

Pistoia Alliance

Open Science Monitor Case Study

Laia Pujol Priego, Jonathan Wareham
January – 2019

EN



Pistoia Alliance- Open Science Monitor Case Study

European Commission

Directorate-General for Research and Innovation

Directorate A — Policy Development and Coordination

Unit A.2 — Open Data Policy and Science Cloud

E-mail Rene.VonSchomberg@ec.europa.eu

RTD-PUBLICATIONS@ec.europa.eu

European Commission

B-1049 Brussels

Manuscript completed in July 2018.

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2018

PDF ISBN 978-92-79-96537-1 doi: 10.2777/073659 KI-05-18-030-EN-N

© European Union, 2018.

Reuse is authorized provided the source is acknowledged. The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

For any use or reproduction of photos or other material that is not under the EU copyright, permission must be sought directly from the copyright holders.

Pistoia Alliance

Open Science Monitor Case Study







Table of contents

Acknowledgments 5

1 Introduction 6

2 Background 7

3 Drivers 12

4 Barriers 12

5 Impact 12
5.1 For Science 12
5.2 For Industry 15
5.3 For Society 15

6 Lessons learned 16

7 Policy conclusions 17

References 18

Acknowledgments

Disclaimer: The information and views set out in this study report are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this case study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

The case study is part of Open Science Monitor led by the Lisbon Council together with CWTS, ESADE, and Elsevier.

Authors

Laia Pujol Priego – Ramon Llull University, ESADE

Jonathan Wareham – Ramon Llull University, ESADE

1 Introduction

The cost of developing a drug, including the cost of failures, has never been higher: it is estimated at 2,6 billion- more than double than just a decade ago (Tufts Center for the Study of Drug Development 2015). The total R&D expenditure globally for the pharmaceutical and biotechnology companies increased from USD 108 billion (2006) to USD 141 billion (2015)(Evaluate Pharma 2017), and the R&D productivity is increasingly declining since the last decade (Lee, 2015). The increasing scientific complexity of diseases areas like rare cancers or neurological conditions, the growing regulatory burden and IP concerns require investing resources in new approaches in early drug discovery that investigate diseases at the molecular and genetic level.

Since the human genome sequencing and the consequent technological advances with the advent of genomics and associated biological sciences such as proteomics and structural biology, biomedical research has become a data-driven science, requiring ready access to primary data and relevant distributed knowledge in a multitude of sources. Aggregation of data and interoperability of such resources are essential for enabling data sharing across research domains and organisational boundaries to accelerate drug discovery and development. Industry players and academia put particular emphasis on a semantics-driven approach to data aggregation, in order to make possible such scientific data aggregation in shared infrastructures to reduce data redundancy and to uncover existing inconsistencies.

The advent of machine learning and artificial intelligence for biomedical research is very promising, however without a necessary amount of metadata, smart annotation of existing data assets, and the appropriate semantic standards, the algorithms cannot make much headway. The generation of such semantic annotation of data assets are hard and resource intensive but an essential collective problem to overcome.

As a result, there are different initiatives emerging in the biomedical field that try to pursue what it is called pre-competitive collaboration between academia and competitors in order to accelerate sharing of resources (i.e., data, technology and expertise) to collectively accelerate the early stages of drug discovery, which is considered "the single most critical barrier in achieving a free flow of drug candidates into the pipeline"(Chaguturu et al. 2014, p.25). The rationale behind such collaborations is to increase gains in R&D efficiency through a reduction of R&D spending by sharing the costs and risks across different organizations. The technological challenges associated with the exploitation of the deluge of biomedical data has led to the emergence of such pre-competitive collaborations, where competitors have incentives to pool common resources and reduce redundant investments in the early R&D stages of drug discovery; later, in de-risked stages, the race starts amongst the different organisations to bring the best drug into the market (Nambisan et al., 2017).

"Ten years ago, pharma companies were not keen to work with each other or with academia. They have now realized that partnerships can increase the likelihood of success" (Ward 2016). According to a Deloitte (2018) study, 334 new R&D consortia have been created from 2005 to 2014, nine times the number formed during the prior decade (Lesser and Hefner, 2017).

Pistoia Alliance is one example of a global pre-competitive collaboration between competitors and academia that work together in pre-competitive research. Members of Pistoia alliance implement pre-competitive collaboration to address issues around aggregating, accessing, and sharing data that are essential to innovation but provide little competitive advantage.

The primary goal of the alliance is to streamline what companies have agreed to consider "non-competitive elements" of the R&D workflow by the specification of common

standards. Senior pharmaceutical industry R&D IT directors, after a meeting in Pistoia (Italy) in 2007), agreed that most of the activities carried out by researchers in biomedical organizations were redundant. "For example, the activities to determine a gene sequence, identify a signal transduction pathway, search a chemical repository or keep abreast of the scientific literature were all, of course, necessary, but remarkably common" (Drug Discovery World, 2011). In the Pistoia meeting, R&D IT managers agreed that much redundant effort by R&D IT departments in the different organisations might be dramatically reduced by sharing data, resources, and capabilities amongst companies about considered precompetitive research activities. The reductions in R&D spending in these early stages of drug development by sharing the costs of pre-competitive research could be invested for each company in more strategic and transformational innovative initiatives.

By collectively working in the definition of industry standards and processes and providing open access to the resulting outputs of the collaboration and web services following an open source framework, the case inspires other potential actions that seek to facilitate a common framework for collaborations between corporates and the broader research community. Most importantly, the case sheds light on the barriers that the different stakeholders experienced when trying to put the grounds for sharing data and building collective knowledge to accelerate scientific discovery. The lessons learned from such an initiative can inspire other industry and academic players while providing some key takeaways for policy-makers on potential interventions to facilitate such pre-competitive collaborations to foster scientific data sharing and re-use.

