

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Perkmann, Markus; Schildt, Henri

## Open data partnerships between firms and universities: The role of boundary organizations

*Published in:*  
Research Policy

*DOI:*  
[10.1016/j.respol.2014.12.006](https://doi.org/10.1016/j.respol.2014.12.006)

Published: 01/01/2015

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Perkmann, M., & Schildt, H. (2015). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, 44(5), 1133-1143. <https://doi.org/10.1016/j.respol.2014.12.006>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



# Open data partnerships between firms and universities: The role of boundary organizations



Markus Perkmann<sup>a,\*</sup>, Henri Schildt<sup>b,1</sup>

<sup>a</sup> Imperial College London, Business School, London SW7 2AZ, United Kingdom

<sup>b</sup> Aalto University School of Business, P.O. Box 21210, FI-00076 Aalto (Helsinki), Finland

## ARTICLE INFO

### Article history:

Received 7 January 2014

Received in revised form 8 December 2014

Accepted 14 December 2014

Available online 26 December 2014

### Keywords:

Open data

Industrial R&D

Selective revealing

Boundary organization

University–industry relations

Open innovation

Research partnership

## ABSTRACT

Science-intensive firms are experimenting with ‘open data’ initiatives, involving collaboration with academic scientists whereby all results are published with no restriction. Firms seeking to benefit from open data face two key challenges: revealing R&D problems may leak valuable information to competitors, and academic scientists may lack motivation to address problems posed by firms. We explore how firms overcome these challenges through an inductive study of the Structural Genomics Consortium. We find that the operation of the consortium as a boundary organization provided two core mechanisms to address the above challenges. First, through mediated revealing, the boundary organization allowed firms to disclose R&D problems while minimizing adverse competitive consequences. Second, by enabling multiple goals the boundary organization increased the attractiveness of industry-informed agendas for academic scientists. We work our results into a grounded model of boundary organizations as a vehicle for open data initiatives. Our study contributes to research on public–private research partnerships, knowledge revealing and boundary organizations.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

‘All human genomic sequence information (...) should be freely available and in the public domain in order to encourage research and development and to maximise its benefit to society’ (Human Genome Project, 1996).

## 1. Introduction

The above quote expresses the ‘open data’ rule that constituted a cornerstone of the Human Genome Project. The disclosure regime of this large-scale research programme was built on the principle of free, unrestricted and timely access to research findings for all interested parties (Murray-Rust, 2008; Molloy, 2011). In the Human Genome Project, public science was pitched against for-profit entities with competing projects based on proprietary intellectual property (Williams, 2010). Yet increasingly firms themselves participate in and even instigate open data initiatives, either by releasing data to academic communities with no restriction or by supporting the generation of open data. Partnerships sponsored

by pharmaceutical companies, such as the SNP<sup>2</sup> consortium and the Genetic Association Information Network (GAIN) have made their data publicly available (Cook-Deegan, 2007; Pincock, 2007; Allarakhia and Walsh, 2011).

Partnerships with universities, aided by public or charity grants, are natural territory for open data practices, given the prominence that public knowledge creation has in the norms and traditions of academic science (Dasgupta and David, 1994). The propagators of open data in corporate R&D argue that by integrating their R&D programmes more closely with those of open academic communities, firms may reap significant benefits for both the quality and the volume of their innovation activity (Melese et al., 2009).

Nevertheless, participation in open data partnerships with universities is likely to complicate firms’ attempts to capture value from research. A first challenge is that firms may fear that proprietary information about their R&D agendas and technologies is publicly disclosed (Alexy et al., 2013), given that open data initiatives operate with minimum intellectual property protection and disclose all research results with no restriction. The second challenge, from a firm’s viewpoint, is to motivate outsiders to work on problems that are valuable to the firm, without being able to

\* Corresponding author. Tel.: +44 207 594 1955.

E-mail addresses: [m.perkmann@imperial.ac.uk](mailto:m.perkmann@imperial.ac.uk) (M. Perkmann),

[henri.schildt@aalto.fi](mailto:henri.schildt@aalto.fi) (H. Schildt).

<sup>1</sup> Tel.: +358 50 413 9442.

<sup>2</sup> SNPs are ‘single nucleotide polymorphisms’. They indicate possible mutations of a gene, and can be used as disease markers.

offer IP-related incentives (von Hippel and von Krogh, 2006; Levine and Prietula, 2014). In other words, in open data initiatives which, unlike traditional firm-sponsored contract research, are strongly aligned with academic conventions, firms may struggle to persuade scientists to work on firm-defined priorities rather than their own personal research agendas.

Extant research provides limited insight into how firms can address these challenges. The literature on research partnerships between firms and universities is largely focused on contexts with traditional, IP-centred appropriation mechanisms in place (Link and Scott, 2005; Bercovitz and Feldman, 2007) but says little about how open data partnerships ought to be structured and governed.<sup>3</sup> In this paper, we therefore address the following research question: what partnership characteristics enable firms to benefit from open data collaboration with academic researchers?

To explore how firms overcome the challenges of open data initiatives, we examined the structures and practices of an international life sciences partnership. We present an inductive study of the Structural Genomics Consortium (SGC) which led an open data programme involving firms and academic scientists. Supported by charity, government and industry funding, the SGC brought together pharmaceutical firms including GlaxoSmithKline, Novartis and Merck, with the Universities of Toronto and Oxford, and the Karolinska Institutet (Stockholm). The SGC's mandate was to determine the three-dimensional shape of proteins and release this knowledge into the public domain without restriction. This information is seen as vital to the discovery of new drugs to combat common human diseases, including cancer, diabetes and inflammation.

We draw on our empirical analysis to develop a grounded model of open data in university–industry partnerships. We propose that open data university–industry partnerships that are structured as boundary organizations (O'Mahony and Bechky, 2008) are particularly adept at generating productive outcomes while mitigating firms' challenges. Boundary organizations accomplish this via two core mechanisms: mediated revealing and the enabling of multiple goals. The former allows firms to reveal their research problems to external problem solvers in a way that reduces the threat of unintended knowledge disclosure and simultaneously allows them to shape the collective research agenda. In turn, by enabling multiple goals – in this case the concurrent pursuit of both industrial and academic goals – the boundary organization broadens the objectives and activities of the partnership so they align with the ambitions and professional practices of academic researchers which in turn helps to ensure their participation.

Our findings contribute to previous work by considering the implications of open data for both the rationales underpinning research partnerships between firms and universities and questions of organization design. In particular, we demonstrate the role that boundary organizations can play in orchestrating industry-informed, large scale scientific work that has the potential to advance and transform the knowledge commons from which science-based sectors draw.

## 2. Open data in university–industry partnerships

Open data partnerships provide universal and free access to research outputs including results, data and sometimes materials (Murray-Rust, 2008; Molloy, 2011). The open data approach is in contrast not only to commercial emphasis on intellectual property rights, but even to classic open science in which only the final

outputs are shared (Boudreau and Lakhani 2015; Franzoni and Sauermann, 2013). Various scientific communities have recently adopted increasing openness, including the free sharing of data on which outputs are based (Reichman et al., 2011).

This development was partly spurred by the increasingly widespread use of computer code and large datasets which makes the large-scale sharing of data both feasible and economical (Boulton et al., 2011). The same technological affordance has facilitated 'crowd science' experiments where problem solving is pursued by a large number of dispersed contributors (Franzoni and Sauermann, 2013). Particularly in the life sciences, a further driver of open data is the trend towards larger scale initiatives designed to address the complex, interconnected nature of biological systems which has tested the limits of the traditional small-scale approach in biology, centred around individual investigators (Swierstra et al., 2013). The Human Genome Project (HGP) absorbed \$3b of funding and used an open data approach to facilitate coordination across thousands of researchers around the world, and the subsequent exploitation of the generated knowledge (Wellcome Trust, 2003). Similarly, the Census of Marine Life project resulted in the Ocean Biogeographic Information System (OBIS) database, the world's largest open access repository of marine life data (Vermeulen et al., 2013).

