terms of scale and impact reflect that their coordination and analytical work also get institutionalized.

Evaluation culture, as a sign of institutionalization of evaluation mechanisms, can be observed from evaluation being increasingly identified by PRIs' staff as a mature and useful instrument for organizational governance. Specifically, all the interviewed staff of MPG agrees that evaluation helps realize the organizational mission to achieve the best fundamental research, to win the best scientists worldwide and to guarantee its person-centred funding framework that has been applied for nearly 100 years. Prioritizing strategic alignment of the highly collaborated programmes across the centres is supported by the interviewed HFG staff in the PoF evaluation. The individual HGF centres are proud to be empowered to conduct centre-level mid-term evaluation themselves to guarantee their scientific quality.

In CAS, the One-Three-Five evaluation mechanism is being institutionalized in a very top down manner. Each of the 104 institutes is required to sign a mission statement including their 'one position' targeted within their scientific community either nationally or internationally, 'three breakthroughs' aimed as long-term scientific achievements and 'five directions' planned for development in the next years. Such a mission statement would be evaluated by the CAS headquarters every 5 years with the help of on-site peer review and recommendations. This 'One-Three-Five' criteria aim at integrating the previously fragmented and short-term evaluations to set long-term and strategic plans for the CAS research system, which gets quick and massive coordination across the different organizational levels.

The mature evaluation culture identified across the cases with a high level of acceptance goes against some studies (e.g. Leisyte, Enders and de Boer, 2010) showing the resistance behaviour of researchers towards institutional changes. We would not state that there is no resistance at all in our cases. Nevertheless, a common reason emerges from the three cases for building such a culture which is researchers' commitment to their organizational missions. For HGF and CAS, strictly defined research missions and fields are prioritized over the assumption of Leisyte, Enders and de Boer (2010) that researchers, in principle, want to design their own research questions and maintain autonomy as much as possible. Scientific autonomy, usually supported by institutional block funding, can only be guaranteed for HGF and CAS scientists when they conduct mission-oriented research. The MPG researchers already embrace a high level of scientific freedom, and evaluations mostly aim at scientific quality and learning without explicit impact towards resource allocation.

The evaluation institutionalization of the three PRIs also involves the influence towards and responses from stakeholders—their coordination on evaluation procedures and acceptability of evaluation culture—as the second condition of the balancing process. We discuss this in the next section.

Increasing interaction and negotiation of stakeholders

Across the three PRIs, their evaluation procedures are found to intentionally and increasingly push multi-level stakeholders to interact and therefore open up deeper dynamics of stakeholders' negotiation. Those interactions consider the diversity of external and internal stakeholder interests beyond the individual evaluation objectives and aim at identifying and addressing both their common and diverging concerns. Participation, conversation and learning enrich the rationality and mutual understanding of the evaluated and stakeholders. The quality of dialogue is thus enhanced, which helps PRIs to remain sufficiently responsive to external and internal changes (Van der Knaap 2006).

A greater involvement and active participation of external stakeholders in internal decision-making process is found in the two German cases through the diverse composition of their Senate members from a wide spectrum of backgrounds, such as politics, academics, business, and industry. However, CAS is given politically defined tasks from its one and only leader, the State Council. Such a centrally planned system setting up research scenarios for CAS is characterized by a close entanglement of scientific and political targets. Other categories of external stakeholders for CAS, such as the academic and industrial collaborators, have increasing involvement but are still much less active than their German counterparts.

All the three evaluation mechanisms enrich the stakeholder dialogues through various phases of evaluation procedures. Political stakeholders have an increasing involvement by imposing evaluation regulations and standards related to organizational missions. External professionals, internal directors, and individual researchers are engaged at the preparatory phase of evaluation in negotiating guidelines, schedules and composition of review committees where they typically emphasize different evaluation criteria. At the phase of on-site inspection, professional presentation and discussion among peers using technical language can legitimate research investment and facilitate tension-related interaction for seeking possible solutions. At the next phase of preliminary evaluation results, the evaluated have a chance to defend themselves and exchange feedback with stakeholders.

At the phase of final results, a balance between confidentiality of evaluation results and justification for external stakeholders is often realized by using different versions of the results with different degrees of confidentiality. The dissemination of the different versions is carefully undertaken by the three PRIs to make sure information with differentiated confidentiality degrees are transferred in an appropriate manner to the corresponding stakeholders.