2 Background

Pre-competitive collaborations in biomedical research are often denominated as "data commons" or "research commons" (Mishra et al. 2016) where the primary goal of the collaboration is to generate technological standards, infrastructures (including data platforms) and novel scientific knowledge, while at the same time they allow organisations to pursue their corporate interests (De Vrueh and Crommelin, 2017). Different challenges emerge when trying to align various organisational goals to collectively speed up scientific data sharing and aggregation to accelerate drug discovery: "there are real issues and difficulties in establishing and maintaining large-scale cross-function and cross-organizational research collaborations. First and foremost is the illogical nature of cooperation." (Chaguturu, 2017 p.12).

There are different types of pre-competitive collaborations depending on their origin and organisational forms: i. *Industry-driven* (i.e., when the collaboration has been created originally by firms); ii. *Policy-driven* (i.e., when the alliance has been created as a policy instrument to support multi-stakeholder long-term R&D programs implementation in specific technology areas (e.g., IMI). Pre-competitive collaborations can also be classified according to the primary focus of their collaboration (Altshuler et al., 2010): i. Generation of *mutual standards* to foster scientific data sharing; ii. *Data generation and aggregation,* where partnerships pool together financial resources and expertise to create, maintain and aggregate data; iii. *Knowledge creation*, where collaborations leverage existing standards and shared tech infrastructures to create new collective knowledge in early drug discovery phases.

Pistoia Alliance is an ***industry-driven*** partnership created by large pharmaceutical companies (GlaxoSmithKline (GSK), AstraZeneca, Pfizer, and Novartis) in 2009 without government intervention to cooperate and share resources in the pre-competitive phase of drug development. The Alliance has been expanding since then accepting applications from both private and public who share the ideology of sharing data and resources of drug discovery research pipeline. Today some of the partners including, but not only: Lundbeck, Accelrys, ChemAxon, ChemITment, BioXPR, EdgeConsultancy, GGA, Rescentris, and UPCO. Pistoia was originated to generate ***mutual standards*** in industry, ontologies and web services that are made available under an open-source framework to academic institutions,

vendors and companies, to facilitate data sharing, data representation, text-mining activities and improve the R&D efficiency of the drug discovery and development process.

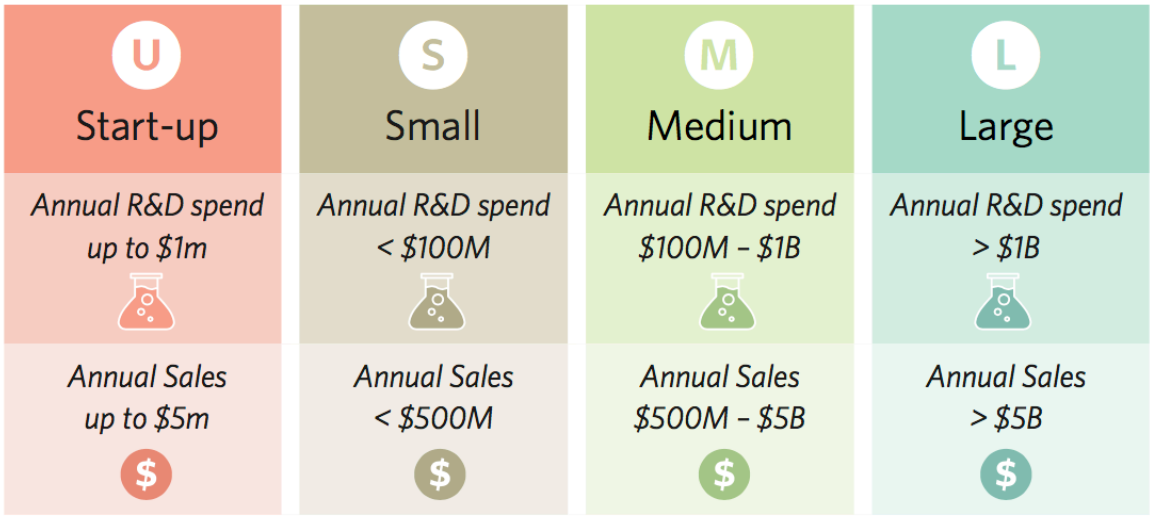
With many projects underway since its foundation, covering from Internet of Things (IoT) to 'ontologies mapping,' the alliance pools resources from many companies, and openly shares the output of its work with the outside research community. In the following section, we describe the R&D process in Pistoia and major R&D outputs.

2.1 R&D Process: Who can contribute

Membership at Pistoia Alliance

The alliance has designed different types of membership in order to engage a diverse set of interest in enhancing life science R&D. In order to do so, the fee that each member pays differs according to his or her annual Sales and annual R&D expenditure (see figure 1 & figure 2). When "R&D spend" category is different from "Sales" category, members must select the larger category to calculate their annual fee to be in Pistoia.

Figure 1. Membership categories according to Annual R&D spend and Annual Sales (Pistoia Alliance, 2018)



For the four different categories, the annual membership fee differs from $1000 per year to $50.000 per year (see figure 2).