The sharing of data in areas such as genetics, clinical trials and climate science is supported by various types of stakeholders, including research funding organizations, patient groups, interest groups and not least academic scientists themselves. They argue that open data enables scientific communities to validate and substantiate the results of previous research and thereby enhance its quality, particularly in areas where conflicts of interests are at play such as pharmaceutical research (Washburn, 2008).

Below, we first contrast the new open data approaches with traditional approaches in university–industry collaboration and then outline the specific challenges that open data collaborative initiatives create for for-profit firms.

### 2.1. Research partnerships between firms and universities

Research partnerships are innovation-based relationships focusing on joint research and development (R&D) activities (Hagedoorn et al., 2000). Firms engage in research partnerships because they allow investments in the creation of new knowledge to be shared across multiple participants. They also provide firms with access to complementary knowledge, broaden the scope of their R&D, and create new investment options in high-risk contexts (Hagedoorn et al., 2000; Perkmann et al., 2011). Especially in science-intensive sectors such as chemicals and pharmaceuticals, universities represent important partners and sources of innovation for firms (Mansfield, 1991; Cohen et al., 2002). Firms tend to view university research as complementary (rather than substitutive) to internal R&D (Rosenberg and Nelson, 1994; Hall et al., 2001). Access to key personnel represents an additional important motive for firms to work with academia, resulting both in "information gifts" from highly specialized academics as well as opportunities for hiring students and staff (Hagedoorn et al., 2000).

Partnerships are not without challenges. Chief amongst these is the concern that a firm may struggle in appropriating the knowledge outputs generated in the partnership (Teece, 1986). Compared to inter-firm partnerships, such concerns are even more pronounced in university–industry partnerships (Hagedoorn et al., 2000). There are two aspects to this problem. First, firms' efforts to appropriate knowledge arising from partnerships may be misaligned with open science practice. Academics may prefer generating publishable research output and contest the formal requirements involved in creating protected knowledge assets (Murray, 2010). At the very least, this may lead to an uneasy

<sup>3</sup> The phenomenon we refer to as 'open data' has also been labelled 'open source science' or 'open access research' (Munos, 2006; Edwards, 2008; Gowers and Nielsen, 2009; Hope, 2009; Melese et al., 2009).

co-existence of open publishing and intellectual property protection (Gittelman and Kogut, 2003). Second, partnerships involving universities often attract grants from government or charities. This means that the universities will in most jurisdictions make ownership claims over intellectual property generated (Kenney and Patton, 2009). Regulations such as the Bayh–Dole Act in the United States stipulate that universities can claim IP ownership over outcomes from government funded research (Mowery et al., 2001). In such a context, the higher the share of public funding in a partnership, the more pronounced firms' concerns about appropriation will become (Hall et al., 2001).

Firms respond in several ways to the appropriability challenges pertaining to partnerships with universities (Panagopoulos, 2003). First, firms tend to prefer larger collaborations when public partners are involved. In this case, appropriability has already been diminished because the presence of a larger number of private and public partners not only stipulates the shared ownership of intellectual property, but also increases the risk of unintended knowledge spill-overs (Link and Scott, 2005). Also, participation in larger partnerships carries a reduced cost, and hence implies an improved balance between risks and rewards (Saez et al., 2002). Second, firms are often given the first right of refusal for licencing intellectual property arising from a partnership (Perkmann and West, 2015). Firms can access the results from joint research with conditions that were determined ex-ante, implying a reduction of uncertainty relating to the appropriation of partnership outputs. Third, firms may choose partnerships in 'pre-competitive' areas where intellectual property appropriation is less important than alternative benefits, such as the development of new areas of expertise (Powell et al., 1996).

Extant research on university–industry partnerships has mostly focused on the question of "primary" appropriability, that is the control and ownership of intellectual property created within the partnership (Ahuja et al., 2013). This focus is mirrored in universities' efforts to assert ownership of the outputs from research collaborations (Kenney and Patton, 2009). Against this background, we lack insight into the benefits accruing to firms from partnerships that entirely relinquish intellectual property in the first place. For firms, open data policies pose a conundrum: On the one hand, the absence of intellectual property rights makes it difficult to gain returns to investments, yet on the other hand the sheer scale of these collaborative efforts makes them too important to ignore, particularly if large numbers of scientists are potentially available to work on topics of interest to firms. Next, we discuss the considerations relevant for firms with respect to participation in open data initiatives.

## 2.2. Challenges facing firms in open data research partnerships

Compared to conventional research partnerships, open data partnerships pose two significant problems which may temper firms' motivation to engage in such initiatives. The first is that of revealing; the more a firm attempts to align the efforts in an open data research programme with its R&D priorities, the more it will have to reveal about the problems it is addressing within its proprietary R&D. Revealing has both advantages and disadvantages for firms – they may benefit from revealing problems or solutions as this may allow them to shape the collaborative behaviour of others, and thereby enhance their competitive position (Alexy et al., 2013). Such benefits have been documented for various contexts, including mining during the industrial revolution, 19th century iron production, and contemporary embedded Linux software (Allen, 1983; Nuvolari, 2004; Henkel, 2006). Revealing information about their technologies may also discourage others from competing in the same technology areas (Clarkson and Toh, 2010). Yet, by guiding the academic community to address specific scientific

problems, a firm discloses information about its active R&D areas to its competitors (Arrow, 1971; Cohen et al., 2000). Overall, in an open data scenario, while an excessive degree of 'problem revealing' (Alexy et al., 2013) may lead to imitation by competitors, an insufficient degree of problem revealing may impair the firm's ability to steer the alignment of outside knowledge with its R&D activities.

The second issue facing firms in open data is an incentive problem. How can firms encourage individuals operating within distributed scientific communities to participate in their open data programmes? The success of the open data approach relies on motivating self-organizing groups of scientists to focus their research efforts on topics of interest to the firm in the absence of effective hierarchical control (Murray and O'Mahony, 2007). Since academic scientists are embedded in the academic status hierarchy and career system that differs considerably from private sector R&D, monetary incentives are unlikely to be effective. The primary objective for many participating researchers will be to improve their standing and position in their chosen academic community, even at the expense of pursuing commercially valuable opportunities or personal monetary gains (D'Este and Perkmann, 2011). An open data initiative will have to provide suitable incentives that are aligned with academic scientists' desire to be rewarded for their work within their respective communities.

Having outlined the challenges for firms arising in open data partnerships, in this study we will explore how they should be organized to enable firms to address the challenges while garnering benefits from the partnership.

## 3. Data and methods

### 3.1. Site: the Structural Genomics Consortium

We studied the Structural Genomics Consortium, a major initiative with laboratories at the Universities of Oxford and Toronto, and the Karolinska Institutet (Stockholm), established in 2004. The consortium was funded by the Wellcome Trust, pharmaceutical companies GlaxoSmithKline, Novartis and Merck, government organizations, and several smaller foundations. The Wellcome Trust is one the world's largest medical research foundations, and the participating firms were in first, third, and sixth position respectively, for global market share of prescription drugs in 2012.<sup>4</sup> The SGC's objective was to identify the three-dimensional shape of thousands of human proteins with potential relevance for drug discovery. The physical shape of proteins affects how they interact with other molecules in the human body. Thus, knowledge of proteins' structural characteristics can aid the discovery of new drugs and exploration of the molecular mechanisms that underpin them.

The pharmaceutical industry provides an ideal setting to study how firms implement open data initiatives and overcome their challenges. Because in this sector proprietary intellectual property has traditionally played a strong role, potential tensions arising from open data were likely to be particularly accentuated.