In both German cases, evaluation rules are openly negotiated by various Senate members involving internal members and external experts from diverse backgrounds. Such a dialogue platform following policy goals in a more top-down way can justify public investment and emphasize organizational alignment with national policies. The bottom-up communication mostly takes place during the on-site evaluation and negotiation of evaluation results. With regards to guideline-making or result negotiation, stakeholders' interaction across the organizational levels, both inside and outside of the PRIs, are found more often in MPG and HGF. But increasing involvement of international experts in the new evaluation mechanism of CAS gives rise to more frequent and deeper interaction between CAS scientists and their international peers.

The question remains whether these two conditions for the balancing process can be only addressed by evaluation rather than other governance tools. The advantages of evaluation taking such a balancing role at PRIs are as follows: first, the internal stakeholders of research institutions get used to peer-review culture and easily accept evaluations as necessary and regular events where evaluation principles, procedures, and cultures can be widely diffused; and secondly there is a higher cost-effectiveness for the PRI to make use of rich evaluation evidence for decision-making since those activities are already carefully organized and invested.

Ideal governance balances expected

One of the significant outcomes from the empirical studies of the three evaluation mechanisms that meet the conditions of good practice is to mitigate the governance tensions in different contexts. In the next section, we present an open-ended discussion on some generalized 'ideal-types' (George and Bennett 2005) of governance balances at the three organizational levels expected by stakeholders.

Governance balance maintenance at L1

Across the three cases, the overall organizational objective at L1 is to maintain organizational governance with two primary types of balance. The first is to empower bespoke scientific governance for the units under the same organizational umbrella, while keeping them under standardized top-down administration. The former relies on the credibility of their individual research evaluations and the latter can be guaranteed by strategic evaluations of multiple institutes or collaborated programmes across institutes. Such a balance between scientifically tailored and administratively standardized governance at L1 helps with the challenge of governing the heterogeneous units at L2 and L3.

The second type of balance acts between the top-down alignment with organizational missions and the bottom-up scientific adaptation and adjustment. The former is addressed by evaluation guidelines and results concerning justification for top-level stakeholders such as government funders; and the latter by evaluation procedures where the evaluated can defend and adjust themselves in their scientific remit. In other words, the ideal outcome is to achieve centralized organizational strategies and self-managed R&D activities. In the longer term, the more satisfactorily the units behave in both strategic alignment and scientific review in one evaluation round, the more resources they will receive that are beneficial for their autonomy in the next round.

Autonomous development of units (centres/institutes) at L2

At L2, the ideal balance between alignment of organizational strategy and the units' autonomous development can be achieved by the complement between strategic evaluations across the units and the self-organized evaluations of the centres or institutes. The scientific autonomy of the units at L2 across the three cases is generally at a high level because their mainstreamed R&D for the whole PRI is legitimized by evaluations. This in turn strengthens their autonomy to conduct other R&D funded by third parties that is also reviewed by high-level peers.

The tension on funding competition between the units is related to their varying degrees of autonomy. Normally the higher the proportion of block funding, the more autonomy the units have.

A common purpose of evaluation in the two German cases is to justify and guarantee their already high level of block funding (85% for MPG and 70% for HGF). In the Chinese case, the new evaluation mechanism claims to contribute to the policy making to increase 20–30% of the institutes' block funding. With a higher level of block funding and autonomy, the institutes under the PRIs umbrella would collaborate more and compete less with each other, and have more flexibility to apply for external funding, which also relieves the internal competition within PRIs.

Protection of freedom and trust for scientists at L3

All the three PRIs are cultivating a culture where the credibility and acceptability of evaluation protect scientific freedom and trust of the evaluated scientists, outlined as the ideal outcome at L3. It is a mature organizational culture that evaluation promotes mutual trust between public funders and scientists, and also between the headquarters and institutes. A sign that can be checked is whether the proportion of block funding for institutes and scientists stays high or increases further. The higher the proportion of block funding, the more freedom researchers have in terms of the selection of research topic and more flexibility to pursue their own initiatives and goals.

Ideally, the R&D excellence and trustworthiness of the scientists can always be justified and guaranteed by regular and rigorous science evaluations. Rather than feeling monitored, evaluated researchers can be motivated to interact actively with peer evaluators and get useful feedback and learning opportunities both for their work and themselves. The interviewed scientists believe that well-skilled researchers in a trustworthy environment with stable block funding and well-established peer review would concentrate more on their work, which leads to excellent outcomes.