Figure 2. Membership fees per category (Pistoia Alliance, 2018)

	Individual/Academic/Non Profit/Gov't agency	Start-up	Small	Medium	Large
R&D SPEND	n/a	Up to $1M	< $100M	$100M – $1B	> $1B
SALES	n/a	Up to $5M	< $500M	$500M – $5B	> $5B
Annual Membership Fee	**No Fee**	**$1,000**	**$15,000**	**$25,000**	**$50,000**

If "R&D spend" category is different from "Sales" category, select the larger category

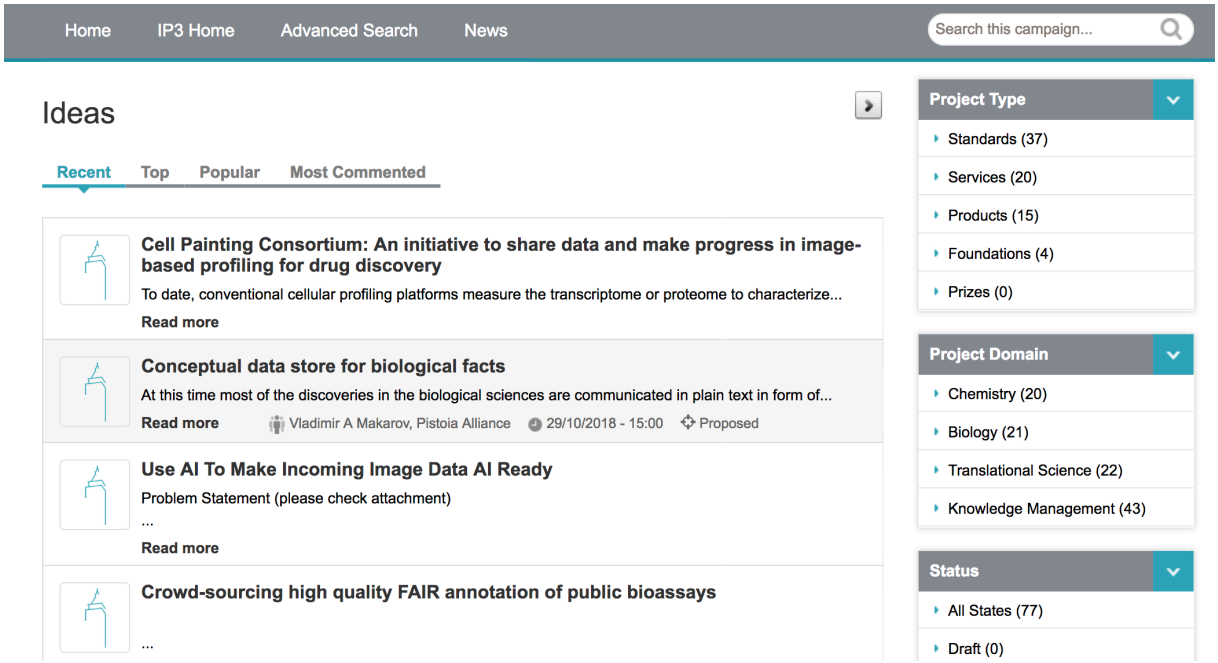
The benefits for members include, but not only: *social networking* with other industry organisations (e.g., attendance to member-only meetings, receptions, and networking events); *shape joint projects* with members of the alliance in common areas of interest (e.g., engage in joint projects specifications workgroups and project teams, participate in cross-company work streams); joint agreement on cross-company requirements for critical vendors.

Shaping the R&D agenda amongst all actors

Pistoia members come together in cross-industry working groups in order to identify common problems and to design projects that seek to implement the corresponding solutions. The alliance facilitates cross-company dialogues amongst their members (via workshops, meetings, and events) in order to identify the overlapping and resource-wasting replication of pre-competitive research activities in order to launch joint project teams to address these common challenges.

Besides the face-to-face opportunities facilitated by the partnership, Pistoia also encourages all members to submit new ideas and discuss existing proposals using an *Interactive Project Portfolio Platform (IP3)*[1] *(figure 3) (see section 2.3).*

Figure 3. *Snapshot from the IP3 platform* *(accessed January 2019)*



When Pistoia Alliance started its activity in 2009, the partnership set out a roadmap or programme of work around four major technology pilots:

- Sequence services: It consisted in "defining and documenting an externally hosted service where organizations can securely store and mine both in-house-derived gene/sequence information as well as the public domain gene databases" (Pistoia Alliance, 2018).

- SESL (Scientifically Enriched Scientific Literature): The idea was to assess the feasibility of *brokering* services that researchers could employ to rapidly gather information on disease-causing genes.

- Electronic Lab Notebook (ELN): Pistoia members set up the goal to define a common standard that would support scientists to use the same query across multiple ELNs.

- VSI (Vocabulary Standards Initiative): The partnership sought to define a standard, vocabulary-based methodology for querying the scientific literature and databases so that all hits could be found against molecular drug target search criteria rather than a subset.

[1] Access to *Interactive Project Portfolio Platform*: ip3.pistoiaalliance.org

2.2 R&D Outputs: Who can access the results of Pistoia collaboration and outputs

Pistoia is a standards development organisation that generates technical specifications to be widely implemented across the pharmaceutical industry. In practical terms, the alliance has the right to distribute its standards and any supporting material to ensure that the different players implement an Alliance standard correctly. To do so, the IP policy of Pistoia is similar to most standards organisations whose members include thousands of technology companies, universities, and end users (Pistoia alliance IPR policy, 2014).

The partnership openly shares the results of their projects for anyone to access and adopt the standards collaboratively produced by the alliance members. The goal of the alliance is to achieve a comprehensive and vast adoption by organisations working in life science so that the standards are useful and effectively help to increase R&D efficiencies supporting data sharing and rapid flows of information across players in the ecosystem. Besides sharing the outcome, typically the results of the project are accompanied by a range of resources that support the applicability and extensive use of the result of the project. An example of R&D output of the Pistoia partnership is the UDM standard.