### 3.2. Data collection

We used an inductive, qualitative approach to study the SGC which is suitable for the in-depth exploration of phenomena that are not well understood (Eisenhardt, 1989). Data collection involved studying archival documents, interviewing and observation. Our archival documents were drawn from the official minutes

<sup>4</sup> Source: IMS Health.

**Table 1**  
Representative quotes grouped according to second-order codes.

|  |  |
|--|--|
| Enabling firms' influence on research agenda | <p>'The real benefit to industry is the ability to nominate targets. That's the benefit that is unique to funders' [pharma sponsor]</p> <p>'What the SGC does is the sort of work I would love to have done in my department, but I knew that it was much earlier phase in the drug discovery process, and it's much higher risk, and there's much more of it to do. We couldn't fund this work internally, but I wanted the results of the SGC because I would then be able to pick up particular items of interest and then use them internally' [pharma sponsor]</p>  |
| Maintaining confidentiality                  | <p>'It was noted that Board members should maintain confidentiality between the SGC and Funders' [board member]</p> <p>'It was noted that the issue of confidentiality of the target list is a significant issue for us' [pharma sponsor]</p> <p>'Knowing that the structure is solved means we can reproduce it and we have a better starting point. The fact that the SGC keeps this to a very small number of people means we have zero risk, essentially, that our list would go to a competitor' [Pharma sponsor]</p> <p>'The SGC maintains the Target List as confidential information; neither consortium members nor the public are aware of the proteins on the Target List' [SGC official communication].</p>  |
| Promoting academic goals                     | <p>'The SGC Oxford asked the Scientific Committee for a recommendation to the Board for publication of the draft manuscript by Smith et al. This would reveal the identity and screening data of a number of kinases (...). The Scientific Committee agreed' [archival].</p> <p>'The Scientific Committee supports the request to the Board for approval to release for publication, an article containing information regarding protein kinase targets whose structures have not been solved. The Scientific Committee considered that the benefits of releasing the information far out-weighted the likely impact of releasing such information.' [archival]</p> <p>'There's some latitude, and the groups themselves have some capability to do their own research. ... [Targets] are not prescribed down to "these have to all have to be done and none others". ... There's some latitude that can allow both scientific curiosity' [SGC management].</p> <p>'Even though the primary focus of the SGC is to deliver structures, there is still an expectation by the community that the SGC should publish in peer reviewed journals'. [SGC management]</p> |
| Adopting academic practices                  | <p>'SGC scientists are all employees of the University of Toronto. A majority of the principal investigators have applied for what's called <i>status only</i> appointment in some of the academic departments and that allows them to apply for other grants if they wish to and participate in the academic life of the department (...). It's a great way to keep a foot in those doors, if you will, and a foot in both camps'. [scientist].</p> <p>'When people collaborate with us or come to study the project, I think they're impressed that we're doing a good job and that's why we have so many [external academic] collaborators'. [SGC scientist]</p>  |

of 16 meetings of key SGC bodies, including the Board of Directors, held between 2005 and 2007 (see [Appendix 1](#)). The minutes provide records of the organization's activities and decisions but also allowed the inference of more subjective agendas and interests of various participants. We also perused additional SGC documents including the Memoranda on Articles of Association, the Funding Agreement, annual reports, press communications and presentations. The total word length of all documents is approximately 100,000.

We further conducted 22 semi-structured interviews with SGC staff, members of the board and the scientific committee, senior management, and scientists. The interviews covered more than half of the individuals involved in the governance and management of the SGC, a sample of the researchers, and an external observer. We asked informants to provide us with their version of the SGC's origins and history as well as their own motives, objectives and role within the consortium. We also requested they describe key organizational processes, specifically those relating to aspects of revealing and motivation. All but four interviews were recorded and transcribed verbatim (see [Appendix 2](#)). Triangulation with archival meeting minutes allowed us to control for potential self-reporting and retrospective bias in the interview evidence.

A third set of data was based on observations and informal conversations in London and Toronto, and by phone, between 2007 and 2011. The informal discussions were with SGC managers, sponsors' representatives, and external observers, including some critics. The first author attended three SGC workshop held in 2007 and 2011, and had numerous informal conversations with participants as well as outsiders. After each interview or conversation, we created a memo, summarizing insights and exploring avenues for theorizing. We sought to obtain external validity by triangulating information across multiple sources, spanning insiders and outsiders.

### 3.3. Data analysis

Our inductive analysis proceeded in several steps. We first generated a case narrative, depicting the SGC's operating context, the organization's development, and its structures and practices. We used this account to generate a 3200-word report which we sent to all interviewees. Two respondents provided detailed feedback and corrected factual mistakes while others provided cursory feedback.

Using the qualitative data analysis software, NVivo, we conducted an initial round of first-order (open) coding ([Corbin and Strauss, 2008](#)) on all archival documents, interview transcripts and memos. Guided by our research question, we coded the activities of the SGC with respect to how these addressed the key challenges associated with open data. Examples for first-order codes included 'pharma members' ability to nominate targets' and 'making allowance for scientific curiosity' (see sample extracts in [Table 1](#)). We validated codes by ensuring they emerged from multiple instances, otherwise we discarded them.

We next moved to second-order (axial) coding and established relationships between the open codes by searching for connections between them. For instance, we grouped the first-order codes 'pharma members' ability to nominate targets', and 'pharma members shape SGC strategy to focus on targets relevant for drug discovery' to form the second-order category of 'enabling firms' influence on research agenda'. Throughout, we constantly moved backwards and forwards between our evidence and the emerging categories, helping us to render our results as robust as possible. Our final step was to work our second-order codes that were still fairly close to the phenomenon into a grounded theory model that abstracts from the specificities of our case, and posits theoretical mechanisms potentially applicable to a wider range of empirical situations.



Below, we first present our raw findings, reflecting the results of our first-order and second-order coding exercise, before presenting our grounded model in the subsequent section.

## 4. Findings

### 4.1. History and features of the Structural Genomics Consortium

During the 2000s, there was an increasing recognition in the pharmaceutical industry that its research productivity was slowing. Despite escalating R&D expenditure, the number of novel drugs failed to rise proportionately (Paul et al., 2010). ‘Big Pharma’ responded by reducing R&D expenditure and engaging in external collaboration (Garnier, 2008; Schuhmacher et al., 2013). In particular, public–private partnerships appeared attractive as many industry insiders believed that by relying on public science they could reduce the high failure rates in drug development (Munos, 2009).

The SGC was designed in response to this need. The organization, founded in 2003, originated from interactions between GlaxoSmithKline scientists and officials at the Wellcome Trust. Similar to the Human Genome Project, the research initiative would involve academic researchers and attract funding from foundations and the State. With funding from the Wellcome Trust, GlaxoSmithKline and the Canadian and UK governments, the SGC initially operated laboratories at the Universities of Toronto and Oxford. From the viewpoint of the charity and government funding organizations, the initiative was to aid the expansion of the knowledge commons in the pharmaceutical sector. The creation of new, publicly shared knowledge was expected to underpin the discovery of new drugs and thereby enhance pharmaceutical companies’ innovativeness. Aled Edwards, a Toronto-based academic scientist with prior involvement in several biotechnology start-ups, was nominated as CEO of the SGC. The organization was governed by a board of directors of some 15 individuals who represented the sponsors. A separate scientific committee of ten scientists comprising both independent academics and sponsor representatives oversaw all scientific decisions.

In 2005, the SGC expanded its activities by establishing a laboratory at Karolinska Institutet in Stockholm as a group of Swedish funders joined as additional sponsors.<sup>5</sup> The pharmaceutical companies, Novartis and Merck also joined as sponsors of the consortium. With an annual turnover of around CAD 30m at the time of study, the organization employed approximately 180 staff with each site hosting several teams led by principal investigators and a chief scientist.