The question arises whether these ideal results are inevitably a result of deploying evaluations under certain methodological conditions. We would suggest that it is much more reasonable to expect these outcomes as an intended achievement but not guaranteed or inherent in the deployment of certain evaluations. The three PRIs studied are successful cases but still a long way away from of achieving the ideal results. Apart from the cases, everything described in this paper rests on the possibility that evaluations may not achieve those ideal results. Yet, we can still discuss practical implications in the next section that could lay the foundations of a better probability of achieving these ideal results, and leave the readers to reflect on the concrete circumstances where evaluations can or cannot achieve those results.

Conclusion and implications

The empirical research in this paper was derived from observed phenomena of interest involving three international, large-scale and scientifically successful PRIs, namely the MPG and the HGF, both in Germany, and the CAS. Unlike universities, MPG, HGF, and CAS are not the subjects of nationally regulated science evaluation systems. They design, fund and use their own evaluation mechanisms, which are peer-review driven and metrics supplemented, for their own internal governance. Their evaluations are found to contribute *de facto* to organizational governance with respect to governance aspects (strategy, funding, and operation) where tensions often occur among different stakeholders. This is not surprising. However, why and how the evaluations contribute to mitigating governance tensions in those concrete cases and at PRIs in general is a question overlooked by the current literature.

The existing literature on evaluation for organizational use also ignores how evaluation should be embedded in organizational governance and lacks systemic consideration of institutional contexts. Little academic discussion can be found that considers practices of world leading PRIs to deal with governance tension via evaluation. This article has made a crucial attempt to attract more attention to the gap where the role of evaluation in governance has a more well-known importance in research institution management practices than research evaluation scholars.

Based on empirical studies of the three PRIs, this article offers an analytical framework (Figure 1) to better understand the two key concepts 'organizational governance' and 'evaluation mechanism', their components, characteristics, and co-evolving relationships with each other. We propose that any other PRI case can be studied following the steps illustrated by our analytical framework. Four key elements of this proposed analytical framework have been discussed.

One, the definition of organizational governance in the context of PRIs includes two components, namely governance aspects as 'forms of government', and 'collective authority of stakeholders' considering their diverging interests in those aspects and power from organizational levels. Therefore, the governance tensions in the PRIs can be described both aspect-wise (strategy, funding, and operation) and level-wise (L1, L2, and L3) from macro-level decision makers, meso-level institutes/centres to micro-level researchers. The three-aspects-to-three-levels matrix, as shown in Table 1, to identify typical governance tensions can be applied for PRIs in general to find out their own problems. These problems that are anticipated to be balanced can be considered at evaluation design phase and observed consistently along with evaluation procedures and results.

Two, the concept of evaluation mechanism is defined from observation on the individual evaluations that are intended to be aggregated in a mechanism so as to respond to the overarching governance requirement. To construct an evaluation mechanism within one single PRI can be considered as an internal control for its organizational management. In the process of being institutionalized, evaluation procedures are professionalized, and the culture on evaluation credibility and acceptability is promoted. Also, the interaction and tension-related negotiation between the various stakeholders are facilitated. These constitute the conditions where evaluation mechanisms can respond to governance requirement. The value of such studies of evaluation roles is that we oppose approaches that treat all peer-review driven evaluations alike. The evaluation culture identified across our cases goes against some studies (e.g. Leisyte, Enders and de Boer, 2010) that show the resistance behaviour of researchers towards institutional changes. We did not observe such resistance due to the high commitment of all the interviewees to the mission-oriented research within their organizations.

Three, the balancing process performs as co-evolution between organizational governance and evaluation mechanism acting dynamically across years or even decades. This is where governance tensions justify evaluation purposes and evaluation procedures and results are systemized to mitigate governance tensions. It is a long process of practicing institutionalization and learning from international peers and own lessons. Our study has shown how the three large and resourceful cases successfully promote such co-evolution in a long-term perspective.

Four, the ideal types of governance balance at the three levels are not inevitable results of deploying evaluation under certain methodological conditions. Instead these are interpreted by us as an intended achievement for the PRI stakeholders. The three PRIs studied are still a long way away from achieving those ideal balances. More case studies are welcome regarding different expectations, whether successfully realized or not, which will together contribute to the development of a more fully grounded strategy of such studies.