Pistoia and UDM standard

Pistoia released a Unified Data Model (UDM). The primary goal of the UDM is to create and publish an open and freely available data format for storage and exchange of experimental data about compound synthesis and biological testing. The project released a first version of the UDM in June 2018, and since then the members worked to customise it for individual Electronic Lab Notebook (ELN) and software vendors. UDM was developed as an industry standard under the stewardship of the Pistoia Alliance project team with funding support from Biovia, Elsevier, GSK, Novartis and Roche. UDM fosters the exchange of data and the collaboration between researchers' organisations, and other stakeholders. It seeks to standardise data to unlock its potential value.[2]

As Pistoia members highlight standards like UDM are essential for advancing R&D, especially in areas where there is increased adoption of AI, ML and deep learning, which rely on the high quality of data. To be effective for Artificial Intelligence (AI) and Machine Learning (ML), there is a need of both volume and quality of data. Towards this goal, the experiments that are run today in early drug discovery are gathered in platforms such as Electronic Lab Notebooks (ELNs) in the industry. However, having an experiment in an ELN does not necessarily mean that the data captured is any better quality than in previous tools. Sometimes there is little-enhanced metadata to support future exploitation and integration and where this metadata is critical for AI/ML to be able to make accurate insights. Through this and other projects, the alliance seeks to foster increased efforts to make these data captured during the different experiments and be enhanced with metadata and essential linkages to already existing data management practices including ontologies and other semantic standards.

At present, the alliance is increasing awareness amongst life science companies to adopt the UDM standard in their organization and support further exchanges of data amongst organisations to increase pre-competitive collaboration. The implementation of UDM may lead to more accessible data and a reduction in duplicate data, and to accelerate research by reducing the time and cost devoted to converting data to other more valuable tasks. As Dr. Roman Affentranger, Head of Small Molecule Discovery Informatics at Roche, and UDM Project Champion explains: "Widespread adoption of the UDM standard has the potential to improve collaboration and data exchange across the entire life science industry."

[2] Source: https://www.pistoiaalliance.org/udm-release-v5/

FAIR in Pistoia Alliance

In its work towards finding semantic standards to foster scientific data sharing and aggregation, the partnership is fostering the FAIR data management of internal preclinical and clinical datasets in Pharma IT. By applying FAIR to internal preclinical and clinical datasets, organisations may enhance the reusability of their data holdings. FAIR data are data, which meet standards of Findability, Accessibility, Interoperability, and Reusability. To do so, members in the partnership are working together towards concrete, practical approaches and tools that exist today to bring existing pharma data assets aligned with the FAIR Guiding Principles, which define characteristics that data resources, tools, vocabularies, and infrastructures may have to support discovery and enable the reuse by third-parties. In March 2016 a consortium of scientists and organisations published the "FAIR Guiding Principles for scientific data management and stewardship" (Wilkinson et al., 2016). Team members have jointly evaluated critical topics around the Findability, Accessibility, and Interoperability of datasets while discussing together the role in this of recent development such as Bioschemas. Bioschemas aims to contribute schem.org, a community initiative launched by the leading search engines like Google or Bing in order to display structured data markup on web pages without modifying the presentation layer. In particular, Bioschemas aims to contribute to such initiative by improving the exposure of structured data in life sciences (e.g., protein sequence) by defining sets of minimum metadata properties. This would result in improved FAIR of biological data.

2.3 Technical Design: IP3, an innovation platform for portfolio generation and transparency

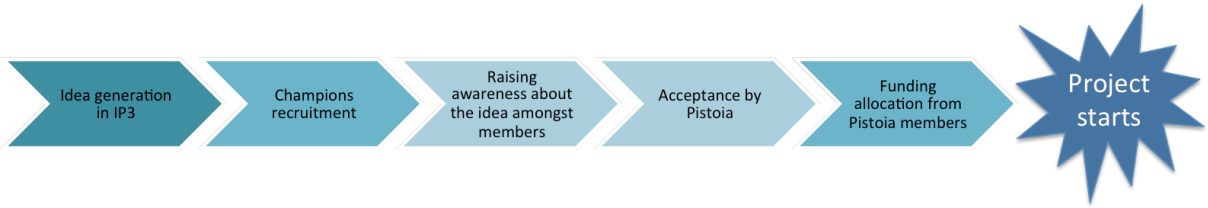
The partnership designed and launched IP3, which consists of an interactive platform that sought to embrace the philosophy of Pistoia about openness, access, and transparency about project processes. The public can access the platform and review the different projects. Individuals willing to contribute only need to register in order to upload a comment on the idea pool, or vote on the viability of the different existing ideas openly shared, or ultimately submit their ideas to increase R&D efficiency of pharma industry by reducing duplication of efforts.

Once the ideas are submitted, Pistoia IP3 team identify champions for the ideas and the champion start proactively raising awareness about the idea through the Pistoia community to engage diverse teams.

When the ideas have been able to generate a sufficient amount of member interest they go through the alliance management team to allocate investment so that they develop in preliminary business cases. Alliance turns to the Pistoia member community to solicit funding to develop the ideas into full projects with clear milestones and deliverables.

While the projects are active, they follow standard management procedures to support their activity. On-Going monitoring and assessment accompany the implementation of the project. IP3 system is used during the project implementation, and it documents background information, business justification, decision points, and day to day progress of the project. (Pistoia Alliance annual report, 2014).

Figure 4. ***From project ideation to completion in IP3 platform***



3 Drivers

As described in section 1, the main drivers for the creation of a partnership in 2007 when pharmaceutical industry R&D IT directors met in Pistoia (Italy), was to lower the barriers to R&D innovation effectively.

They identified similar tasks and activities performed across companies and agreed that by collectively working together in generating common standards that could accelerate data sharing within and across companies they would significantly reduce R&D costs and thus facilitate to re-allocate funding towards more transformative projects within the companies. By identifying redundancies and shared problems across the firms, they decided to create a partnership with academic groups, vendors and publishers to collectively lower the barriers for R&D innovation and accelerate the drug discovery process by generating common standards to facilitate data and information exchanges and smoothing the R&D process.