The consortium practiced an open data approach and deposited all newly resolved protein structures in the open Protein Data Bank (PDB), with no advance access given to consortium members. In loose analogy with open source software, open data was to facilitate a process whereby self-motivated innovators could freely build on the work of others (Edwards, 2008). The funders of the consortium believed that this approach would accelerate the collective creation of knowledge underpinning the discovery of new drugs. Edwards warned that ‘the predominant methods of drug research are too patent heavy, leading to duplicated effort and lost opportunities for significant productivity (...) Intellectual property

is killing the process of drug discovery’ (SGC press communication). Whilst close interaction with academia had been common in the pharmaceutical industry for many years (Cockburn et al., 1999), the SGC model differed from traditional collaboration: participants were committed to relinquishing intellectual property ownership; collaboration among competing pharmaceutical firms was emphasized; and research funds were sourced internationally via an organization that maintained flexible ties with universities.

In Table 2, we provide a summary of the interests held by each type of consortium participant, and their perceived benefits. While the table illustrates that the various parties’ agendas partly diverged, it also shows that each derived benefits from participation.

We use the term ‘sponsors’ to refer to organizations that provided funding to the SGC. We use the term ‘participants’ to refer to individuals involved in the SGC, including its lead scientists (CEO and chief scientists), members of the board of directors (the ‘directors’), members of the scientific committee and the scientists working for the SGC. When we speak of the SGC as an organizational actor, we refer to the collective actions of its principal officers (the lead scientists), the directors and the scientific committee members.

### 4.2. Revealing and confidentiality

As part of our investigation, we explored how the pharmaceutical companies aligned SGC’s research with their own proprietary R&D activities and the implications this had for their revealing behaviour. The challenge was to steer the SGC’s work towards areas of importance to a firm but simultaneously avoid detailed problem revealing to the extent that it would benefit competitors (Alexy et al., 2013). Within the SGC, this challenge was addressed in two ways. First, the SGC created procedures through which the participating firms could influence its work programme. Second, the SGC designed these procedures in a way that restricted information spill-overs.

#### 4.2.1. Enabling firms’ influence on the research agenda

The consortium designed a decision making process that enabled the pharmaceutical companies, like the other sponsors, to shape the organization’s research programme. The process entailed compiling a list of proteins to be resolved by the SGC scientists. Every sponsor was allowed to submit a ‘wish list’ of 200 targets of which 20 could be designated as priorities. The nominations were examined by the scientific committee, which produced a master list of proposed targets for final approval by the board. At the time of our study, the list included a few thousand proteins.

For the pharma companies, the ability to shape the research agenda was a critical objective. They were keen to focus the SGC’s work on proteins that were likely to be relevant for human health, rather than those that were of more general scientific interest. The focus on such ‘human targets’ was attractive because of their potential for informing the development of new drugs.

#### 4.2.2. Maintaining confidentiality

The wish lists proposed by the SGC’s pharma members contained those proteins regarded as important for their R&D activities. Revealing these lists openly may have allowed competitors to infer a company’s R&D priorities. So, despite its insistence on openness in many other respects, the SGC kept these lists confidential even from the board of directors and scientific committee; in this way, a sponsor’s interest in a particular protein was never revealed to another sponsor. The office of the CEO combined the individual wish lists into an anonymous ‘master list’ that was circulated among the management, board of directors and the scientific committee.

<sup>5</sup> The full list of sponsors at the time of our study includes: The Canada Foundation for Innovation (CFI), The Canadian Institutes of Health Research (CIHR), Genome Canada, GlaxoSmithKline plc. (GSK), Karolinska Institutet, The Knut and Alice Wallenberg Foundation, Merck & Co., Inc., Novartis, The Ontario Genomics Institute (OGI), The Ontario Ministry of Research and Innovation (MRI), The Swedish Foundation for Strategic Research, VINNOVA (The Swedish Governmental Agency for Innovation Systems), The Wellcome Trust. The Wellcome Trust was the largest contributor of funding, and the pharmaceutical firms contributed approx. 5% each.

**Table 2**  
Summary of SGC participants' interests and perceived benefits.

| Participants                    | Interests and goals  | Benefits from SGC   |
|---------------------------------|--|---|
| Firms                           | Exploitation of scientific knowledge for development of commercially viable drugs<br>Cost-effective discovery of new drugs | Influence on generation of public knowledge on proteins relevant to drug discovery<br>Access to subsidies from charities and governments                  |
| Academics                       | Peer recognition and career advancement through peer-reviewed articles   | Participation in state-of-the-art programme on previously uncharacterized proteins<br>Opportunities for conducting follow-on research on these proteins   |
| Charitable research foundations | Advancement of human health by enabling projects not supported by the market   | Development of knowledge commons underpinning drug discovery and thus human health<br>Solicit industry input on promising directions of work              |
| Government                      | Support for science with high economic and social impact   | Generate basis for wide-ranging follow-on research supporting human wellbeing and economic growth<br>Development of scientific talent in national economy |

As an additional safeguard, the master list was not publicly disclosed. Only when a target protein was resolved was its structural information openly deposited, but the identity of proteins that the consortium failed to resolve was never disclosed. The confidentiality requirements also extended to parties collaborating with SGC scientists and external collaborators were required to sign a confidentiality agreement. The confidentiality formula was regarded as a decisive benefit by the private-sector sponsors, particularly with respect to a small number of high-priority targets. The presence of this rule meant that pharma members were prepared to entrust the consortium with more of their sensitive high priority targets, than had the wish lists or the master list been publicly disclosed.

The confidentiality formula was maintained even though it was criticized by some academic outsiders. Because it was not known which proteins the SGC was working on, these individuals feared that they may be expending parallel effort on resolving the same proteins and that the SGC would likely succeed in doing so first because of its economies of scale. Even in the face of such criticism, the SGC leadership maintained that this trade-off was necessary in order to maintain the partnership with corporate sponsors. Exceptions to the confidentiality formula were only made when the public interest overrode confidentiality concerns, as for instance when targets related to malaria were prioritized for global health impact, or when the information was an essential part of journal articles to be imminently published.

#### 4.3. Motivating academic researchers

The SGC pursued several strategies for attracting academic scientists to participate in its endeavour. Some of these were aimed at the SGC-internal scientist community while others were tailored for engagement with external scientists.

##### 4.3.1. Promoting academic goals

The SGC encouraged its academic workforce to engage in research beyond mapping the proteins on the master list. Such activities were not always aligned with sponsoring firms' interests, but facilitated the career progression of participants and increased the prominence of SGC as a research institution.

First, SGC scientists were encouraged to pursue 'follow-on' research on the characterized proteins and publish results in peer-reviewed articles. This meant studying how they linked to and reacted with other molecules, such as inhibitors. The investigation of these mechanisms was seen in the academic community as more demanding and interesting, than resolving the protein structures per se. On one occasion, the SGC leadership reduced the amount of proteins to be resolved by 15% each quarter. This measure was enacted to 'enable [the SGC staff] to utilize their intellects fully' by increasing the time at their disposal to pursue personal research programmes, and thus achieve 'higher overall scientific impact' of

the SGC (d4). This was thought to improve staff retention because it enabled the researchers to publish more high impact articles and thereby improve their career prospects in academia.

Second, the SGC allowed the scientists to tackle proteins that were not on the master list, even though the SGC had solicited suggestions from the academic community when it compiled the list. The freedom to explore structures outside the master list was granted when it promised to add 'significantly to scientific understanding' with a prospect of academic publications. At one site, 17 of its 114 resolved structures were outside the master list and had been chosen by the researchers themselves. These structures were of less immediate interest to drug discovery but of more general scientific relevance. In effect, the SGC provided its scientists with organizational slack so they could pursue their scientific curiosity, leading not only to publications, but also to increased knowledge and capabilities.