To generalize the case findings, a proper contextualization of evaluation design and use in relation to institutional characteristics would be a big concern as well as a challenge. However, we provide practical recommendations for PRIs in general, especially for large ones, if they also expect the above balancing process and results. The primary recommendations are as follows: (1) to justify evaluation purposes with identified or potential governance tensions at an early stage and in a transparent way, such as via printed texts in evaluation guidelines; (2) to systemize individual evaluation activities and make them complement each other at the design phrase with the help of the internal units responsible for evaluation coordination; (3) to institutionalize evaluation procedures and culture with a long-term perspective by developing evaluation guidelines and coordination teams consistently; and (4) to increase the quality and impact of stakeholder interaction in evaluation practices, for instance, by making evaluation data collection and analytical processes more transparent, at least within PRIs. The ideal types of governance balances at the three levels can be self-checked regularly for the long-term development within individual PRIs.

The above recommendations may not be valuable for small PRIs with limited types of evaluation ongoing but will be useful for national policy makers on research evaluation. National science policy making can be more effective when addressing categorized research orientations (fundamental, strategic, and applied) and the bespoke governance of the involved institutions. This is particularly true if the existing science landscape covers all categories and aims to develop comprehensively and diversely, like Germany and China. The complementarity of distinct evaluation evidence from differentiated PRIs can be used for national decision-making, just as the distinct research orientations and evaluations of MPG and HGF together contribute to the German science landscape. That is why China's science policy, as well as that within CAS, has undergone reforms in recent years aiming at a clearer classification of the mixed and fragmented research orientations. The potential role of evaluation mechanisms that can mitigate governance tensions in both clearly categorized and non-categorized science landscape at national or EU-like system level is definitely worth more in-depth research.

Finally, the high cost and possible negative consequences of systemizing evaluations in the three cases are not discussed yet, largely due to the lack of evidence, but these can be researched in any future studies. In addition, future research can explore whether there are some attempts to change the existing balance of PRIs in favour of new ways of evaluation.

Notes

1. EFI (Expertenkommission Forschung und Innovation) was established by the German government and its reports are reliable for understanding the German R&D landscape.
2. Wissenschaftsrat (WR) is funded by the federal and the 16 state governments and has 32 members representing all major stakeholder organizations in Germany.
3. The Harnack Principle was named after Adolf von Harnack (1851–930) who was the first president of the Kaiser Wilhelm Society, the predecessor of MPG.
4. The six fields of the HGF research include Aeronautics, Space and Transport; Earth and Environment; Energy; Health; Key Technologies; and Structure of Matter.

Acknowledgement

We highly appreciate the participation and comments of all the anonymized interviewees and reviewers.

Funding

This work was supported by the China Scholarship Council (CSC) [2011613006].

Conflict of interest statement. None declared.

References

Arnold, E. (2004) 'Evaluating Research and Innovation Policy: A Systems World Needs Systems Evaluations', *Research Evaluation*, 13: 3–17.

Arundel, A. et al. (2007) 'How Europes Economies Learn: A Comparison of Work Organization and Innovation Mode for the EU–15', *Industrial and Corporate Change*, 16: 1175–210.

Bai, C. (2012) 'Reform of CAS S&T Evaluation: Towards a Major R&D Outcome-Orientated System', *Internal Journal of CAS, 27 (4), Speical Issue Innovaion*, 2020: 407–10.

Bryson, J. M., Patton, M. Q., and Bowman, R. A. (2011) 'Working with Evaluation Stakeholders: A Rationale, Step-Wise Approach and Toolkit', *Evaluation and Programme Planning*, 34: 1–12.

Chinese Academy of Sciences (2012) *KIP Evaluation Report: Experiences of China's National Innovation System*. Beijing, China: Science Press.

Chinese Academy of Sciences (2014) *Annual Report 2014* <http://english.cas.cn/about_us/reports/> accessed 26 March 2017.

Chinese Academy of Sciences Evaluation Centre (2013) *An Introduction to Expert Diagnosis Assessments for CAS Institutes*. Unpublished report, CAS, Beijing, China.

Cozzens, S. (2007) 'Death by Peer Review', in Whitley R. and Gläser J. (eds.) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, pp. 225–42. Sociology of the Sciences Yearbook. Springer Nature: Dordrecht, The Netherlands.