4 Barriers

The main barriers experienced by Pistoia Alliance to foster pre-competitive collaboration amongst partners have been the **lack of engagement of top corporate management in the alliance** in order to speed up the allocation of funding to the different project initiatives and to unlock some bottlenecks for sharing information and knowledge smoothly across the members of the alliance. The engagement of senior managers of R&D IT has been undoubtedly crucial but sometimes not enough to scale-up some projects, to accelerate collaborations amongst members, or to deepen collaborative strategies amongst the participants.

Another barrier, intrinsically related with the one mentioned above, has been the process of allocating **funding** for the different initiatives the alliance. The process still requires time and high commitment and engagement by the champion of the idea and the management team at Pistoia in order to effectively secure investment from the members to the idea. This affects not only the initial stage of the project (before it starts) but also when project participants try to expand the scope of the initiative and scale it up.

Finally, raising awareness and **increasing adoption of the open standards** developed, although happening at some level spontaneously and in a voluntary basis, it also requires active engagement, especially at the beginning when it is released, of both the participants and the Pistoia management team. The advantages of standards can only be grasped if widely adopted by a vast number of players. Thus, achieving such adoption requires time and investing in raising awareness within the biomedical community, while supporting its effective implementation.

5 Impact

5.1 For Science

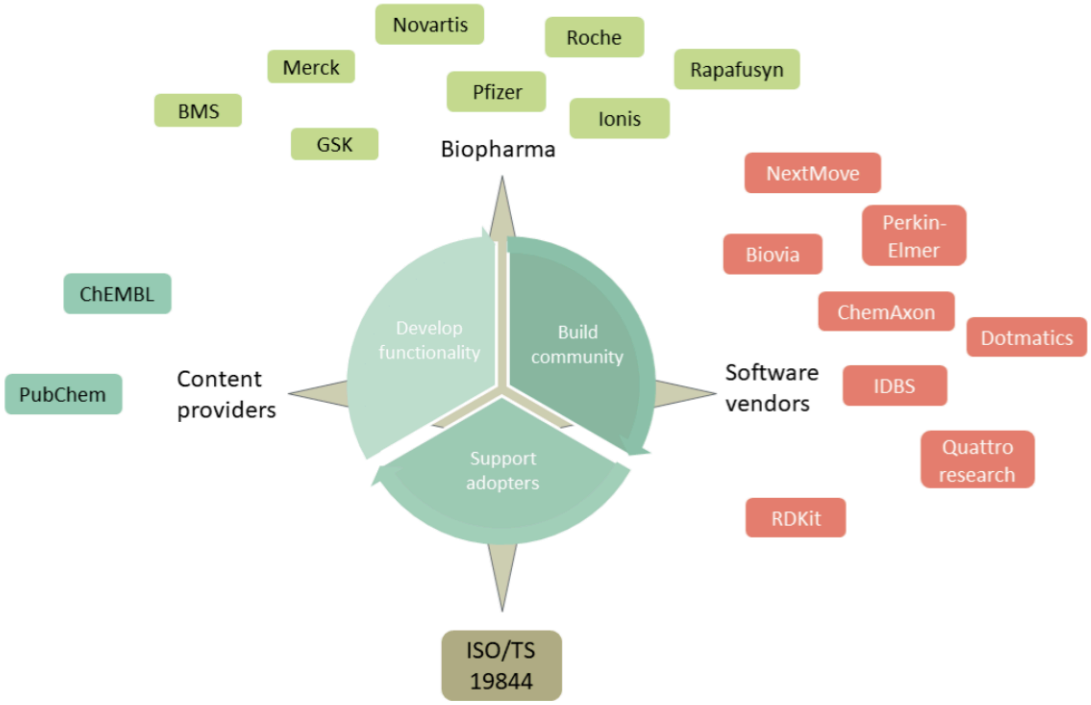
Pistoia Alliance has a project portfolio, which is continuously evolving with projects entering the pipeline periodically. We select one use case (project) in the history of Pistoia collaboration to show the impact of the collaboration:

1. **Hierarchical Editing for Large Molecules (HELM)**, which is a global standard notation that is machine-readable for large molecules originated at Pistoia Alliance[3]. HELM enables the representation of multiple types of biomolecules (e.g., proteins, nucleotides, antibody drug conjugates) whose size and complexity render sequence-based informatics methodologies not possible to apply. HELM has found a solution to such problem through a hierarchical notation that displays the complex macromolecules as polymeric structures and also supports unnatural components, such as unnatural amino acids, and chemical modifications (Zhang et al., 2012). HELM is now a FAIR standard.

   Scientists at Pfizer originated the idea of HELM in 2008. The problem that they were facing is that over the past years the pharmaceutical industry has deployed a wide range of increasingly complex bimolecular entities when looking for active therapeutic agents. Scientists at Pfizer realized that their bioinformatics infrastructure and methodology was not able to deal with these types of complex molecules and there was no commercial solution available. As a result, they came up with the idea of HELM, which soon gathers much interest by other Pistoia members facing the same problem.

   The alliance launched HELM as an open source project that can be found on GitHub. The project team included members from 23 different organisations in the life science space integrating a large and diverse ecosystem.