##### 4.3.2. Adopting academic practices

The SGC sought to emulate core features of academic environments even though strictly speaking it was autonomous from academia. The SGC did not employ researchers directly but disbursed funds to universities so they could in turn employ the researchers using terms and conditions familiar in academic contexts. The SGC also sponsored academic activities, such as seminars and visiting scholarships, it actively encouraged collaboration between its scientists and other researchers at its own university sites and in other universities (i16). In September 2009, SGC Oxford had 70 collaborating researchers and SGC Toronto had 37 active collaborations (d23). The SGC also ensured its staff could obtain honorary appointments within the respective departments at the universities hosting the SGC laboratories. The academic outlook of the organization was reinforced by the fact that while industrial sponsors shaped the overall work agenda through their boardroom representation, they had minimal impact on the day-to-day pursuit of the research.

Overall, the design of the SGC as an autonomous organization meant that it was able to provide a work context that differed little from traditional academic settings. By using these practices, the SGC could attract high calibre researchers who used their employment to underpin a career in mainstream academia. The important overall insight is that the SGC achieved this by broadening the agenda from a pure industrial focus to accommodate both industrial and academic goals.

## 5. A model of open data partnerships between firms and universities

The Structural Genomics Consortium shares many features with organizations characterized in previous literature as boundary organizations. These stand between parties with divergent

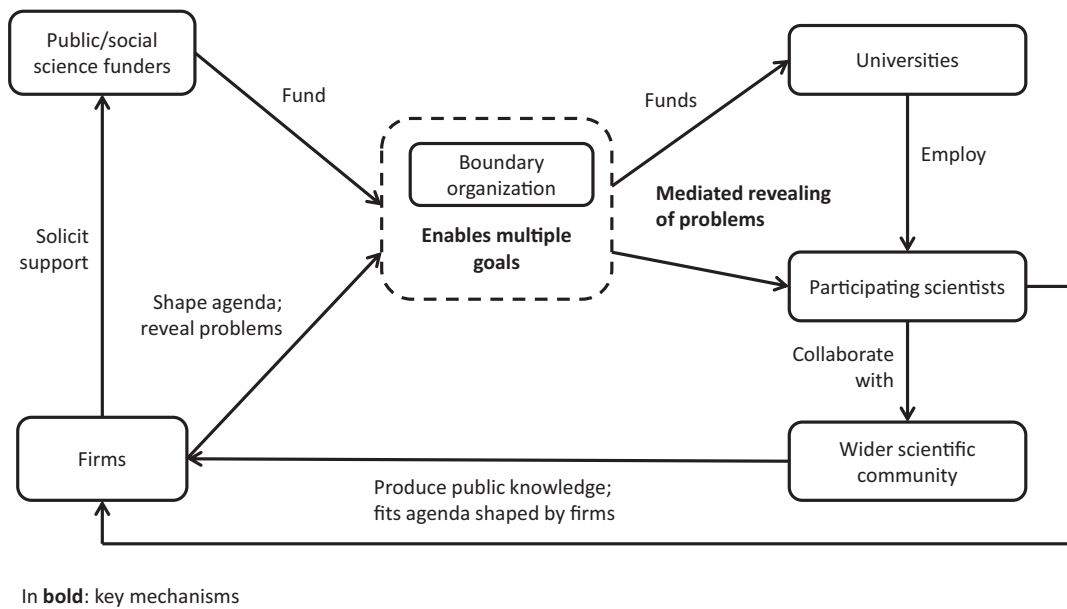


Fig. 1. Model of open data in university–industry partnerships.

interests and allow them to collaborate (Guston, 2001; O'Mahony and Bechky, 2008; Miller, 2001; Howells, 2006). They provide solutions to governance problems in situations where parties with different interests interact (O'Mahony and Bechky, 2008). In the next section, we use our findings to develop a model of boundary organizations in the specific context of open data partnerships.

#### 5.1. Mechanisms enabled by the boundary organization

Generalizing our findings, we identify two key mechanisms by which boundary organizations enable open data partnerships (see Fig. 1). The first is 'mediated revealing', which involves an intermediary to aggregate and anonymize information before it is passed on to a different party. The SGC aggregated firms' wish lists of proteins, and compiled a master list for the scientists that did not disclose which target was nominated by whom. Through this mechanism the firms fundamentally shaped the direction of work pursued by the academic researchers in the consortium and their external collaborators', without revealing which proteins specifically they were interested in. Previous work has pointed out that firms use 'selective revealing' (Henkel, 2006; Alexy et al., 2013) as a means of balancing the benefits of disclosing information to externals and the risk for this information to be adversely used by their competitors. While in selective revealing firms' information is directly exposed to rivals, mediated revealing establishes an additional safeguard by inserting a boundary organization between firms and externals. This arrangement enables firms to disclose information to externals that may be too sensitive to be disclosed directly can be disclosed to an intermediary. This in turn increases the potential benefits from revealing as firms are prepared to share their more important problems.

##### 5.1.1. Mediated revealing

Mediated revealing requires that the interacting parties trust the boundary organization. Trust refers to the confidence the involved parties have that an actor will adhere to mutually agreed lines of action (Nooteboom, 1996). The concept of 'trusted intermediary' (Rai et al., 2008) aptly encapsulates the specific role of the SGC. In information systems, trusted intermediaries allow system owners to ensure that certain types of information are separated from others (Pavlou and Gefen, 2004). The SGC played an equivalent role by

brokering information between firms and academic researchers. Like many trusted intermediaries, its mission was confined to pursuing a specific objective – managing an open data research agenda. By confining its activities to this focused objective, and keeping its distance from the organizations involved, the SGC succeeded in acquiring trust.

The kind of boundary organization characterized above shares some features with specialized innovation intermediaries that 'crowd-source' problem solutions via broadcast search and innovation contests (Jeppesen and Lakhani, 2010) or orchestrate the trading of knowledge (Dushnitsky and Klueter, 2011). Like the SGC, these intermediaries may anonymise the identity of the problem owner and they are trusted by the revealing parties. However, they do not practice open data and often require the problem solvers to assign their rights to the problem owner for a reward. Unlike these entities, boundary organizations engaged in open data initiatives face the additional challenge of having to engage potential innovators by offering non-monetary incentives. This consideration connects with issues of motivation addressed below.

##### 5.1.2. Enabling multiple goals

The second key mechanism consists in enabling multiple goals. In an open data scenario, firm benefits from openly generated scientific knowledge will depend on whether they can pique the interest of academic researchers in the topics they propose. We suggest that boundary organizations can accomplish this by enabling multiple goals to be pursued in the context of the collaboration. While for many of the SGC's activities the interests of academia and industry were aligned because the scientifically interesting proteins were also those likely to inform drug discovery, the SGC allowed some activities to be of purely academic interest. In other words, while goals sometimes overlapped, this was not always and not necessarily the case.

The SGC resolved goal conflicts by allowing multiple goals to co-exist instead of optimizing its activities and costs around either purely industrial or purely academic goals. The SGC pursued the sponsor firms' primary goal of mapping the protein structures but also encouraged the pursuit of goals concomitant with academic science research driven by curiosity and academic publishing (Owen-Smith, 2003). Goal co-existence allowed the SGC to attract high-calibre academic scientists and, as a second-order benefit,



helped it connect with academic collaborators further afield. While the concurrent pursuit of multiple partially incompatible goals may be seen as ineffective (Simon, 1964), it allows a boundary organization to garner more financial resources and access a wider spectrum of human capital.