Cyranoski, D. (2014) 'Chinese Science Gets Mass Transformation', *Nature*, 513: 468–9.

Dahler–Larsen, P. (2011) *The Evaluation Society*. Stanford, CA: Stanford University Press.

De Jong, S. et al. (2014) 'Understanding Societal Impact through Productive Interactions: ICT Research as a Case', *Research Evaluation*, 23: 89–102.

Eden, C., and Ackermann, F. (1998) *Making Strategy the Journey: The Journey of Strateaic Management*. London: Sage Publications.

Edler, J. et al. (2010) *INNO-Appraisal: Understanding Evaluation of Innovation Policy in Europe*. Brussels and Manchester: European Commision, DG Enterprise.

Edler, J. et al. (2012) 'The Practice of Evaluation in Innovation Policy in Europe', *Research Evaluation*, 21: 167–82.

EFI (Expertenkommission Forschung und Innovation) (2012; 2015) *Research Innovation and Technological Performance in Germany*. Berlin: Germany.

Freeman, R. E. (1984) *Strateaic Management: A Stakeholder Approach*. Boston, MA: Pinnan.

Gazlig, T. (2009) *The Strategy of the Helmholtz Association: Top-Level Research for Society, Science and the Economy*. HGF: Berlin, Germany.

Gazlig, T. (2012) *Helmholtz–Roadmap for Research Infrastructures*. HGF: Berlin, Germany.

George, A., and Bennett, A. (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA, USA: Harvard University.

Golden–Biddle, K., and Rao, H. (1997) 'Breaches in the Boardroom: Organizational Identity and Conflicts of Commitment in a Nonprofit Organization', *Organization Science*, 8: 593–611.

Guerrieri, P., and Tylecote, A. (1997) 'Interindustry Differences in Technical Change and National Patterns of Technological Accumulation', in Edquist C. (ed.) *Systems of Innovation: Technologies, Institutions and Organizations*, pp. 107–25. London: Pinter Publishers/Cassell Academic.

Hage, J., Jordan, G., and Mote, J. (2007) 'A Theory-Based Innovation Systems Framework for Evaluating Diverse Portfolios of Research, Part Two: Macro Indicators and Policy Interventions', *Science and Public Policy*, 34: 731–41.

Hansen, H. F. (2013) 'Systemic Evaluation Governance: New Logics in the Development of Organizational Fields', *Scandinavian Journal of Public Administration*, 16: 47–64.

Hansson, F. (2006) 'Organizational Use of Evaluations: Governance and Control in Research Evaluation', *Evaluation*, 12: 159–78.

Heinze, T., and Kuhlmann, S. (2008) 'Across Institutional Boundaries? Research Collaboration in German Public Sector Nanoscience', *Research Policy*, 37/5: 888–99.

Hellstern, G. M. (1986) 'Assessing Evaluation Research', in Kaufmann F.-X., Majone G., Ostrom V. and Wirth W. (eds) *Guidance, Control and Evaluation in the Public Sector*. Berlin: Walter de Gruyter.

Helmholtz (2014) *Annual Report 2013: Helmholtz-Research for Change* <https://www.helmholtz.de/fileadmin/user_upload/2013_AnnualReport_HelmholtzAssoication_EN_web.pdf> accessed 26 March 2017.

Hermanson, D. R., and Rittenberg, L. E. (2003) 'Internal Audit and Organizational Governance', in Bailey A. D., Audrey A., and Sridhar R. (eds) *Research Opportunities in Internal Auditing*, pp. 25–71. Altamonte Springs, FL: The Institute of Internal Auditors Research Foundation.

Højlund, S. (2014) 'Evaluation Use in the Organizational Context—Changing Focus to Improve Theory', *Evaluation*, 20/1: 26–43.

Jacob, S., Speer, S., and Furubo, J. E. (2015) 'The Institutionalization of Evaluation Matters: Updating the International Atlas of Evaluation 10 Years Later', *Evaluation*, 21: 6–31.

Jansen, D. (2007) *New Forms of Governance in Research Organizations. Disciplinary Approaches, Interfaces and Integration*. Dordrecht: Springer.

Kuhlmann, S. (1999) 'Moderation of Policy-Making? Science and Technology Policy Evaluaiton beyond Impact Measurement—The Case of Germany', *Evaluation*, 4: 130–48.