Figure 5. HELM *ecosystem* (source: HELM wiki, 2018)



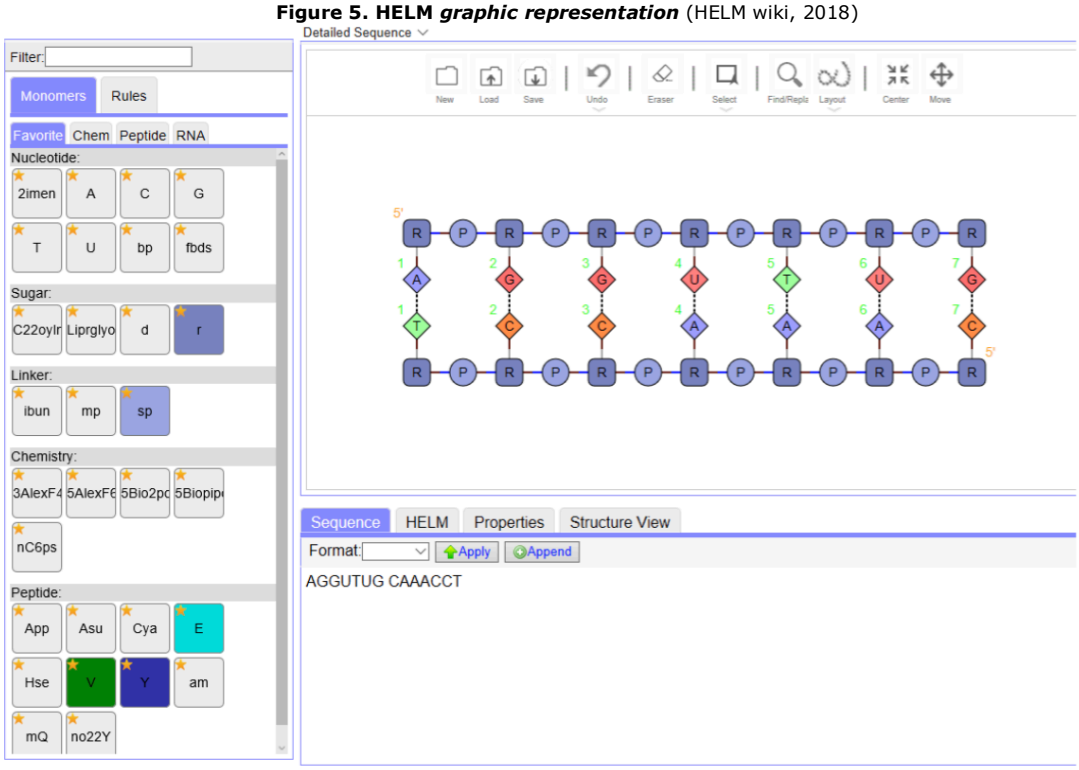
The HELM technology was officially released into the open source community on June 18, 2013. The release consisted of a GitHub repository containing the source code of the toolkit and editor, along with a web site (www.openhelm.org) containing the notation language specification, a free applet version of the editor, training videos and user guides. In 2014 Exchangeable HELM and in-line monomer definitions were implemented, and Roche released the HELM Antibody Editor (HAbE) as open source code in March 2015.

[3] About HELM: www.openhelm.org

The result of such collective effort is the official release of HELM technology on June 18, 2013. The release consisted of a GitHub repository that contained the source code of the toolkit and editor, together with a web site containing the notation language specification, a free applet version of the editor, supported with some training videos and user guidelines (HELM website).

The HELM editor is a graphical user interface that supports users "to drag and drop monomers onto a canvas, enter the structure as HELM notation or a variety of industry-standard sequence representations. It includes a range of functions supporting common actions such as the automatic creation of complementary RNA strands. It also includes an editor for the management of monomer definition and structure" (HELM leaflet, 2018)

Figure 5. HELM *graphic representation* (HELM wiki, 2018)



HELM notation became an open standard, published openly and for free. HELM is increasingly being adopted through the companies and academic organisations working on biopharma, including also solution providers and regulatory agents (Pistoia Alliance, 2018).

"*Recognised by the FDA as an acceptable submission format, HELM is nearing a critical mass of users as a standard notation to represent large therapeutic molecules in bioregistration systems. It is becoming more common for software developers to include APIs that make it easier to provide registration of chemicals and biologics from third-party applications such as electronic lab notebooks (ELNs)." (Moran, 2016).*

The uptake of HELM has been rapid and includes today a vast range of organisations as content providers, informatics vendors, and life science companies. The benefits of using HELM include facilitating the work of scientists willing to employ computational manipulations and calculations with complex molecules. HELM makes easier researcher tasks. Public databases such as ChEMBL have started to include some of their biomolecules structure using HELM standard. ChEMBL is a publicly available database of drugs, drug-like molecules and their targets, considered unique because of its focuses on all aspects of drug discovery and due to its size (ChEMBL contains information on more than 1.4 million compounds and over 12 million records of their effects on biological systems) (EMBL-EBI, 2018).

By doing this, HELM facilitates the integration of private and public data easier using the standard.

"The majority of ChEMBL peptides are now available in HELM, many major chemical software vendors have released HELM functionality as part of their suites, and two further major pharmaceutical companies have completed internal projects that use HELM to support their internal systems. The project is also aware of a large number of other organizations who are in the process of completing work" (HELM website, 2018) *(see section 5.2 about industry impact of HELM).*

As Sergio Rotstein, Director of Business Technology at Pfizer, besides the main impact of HELM, the process of developing an open standard together with a range of players has rendered benefits for the participants: *"The HELM project has exceeded my expectations in the execution of the project and the benefits derived. Additionally, there has been the unexpected community aspect of participating in such an alliance: Having the opportunity to work in a team comprised of your smartest peers, colleagues –and indeed competitors– from across the industry is an extremely rewarding experience for all involved."* (HELM, 2018)

5.2 For Industry

Regarding the open standard HELM developed in the framework of Pistoia Alliance, this use case offers an example of the potential industry impact of the pre-competitive collaboration of Pistoia and the effect of collectively agreeing in open standards that generate efficiencies in the R&D activity of organisations, while accelerating their research processes. Opening up HELM to the whole community provided freedom to the scientific software, content providers and other organisations working in life science to integrate the standard into their offerings.