Boundary organizations are well suited to enable the parallel pursuit and alignment of separate goals because their interstitial position allows them to manage operations by maintaining the social boundaries between the different participants, thereby avoiding potential conflicts that could arise from direct coordination efforts (Lamont and Molnar, 2002). A further benefit of maintaining social boundaries is that a boundary organization has a relative prerogative over how resources are allocated and how production is controlled (O'Mahony and Bechky, 2008), allowing for resources to be earmarked for the pursuit of different goals. Finally, boundary organizations are likely to be familiar with the norms and practices prevalent in the different domains in which their stakeholders operate. In our case, this allowed the SGC to build organizational procedures that created a context familiar to academic researchers and simultaneously remain aware of the industrially led priorities advocated by the firms. While these capabilities of boundary suggestions have been characterized by previous research, the novel insight emerging from our study is that boundary organizations can also act as trusted information brokers that aggregate and selectively distribute information. In the case of the SGC, this function that underpinned mediated revealing was a crucial ingredient of the organization's role in enabling open data with industry involvement.

## 5.2. Implications for university–industry partnerships

By coordinating mediated revealing and enabling multiple goals, boundary organizations such as the SGC provide an organizational solution to the challenges that firms face when seeking to instigate and shape large-scale open access initiatives. They help to attract high-calibre academic scientists to contribute to scientific grand challenges because they can reach a larger 'workforce' of academic researchers than conventional, un-mediated partnerships, and thereby achieve enhanced economies of scale. By using the potential of open data to draw on larger 'crowds' of scientists (Franzoni and Sauerermann, 2013), firms may participate in shaping the knowledge commons of industries or sectors in a way not achievable by conventional research collaborations.

Previous research has found that participating in the development of open science can be an important motive for firms to engage in collaboration with universities (Powell et al., 1996; Cohen et al., 2002; Murray, 2002; Simeth and Raffo, 2013). Our study contributes to this body of work in two ways. First, we provide a framework for understanding the conditions under which firms would participate in open data which compared to open science represents a more radical arrangement in terms of publicly sharing information and results. Second, we emphasize the nexus between the firms' ability to appropriate benefits from public science and their ability to shape its course. In face of the generally inverse relationship between the extent to which firm shape a research programme and their willingness to share the results publicly, our model suggests how boundary organizations can moderate this relationship in ways that allows firms greater influence of public research programmes without compromising their commercial interests.

There are however boundary conditions that will limit the applicability of open data partnerships. First, the described benefits will apply to open data initiatives particularly involving multiple firms. A single firm wishing to attract academic scientists can ensure confidentiality by setting up suitable contracts with specific universities or groups of universities. When several firms are

involved, however, they face the challenge of having to disclose potentially sensitive information to each other. Here a boundary organization provides an organizational solution that protects firms' sensitive information from being disclosed to each other. The more competitive the research context is, the more relevant boundary organizations and mediated revealing will become for accomplishing joint open data initiatives.

In contrast with proprietary research collaboration, the ability to benefit from open data is particularly contingent upon time-based competition and superior complementary assets (Teece, 1986). In the case of the Structural Genomics Consortium, the pharmaceutical companies acquiesced to the unrestricted release of knowledge produced about proteins only because they held downstream assets for exploiting this knowledge. Complementary assets availability represents a boundary condition for firm participation in open data initiatives, implying that the formula used by the SGC will apply primarily to basic science initiatives. For open access to advance further down the R&D value chain, boundary organizations will have to offer additional quid-pro-quos to participating firms, such as advance access to research results or commercial control over parts of the knowledge being generated. Future research should examine under what conditions and organizational terms firms will engage in open data collaborations that are more applied than the basic science oriented collaboration studied here.

From the viewpoint of charitable and governmental research funders, our study suggests that large scale open data initiatives do not have to be purely public but can be extended to public–private partnerships. It is by now accepted that new models of collaboration in the wake of the Human Genome Project, using 'weak' intellectual property systems and loose large-scale coordination offer pathways to enhance the overall societal impact of public science certainly within its own boundaries (Rhoten and Powell, 2008; Rai et al., 2008; Kenney and Patton, 2009). In this way, they can mitigate the delay or reduction of follow-on academic research and product innovation caused by the presence of intellectual property protection (Murray and Stern, 2007; Williams, 2010).

Our study not only demonstrates that the benefit of enhanced cumulative innovation provided by open data can be achieved even when for-profit firms are involved. Furthermore, one may plausibly argue that the involvement of industrial partners increases the relevance of the science produced for tackling societal challenges such as the development of new drugs. While industry often fails to cover areas of societal need – as exemplified by the failure in orphan drugs – the involvement of science users in shaping the agenda for scientific challenges will promote the creation of knowledge potentially instrumental for developing future innovations. Simultaneously, the risk of publicly funded research being captured by particular industry interests is reduced because no intellectual property is produced and no privileged access to research results is provided. In this sense, the SGC exemplifies a way in which large-scale, decentralized and open scientific collaborations can be made more impactful by involving industry participants.

## 6. Conclusions and policy implications

The recent "supersizing" of science (Vermeulen, 2010) is likely to change the parameters under which firms engage in partnerships with universities. The Human Genome Project and other large-scale life science programmes have demonstrated, inside science, how the open sharing of data can dramatically increase the volume and the speed of research. We consider for the first time how these developments in science may impact on research partnerships between firms and universities. We suggest that boundary organizations can be an effective tool for firms, academic researchers and science funders to obtain benefits from

open data collaboration. Boundary organizations perform mediated revealing, allowing firms to disclose their research problems to a broad audience of innovators and shape their research agenda, and simultaneously minimize the risk that this information would be adversely used by competitors. Moreover, by enabling multiple goals boundary organizations can attract external innovators to collaborations because this allows them to pursue their own goals and objectives. By providing a formula for bringing together universities and industry to pursue big science, boundary organizations may prove an effective tool for helping to advance the knowledge commons underpinning science-based industries.

Our study has important implications for policy particularly as open data has hitherto been primarily used inside public science. The Human Genome Project has impressively demonstrated the power of open, speedy data sharing (Williams, 2010). While for-profit firms have been slower to embrace this principle in their partnerships with academia, our study offers insight to policy makers on how to obtain greater participation by industry in open data initiatives. From a policy viewpoint, this would result in two significant benefits. First, industrial participation in large-scale scientific collaborations can guide scientific enquiry towards greater societal relevance. Industrial input into decision-making on work programmes will likely focus on knowledge creation in areas expected to contribute to new technologies and products. In the case of the SGC, for instance, industry's target lists focused on those proteins that were seen as the most likely to contribute to the discovery of new drugs. From policy makers perspective, this outcome can be perceived as the expansion of the industry commons because all of the new knowledge created by the partnership was made openly accessible to all industry participants, including those not part of the consortium. Second, involving industry reduces the cost to the public purse of large-scale scientific collaborations as industrial participants can be asked to contribute to the total cost. The justification for pecuniary contributions is that firms that participate are given the opportunity to shape the direction of research to be conducted within an open data consortium, and hence can be expected to leverage their complementary assets to exploit the generated knowledge even though it is released into the public domain.

Given the potential benefits, our insights on the mechanisms for open data may provide policy makers, as well as managers in science-intensive firms, with a possible blueprint for organizing public–private research partnerships. Much of the public support for industry involvement in academic science currently directly subsidizes projects between firms and universities. These are often relatively small-scale projects and are also likely to involve intellectual property provisions. Support for open data initiatives led by boundary organizations such as the SGC has the potential to considerably amplify the societal benefits from government subsidies because (a) initiatives can be larger-scale as they can more easily involve multiple universities and (b) open data principles can be more easily negotiated (with firms as well as universities) by specialist boundary organizations.

## Acknowledgements

We are grateful for comments by our editor Martin Kenney, the anonymous reviewers, as well as Oliver Alexy, Eva Boxenbaum, Giuseppe Delmestri, Lars Frederiksen, Gerry George, Joachim Henkel, Ilze Kivleniece, Matt Kraatz, Tom Lawrence, Mike Lounsbury, Anita McGahan, Maureen McKelvey, Woody Powell, Joel West and Mike Wright. Previous versions were presented at the Academy of Management Meeting 2014, the European Group of Organization Studies Colloquium 2014, the ABC Research Network Conference 2012, a meeting of the Organization Theory Research Group (OTREG) and seminars at the universities of Gothenburg, Linz

and Nottingham. Markus Perkmann acknowledges funding from the UK Economic and Social Research Council (ESRC) via an AIM Fellowship (RES-331-27-0063).