Kuhlmann, S. (2003) 'Evaluation of Research and Innovation Policies: A Discussion of Trends with Examples from Germany', *Technology Management*, 26: 131–49.

Kupferschmidt, K., and Vogel, G. (2014) 'Doing the Math in Berlin', *Science*, 344: 791–2.

Leisyte, L., Enders, J., and de Boer, H. (2010) 'Mediating Problem Choice: Academic Researchers' Responses to Changes in Their Institutional Environment', in Whitley R., Gläser J., and Engwall L. (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 266–90. Oxford, GB: Oxford University Press.

Luo, J. (2016) 'The Balancing Role of Evaluation Mechanisms: Cases of Publicly Funded Research Institutions: MPG, HGF, and CAS', PhD thesis, University of Twente, Enschede, the Netherlands. <https://doi.org/10.3990/1.9789036541213>.

Luo, J., Ordóñez–Matamoros, G., and Kuhlmann, S. (2015) 'Aggregated Governance by R&D Evaluation Mechanism—Case Study of Chinese Academy of Sciences', *Asian Research Policy*, 6: 56–72.

Mark, M. M., Henry, G. T., and Julnes, G. (1999) 'Towards an Integrative Framework for Evaluation Practice', *American Journal of Evaluation*, 20: 177–98.

Max Planck Society (2013) *Annual Reports of the Max Planck Society* <https://www.mpg.de/annual-report> accessed 26 March 2017.

Max Planck Society (2015) *Rules for Scientific Advisory Boards* <https://www.mpg.de/197429/rulesScientificAdvisoryBoards.pdf>accessed 26 March 2017.

Moura, H. M., and Teixeira, J. C. (2010) 'Managing Stakeholders Conflicts', in Chinyio E. and Olomolaiye P. (eds) *Construction Stakeholder Management*, pp. 286–314. Oxford: Wiley-Blackwell.

Nature Publishing Index (2014) 'Top 200 Institutions', *Nature*, 515: 98–108.

OECD (2013) *Principles for the Governance of Regulators Public Consultation Draft*. Paris: OECD Publishing.

Peters, B. G., and van Nispen, F. K. N. (eds). (1998) *Public policy instruments: Evaluating the tools of public administration*. Cheltenham: Edward Elgar Press.

Porter, M. (1996) 'What is Strategy?', *Harvard Business Review*, 74: 61–78.

Raina, R. S. (2003) 'Disciplines, Institutions and Organizations: Impact Assessments in Context', *Agricultural Systems*, 78: 185–211.

Sanders, B. A. (2003) 'Maybe There's No Such Thing as a "Good Cop": Organizational Challenges in Selecting Quality Officers', *Policing: An International Journal of Police Strategies & Management*, 26: 313–28.

Schiene, C., and Schimank, U. (2007) 'Research Evaluation as Organizational Development', in Whitley R. and Gläser J. (eds) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, pp. 171–88. *Sociology of the Sciences Yearbook*. Dordrecht, The Netherlands: Springer Nature.

Schruff, H. (2012) *Evaluation Procedures of the Max Planck Society* < http://www.cnr.it/sitocnr/UPO/documenti/ep2223ott.pdf> accessed 26 March 2017.

Simon, D., and Knie, A. (2013) 'Can Evaluation Contribute to the Organizational Development of Academic Institutions? an International Comparison', *Evaluation*, 19: 402–18.

Sun, Y. T., and Cao, C. (2014) 'Demystifying Central Government R&D Spending in China', *Science*, 345: 1006–8.

Van der Knaap, P. (2006) 'Responsive Evaluation and Performance Management: Overcoming the Downsides of Policy Objectives and Performance Indicators', *Evaluation*, 12: 278–93.

Whitley, R., and Gläser, J. (2007) 'The Changing Governance of the Sciences: The Advent of Research Evaluation Systems'. *Sociology of the Sciences Yearbook*. Dordrecht, The Netherlands: Springer Nature.

Wilsdon, J. et al. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, HEFCE, UK. DOI: 10.13140/RG.2.1.4929.1363.

Wissenschaftsrat Evaluation Report (2001) *Statement of the German Science Council on the Hermann von Helmholtz Association of German Research Centres*. Köln, Germany <http://www.wissenschaftsrat.de/download/archiv/4755-engl.pdf> accessed 26 March 2017.