For instance, in the case of HELM, the alliance commissioned a study which estimated the industry impact of HELM. The study calculated an annual productivity saving of over $225m. These productivity savings were obtained through the combination of (1) eliminating the need for other companies to develop their own approach; (2) making possible data sharing across companies in an efficient way, without costs of converting such data, and overall better and smooth flows of information between organisations and (3) providing more sophisticated tech tools available to scientists to develop their research. The estimated Return on Investment (ROI) assumed the adoption of HELM by the industry, which is currently growing, and illustrates the potential of Pistoia in generating R&D efficiencies within the innovation process of diverse actors.

Finally, as Sergio Rotstein, Director of Business Technology at Pfizer describes: "By working with the Pistoia Alliance to build an open- source community around HELM, Pfizer was able to rapidly transform an internal technology into a global open standard that is benefitting the entire industry." As he furthers explains: "Before HELM, our systems had no way of establishing deterministically if one ADC was identical to another ADC, a situation that seemed quite outrageous for this day and age!" (HELM website, 2018).

5.3 For Society

Finally, pre-competitive collaboration among competitors and academia to facilitate data exchange and smooth information flows amongst players has a direct impact on society.

Society at large benefit from "research commons" as it reduces duplication efforts and aggregated investment waste by breaking individual silos, and facilitates collective work that accelerates drug discovery processes, which are long and convoluted with high attrition rates. By facilitating scientist's work and corporate R&D processes in early

drug discovery, the Public benefits from higher probabilities of drugs going through the development process by increasing and improving the knowledge base on the diseases and their targets.

By pooling resources together and combining investments for common problems experienced in early stages of drug discovery, companies can liberate resources for other innovative, transformational projects in later stages of the R&D process where the race starts to bring the best drugs in the market. Society at large benefit from such efficiencies and such increase in the research quality that supports drug development processes.

6 Lessons learned

Pharma R&D is a high-risk investment and companies need to partner with competitors and academia to find common grounds, shared problems where common solutions can lead to a significant reduction of costs by combining a pool of resources to accelerate the drug discovery process. Since the 'data deluge' in biomedical research in the last decade, the generation of common standards is a central piece to facilitate data sharing in a data-driven research area. Shared technological tools and methods facilitate both scientists work, but also R&D departments to move faster to later stages of the drug development process.

Pistoia Alliance is an industry-driven pre-competitive collaboration that since its foundation in 2008 has developed impactful projects resulting in open standards that have led to an effective increase of R&D efficiencies. HELM is an example of how an open standard can facilitate daily work of scientists and companies when collaborating with each other.

At the same time, since its foundation, the alliance has learned some critical lessons in the process of generating and implementing the different projects:

- It is essential to offer a 'shared *space'* for industry players to exchange ideas

  One of the main benefits that Pistoia members highlight when describing the partnership is the possibility to exchange ideas with other companies in the same field, vendors, scientists in academia in order to increase interoperability of systems and facilitate the exchange of information amongst organisations. The possibility of a network with other players working in similar issues and experiencing similar problems generates synergies that lead to projects that can practically make more natural researcher work both in the firms and academia.

- Academia and firms across the pharmaceutical industry need to share and adopt common standards to accelerate discovery

  To grasp the benefits of AI/ML still there is the challenge of increasing volume but most importantly the quality of the data. In order to actually facilitate data exchanges, the lack of standards is still a bottleneck and a central piece of work to be done collectively amongst actors. Standards are resource intensive because require hours of distributed teams working together to agree on common tech specifications and solutions. However, once the standard is out, if it successfully implemented by a wide range of players, it can dramatically reduce inefficiencies and accelerate information sharing within and outside organizational boundaries.

- Face-to-face exchanges are still crucial and need to be combined with an interactive platform (IP3)

  The generation of a platform to orchestrate the collaboration of Pistoia was an important milestone. It supports all phases of the collaboration from idea generation to project implementation and evaluation. Besides its undoubted usefulness, it needs

to be combined with on-going face-to-face exchanges of the participants to facilitate dialogue and exchanges. The collective identification of problem- solution pairs amongst members still require to combine both digital platform and tools with face-to-face workshops, seminars, and meetings that foster a common dialogue.

- A strong management team is needed to support the collaboration of the participants

  Pistoia team supports the project development with a high-professional team of consultants with long-standing experience in biomedical R&D projects. They guarantee that the collaboration is effective. Distributed teams working together coming from different corporate and organisational cultures is challenging and having in place professional "brokers" and efficient managers to adequately support the projects is a relevant piece explaining the achievements of the collaboration.

- Still a challenge: How to convince your managers about allocating funding for 'pre-competitive' projects?

  Finally, the engagement of top managers in pre-competitive partnerships is crucial in order to accelerate the collaboration between participants and to scale-up the collaboration and achieve high-impact projects with the required level of funding and commitment of companies. It remains a challenge to allocate funding to each project and requires time to convince, in a project base, top managers to invest in the different pre-competitive initiatives.

7 Policy conclusions

"*The life sciences industry is at a real crossroads: heightened global competition, tougher regulatory standards, and continued cost pressures make efficient innovation in R&D more imperative than ever*," said Christopher McKenna, Global Head of Life Sciences Professional Services & Consulting at Clarivate Analytics (The Life Sciences Innovation Report, 2018)

Increasing awareness about already existing standards agreed by a wide range of industry and academic players is a task where policy-makers and regulatory agencies can be significant of help. Additionally, fostering synergies between different standards and similar initiatives in the ground is a crucial activity where policy-makers can make a difference to help accelerate the research processes. In this line, fostering FAIRfication of biomedical research activities (making data FAIR) is helping to reduce inefficiencies in the drug discovery and development process. Supporting FAIR guiding principles implementation combined with the adoption of industry-academic standards to increase scientific data sharing and exchange is a relevant task where policy-makers can play a fundamental role.