## Appendix 1. List of meeting minutes and other archival documents

| Code | Date     | Title   | Word length |
|------|----------|---|-------------|
| d1   | 01/02/05 | Scientific Committee Meeting Minutes                    | 5406        |
| d2   | 02/08/05 | Scientific Committee Meeting Minutes                    | 4838        |
| d3   | 07/02/06 | Scientific Committee Meeting Minutes                    | 4370        |
| d4   | 02/05/06 | Scientific Committee Meeting Minutes                    | 4273        |
| d5   | 30/05/06 | Audit and Risk Meeting Minutes                          | 982         |
| d6   | 06/06/06 | Board Meeting Minutes                                   | 1151        |
| d7   | 31/06/06 | Business Committee Meeting Minutes                      | 1346        |
| d8   | 06/08/06 | Scientific Committee Meeting minutes                    | 3651        |
| d9   | 05/09/06 | Business Committee Meeting Minutes                      | 522         |
| d10  | 05/09/06 | Board Meeting Minutes                                   | 2991        |
| d11  | 07/11/06 | Scientific Committee Meeting Minutes                    | 3451        |
| d12  | 05/12/06 | Board Meeting Minutes                                   | 3107        |
| d13  | 06/02/07 | Scientific Committee Meeting Minutes                    | 3192        |
| d14  | 06/03/07 | Board Meeting Minutes                                   | 2656        |
| d15  | 30/04/07 | Board Meeting Minutes                                   | 1754        |
| d16  | 05/07/07 | Board Meeting Minutes                                   | 2126        |
| d17  | 05/07/07 | Memorandum Articles of Association                      | 9781        |
| d18  | 31/05/07 | Funding Agreement                                       | 18,516      |
| d19  | 11/05/09 | Presentation by SGC lead scientist 1                    |             |
| d20  | 2003–09  | SGC press communications                                | 1889        |
| d21  | 29/09/11 | Presentation by SGC lead scientist 1                    |             |
| d22  | 09/2009  | Commentary by SGC scientists in Nature Chemical Biology | 4481        |
| d23  | 2009–12  | SGC website   |             |
| d24  | 17/2/11  | Economist intelligence unit                             | 824         |
| d25  | 29/09/11 | Presentation by SGC lead scientist 2                    |             |
| d26  | 14/3/07  | Presentations by SGC lead scientist 3                   |             |
| d27  | 12/3/07  | Presentation by SGC scientist                           |             |

## Appendix 2. List of interviews and recorded duration in minutes

| No. | Date     | Affiliation        | Description                        | '               |
|-----|----------|--------------------|------------------------------------|-----------------|
| i1  | 28/11/07 | Pharma member      | Face to face, UK                   | 80              |
| i2  | 13/11/07 | SGC lead scientist | Face to face, UK                   | 37              |
| i3  | 20/12/07 | SGC lead scientist | Phone                              | 60 <sup>a</sup> |
| i4  | 07/02/08 | Pharma member      | Phone                              | 49              |
| i5  | 11/02/08 | Foundation member  | Face to face, UK                   | 54              |
| i6  | 13/02/08 | Pharma member      | Phone                              | 48              |
| i7  | 15/02/08 | Pharma member      | Phone                              | 55              |
| i8  | 25/02/08 | Pharma member      | Phone                              | 51              |
| i9  | 25/03/08 | SGC scientist      | Phone                              | 60              |
| i10 | 11/02/08 | Foundation member  | Face to face, UK                   | 30              |
| i11 | 23/06/08 | SGC lead scientist | Face to face, Canada               | 78              |
| i12 | 23/06/08 | SGC scientist      | Face to face, Canada               | 50              |
| i13 | 23/06/08 | Government sponsor | Face to face, location undisclosed | 64              |
| i14 | 23/06/08 | SGC scientist      | Face to face, Canada               | 59              |

## Appendix 2 (Continued)

| No. | Date     | Affiliation            | Description      |                  |
|-----|----------|------------------------|------------------|------------------|
| i15 | 03/12/08 | SGC lead scientist     | Phone            | 48               |
| i16 | 23/11/09 | SGC lead scientist     | Face-to-face, UK | 120 <sup>a</sup> |
| i17 | 07/04/10 | SGC lead scientist     | Phone            | 69               |
| i18 | 08/04/10 | Sponsor representative | Face-to-face, UK | 120 <sup>a</sup> |
| i19 | 10/05/10 | Sponsor representative | Phone            | 45               |
| i20 | 04/06/10 | Foundation member      | Face-to-face, UK | 65               |
| i21 | 15/09/11 | External scientist     | Face-to-face, UK | 60               |
| i22 | 21/09/11 | SGC scientist          | Phone, UK        | 30 <sup>a</sup>  |

<sup>a</sup> Interviews based on notes rather than recordings at the request of interviewees. All interviewees were assured that their identity would remain confidential.

## References

- Ahuja, G., Lampert, C., Novelli, E., 2013. The second face of appropriability: generative appropriability and its determinants. *Academy of Management Review* 38, 248–269.
- Alexy, O., George, G., Salter, A.J., 2013. Cui bono? The selective revealing of knowledge and its implications for innovative activity. *Academy of Management Review* 38, 270–291.
- Allarakhia, M., Walsh, S., 2011. Managing knowledge assets under conditions of radical change: the case of the pharmaceutical industry. *Technovation* 31, 105–117.
- Allen, R.C., 1983. Collective invention. *Journal of Economic Behavior and Organization* 4, 1–24.
- Arrow, K.J., 1971. Classificatory notes on the production and transmission of technical knowledge. In: Arrow, K.J. (Ed.), *Essays in the Theory of Risk-Bearing*. Markham, Chicago.
- Bercovitz, J.E.L., Feldman, M.P., 2007. Fishing upstream: firm innovation strategy and university research alliances. *Research Policy* 36, 930–948.
- Boudreau, K., Lakhani, K.R., 2015. “Open” disclosure of innovations, incentives and follow-on reuse: theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy* 44, 4–19.
- Boulton, G., Rawlins, M., Vallance, P., Walport, M., 2011. Science as a public enterprise: the case for open data. *The Lancet* 377, 1633–1634.
- Clarkson, G., Toh, P.K., 2010. ‘Keep out’ signs: the role of deterrence in the competition for resources. *Strategic Management Journal* 31, 1202–1225.
- Cockburn, I.M., Henderson, R., Stern, S., 1999. *Balancing Incentives: The Tension Between Basic and Applied Research*. National Bureau of Economic Research Working Paper Series.
- Cohen, W.M., Nelson, R.R., Walsh, J., 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not). NBER Working Paper W7552. National Bureau of Economic Research, Cambridge, MA.
- Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and impacts: the influence of public research on industrial R&D. *Management Science* 48, 1–23.
- Cook-Deegan, R., 2007. The science commons in health research: structure, function, and value. *Journal of Technology Transfer* 32, 133–156.
- Corbin, J.M., Strauss, A.L., 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd ed. Sage, Newbury Park.
- D’Este, P., Perkmann, M., 2011. Why do academics engage with industry? The entrepreneurial university and individual motivations. *Journal of Technology Transfer* 36, 316–339.
- Dasgupta, P., David, P.A., 1994. Toward a new economics of science. *Research Policy* 23, 487–521.
- Dushnitsky, G., Klueter, T., 2011. Is there an eBay for ideas? Insights from online knowledge marketplaces. *European Management Review* 8, 17–32.
- Edwards, A., 2008. Open-source science to enable drug discovery. *Drug Discovery Today* 13, 731–733.
- Eisenhardt, K.M., 1989. Building theories from case study research. *Academy of Management Review* 14, 532–550.
- Franzoni, C., Sauermann, H., 2013. Crowd science: the organization of scientific research in open collaborative projects. *Research Policy* 43, 1–20.
- Garnier, J.-P., 2008. Rebuilding the R&D engine in big pharma. *Harvard Business Review* 86, 68–77.
- Gittelman, M., Kogut, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science* 49, 366–382.
- Gowers, T., Nielsen, M., 2009. Massively collaborative mathematics. *Nature* 461, 879–881.
- Guston, D.H., 2001. Boundary organizations in environmental policy and science: an introduction. *Science, Technology, and Human Values* 26, 399–408.
- Hagedoorn, J., Link, A.N., Vonortas, N.S., 2000. Research partnerships. *Research Policy* 29, 567–586.
- Hall, B.H., Link, A.N., Scott, J.T., 2001. Barriers inhibiting industry from partnering with universities: evidence from the Advanced Technology Program. *Journal of Technology Transfer* 26, 87–98.
- Henkel, J., 2006. Selective revealing in open innovation processes: the case of embedded Linux. *Research Policy* 35, 953–969.
- Hope, J., 2009. *Biobazaar: The Open Source Revolution and Biotechnology*. Harvard University Press, Cambridge (MA).
- Howells, J., 2006. Intermediation and the role of intermediaries in innovation. *Research Policy* 35, 715–728.
- Human Genome Project, 1996. *Summary of Principles Agreed Upon at the First International Strategy Meeting on Human Genome Sequencing* (Bermuda, 25–28 February 1996).
- Jeppesen, L.B., Lakhani, K.R., 2010. Marginality and problem solving effectiveness in broadcast search. *Organization Science* 21, 1016–1033.
- Kenney, M., Patton, D., 2009. Reconsidering the Bayh–Dole Act and the current university invention ownership model. *Research Policy* 38, 1407–1422.
- Lamont, M., Molnar, V., 2002. The study of boundaries in the social sciences. *Annual Review of Sociology* 28, 167–195.
- Levine, S.S., Prietula, M.J., 2014. *Open Collaboration for Innovation: Principles and Performance* (in press).
- Link, A.N., Scott, J.T., 2005. Universities as partners in US research joint ventures. *Research Policy* 34, 385–393.
- Mansfield, E., 1991. Academic research and industrial innovation. *Research Policy* 20, 1–12.
- Melese, T., Lin, S.M., Chang, J.L., Cohen, N.H., 2009. Open innovation networks between academia and industry: an imperative for breakthrough therapies. *Nature Medicine* 15, 502–507.
- Miller, C., 2001. Hybrid management: boundary organizations, science policy, and environmental governance in the climate regime. *Science, Technology and Human Values* 26, 478.
- Molloy, J.C., 2011. The open knowledge foundation: open data means better science. *PLoS Biology* 9, e1001195.
- Mowery, D.C., Nelson, R.R., Sampat, B.N., Ziedonis, A.A., 2001. The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole Act of 1980. *Research Policy* 30, 99–119.
- Munos, B., 2006. Can open-source R&D reinvigorate drug research? *Nature Reviews Drug Discovery* 5, 723–729.
- Munos, B., 2009. Lessons from 60 years of pharmaceutical innovation. *Nature Reviews Drug Discovery* 8, 959–968.
- Murray, F., 2002. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research Policy* 31, 1389–1403.
- Murray, F., 2010. The oncomouse that roared: hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. *American Journal of Sociology* 116, 341–388.
- Murray, F., O’Mahony, S., 2007. Reconceptualizing the institutional foundations of cumulative innovation. *Organization Science* 18, 1006–1021.
- Murray, F., Stern, S., 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior and Organization* 63, 648–687.
- Murray-Rust, P., 2008. Open data in science. *Serials Review* 34, 52–64.
- Nooteboom, B., 1996. Trust, opportunism and governance: a process and control model. *Organization Studies* 17, 985–1010.
- Nuvolari, A., 2004. Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine. *Cambridge Journal of Economics* 28, 347–363.
- O’Mahony, S., Bechky, B.A., 2008. Boundary organizations: enabling collaboration among unexpected allies. *Administrative Science Quarterly* 53, 422–459.
- Owen-Smith, J., 2003. From separate systems to a hybrid order: accumulative advantage across public and private science at Research One universities. *Research Policy* 32, 1081–1104.
- Panagopoulos, A., 2003. Understanding when universities and firms form RJVs: the importance of intellectual property protection. *International Journal of Industrial Organization* 21, 1411–1433.
- Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., Schacht, A.L., 2010. How to improve R&D productivity: the pharmaceutical industry’s grand challenge. *Nature Reviews Drug Discovery* 9, 203–214.
- Pavlou, P.A., Gefen, D., 2004. Building effective online marketplaces with institution-based trust. *Information Systems Research* 15, 37–59.
- Perkmann, M., Neely, A., Walsh, K., 2011. How should firms evaluate success in university–industry alliances? A performance measurement system. *R&D Management* 41, 202–216.
- Perkmann, M., West, J., 2015. Open science and open innovation: sourcing knowledge from universities. In: Link, A.N., Siegel, D.S., Wright, M. (Eds.), *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship*. The University of Chicago Press, Chicago (in press).
- Pincock, S., 2007. Pharma Goes Open Access, The Scientist Online resource <http://www.thescientist.com/news/home/52891> (accessed 17.05.08).
- Powell, W.W., Koput, K.W., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. *Administrative Science Quarterly* 41, 116–145.
- Rai, A.K., Reichman, J.H., Uhler, P.F., Crossman, C.R., 2008. Pathways across the valley of death: novel intellectual property strategies for accelerated drug discovery. *Yale Journal of Health Policy, Law, and Ethics* 8, 53–89.
- Reichman, O., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. *Science* 331, 703–705.
- Rhoten, D., Powell, W.W., 2008. The frontiers of intellectual property: expanded protection versus new models of open science. *Annual Review of Law and Social Science* 3, 345–373.

- Rosenberg, N., Nelson, R.R., 1994. American universities and technical advance in industry. *Research Policy* 23, 323–348.
- Saez, C.B., Marco, T.G., Arribas, E.H., 2002. Collaboration in R&D with universities and research centres: an empirical study of Spanish firms. *R&D Management* 32, 321–341.
- Schuhmacher, A., Germann, P.-G., Trill, H., Gassmann, O., 2013. Models for open innovation in the pharmaceutical industry. *Drug Discovery Today* 18, 1133–1137.
- Simeth, M., Raffo, J.D., 2013. What makes companies pursue an Open Science strategy? *Research Policy* 42, 1531–1543.
- Simon, H.A., 1964. On the concept of organizational goal. *Administrative Science Quarterly* 9, 1–22.
- Swierstra, T., Vermeulen, N., Braeckman, J., van Driel, R., 2013. Rethinking the life sciences. *EMBO Report* 14, 310–314.
- Teece, D.J., 1986. Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. *Research Policy* 15, 285–305.
- Vermeulen, N., 2010. *Supersizing Science: On Building Large-Scale Research Projects in Biology*. Dissertation.com, Boca Raton.
- Vermeulen, N., Parker, J.N., Penders, B., 2013. Understanding life together: a brief history of collaboration in biology. *Endeavour* 37, 162–171.
- von Hippel, E., von Krogh, G., 2006. Free revealing and the private-collective model for innovation incentives. *R&D Management* 36, 295–306.
- Washburn, J., 2008. *University, Inc.: The Corporate Corruption of Higher Education*. Basic Books, New York.
- Wellcome Trust, 2003. Sharing data from large-scale biological research projects: a system of tripartite responsibility. In: Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale, USA.
- Williams, H.L., 2010. Intellectual property rights and innovation: evidence from the human genome, National Bureau of Economic Research Working Paper 16213.