References

Amaka, 2015. *Against the odds, Yale improves access to data.* Retrieved 15th July from https://yaledailynews.com/blog/2015/01/28/against-the-odds-yale-improves-access-to-data/

Altshuler, J. S., Balogh, E., Barker, A. D., Eck, S. L., Friend, S. H., Ginsburg, G. S., Herbst, R. S., Nass, S. J., Streeter, C. M., and Wagner, J. A. 2010. Opening Up to Precompetitive Collaboration. Science Translational Medicine (2:52), pp. 52cm26-52cm26. (https://doi.org/10.1126/scitranslmed.3001515).

Carolyn Y. Johnson. How pharma kept a trove of drug trials out of public view on November 12, 2015. The Washington Post

Chaguturu, R., Murad, F., and Murad, F. 2014. Collaborative Innovation in Drug Discovery: Strategies for Public and Private Partnerships. Somerset, United States: John Wiley & Sons, Incorporated. (http://ebookcentral.proquest.com/lib/georgetown/detail.action?docID=1662193).

Chen R, Desai NR, Ross JS, et al. Publication and reporting of clinical trial results: a cross-sectional analysis across academic medical centers. BMJ 2016;352:i637.

De Vrueh, Remco LA, and Daan JA Crommelin. 2017. Reflections on the future of pharmaceutical public-private partnerships: from input to impact. Pharmaceutical Research 34.10 (2017): 1985-1999.

Drug Discovery World, 2011: http://www.ddw-online.com/business/p149291-what-are-the-obstacles-to-innovation-in-the-pharmaceutical- industrysummer-11.html. Accessed January 2019

Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One 2008;3:e3081.

Ebrahim, Shanil, Zahra N. Sohani, Luis Montoya, Arnav Agarwal, Kristian Thorlund, Edward J. Mills, and John PA Ioannidis. "Reanalyses of randomized clinical trial data." Jama312, no. 10 (2014): 1024-1032.

EMBL-EBI. "What is ChEMBL" https://www.ebi.ac.uk/training/online/course/chembl-quick-tour/what-chembl. Accessed January 2019.

Evaluate Pharma. 2017. World Preview 2017. (http://info.evaluategroup.com/rs/607-YGS-364/images/WP17.pdf).

HELM wiki. 2018. https://pistoiaalliance.atlassian.net/wiki/spaces/PUB/pages/8716303/HELM+Resources. Accessed January 2019.

Krumholz HM. Open science and data sharing in clinical research: basing informed decisions on the totality of the evidence. Circ Cardiovasc Qual Outcomes 2012;5:141–2.

Krumholz HM, Terry SF, Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. JAMA 2013;309:1355–6.

Lee, W. H. 2015. Open Access Target Validation Is a More Efficient Way to Accelerate Drug Discovery. PLOS Biology (13:6), p. e1002164. (https://doi.org/10.1371/journal.pbio.1002164).

Lesser, N., and Hefner, M. 2017. *How Biopharmaceutical Collaborations Are Fueling Biomedical Innovation*. Deloitte US. (https://www2.deloitte.com/us/en/pages/life-

sciences-and-health-care/articles/how-biopharma-collaborations-are-fueling-biomedical-innovation.html).

Mishra, A., Schofield, P. N., and Bubela, T. M. 2016. Sustaining Large-Scale Infrastructure to Promote Pre-Competitive Biomedical Research: Lessons from Mouse Genomics. *New Biotechnology* (33:2), pp. 280–294. (https://doi.org/10.1016/j.nbt.2015.10.002).

Moran, 2016. "Establishing Industry Best Practices for Working with Novel Biotherapeutic Entities." https://www.rdmag.com/article/2016/08/establishing-industry-best-practices-working-novel-biotherapeutic-entities. Accessed January 2019

Navar, Ann Marie, Michael J. Pencina, Jennifer A. Rymer, Darcy M. Louzao, and Eric D. Peterson. Use of open access platforms for clinical trial data. Jama 315, no. 12 (2016): 1283-1284.

The Economist. Spilling the Beans. Jul 25th, 2015

The Life Sciences Innovation Report, 2018. https://clarivate.com/g/life-sciences-innovation-report. Accessed January 2019.

The YODA Project. The YODA Project website.http://yoda.yale.edu. Accessed July 2018

Tufts Center for the Study of Drug Development. 2015. Personalized Medicine Gains Traction but Still Faces Multiple Challenges, Impact Report.

Wallack, 2017. Advancing meta-research through data sharing and transparency. Accessed July 2018: http://blogs.biomedcentral.com/on-medicine/2017/02/15/advancing-meta-research-through-data-sharing-and-transparency/

Wilcox MA, Savitz AJ, Addington AM, Gray GS, Guinan EC, Jackson JW, Lehner T, Normand SL, Ranu H, Senthil G, Spertus J, Valeri L, Ross JS. The Open Translational Science in Schizophrenia (OPTICS) project: an open-science project bringing together Janssen clinical trial and NIMH data. NPJ Schizophr. 2018;4(1):14.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; 't Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend (15 March 2016). "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data*. p. 160018. doi:10.1038/sdata.2016.18.

Zhang, Tianhong; Li, Hongli; Xi, Hualin; Stanton, Robert V.; Rotstein, Sergio H. (2012). "HELM: A Hierarchical Notation Language for Complex Biomolecule Structure Representation." *J. Chem. Inf. Model*. **52** (10): 2796–2806. doi:10.1021/ci3001925. PMID 22947017.

Getting in touch with the EU

IN PERSON
All over the European Union, there are hundreds of Europe Direct Information Centres.
You can find the address of the center nearest you at: http://europa.eu/contact

ON THE PHONE OR BY E-MAIL
Europe Direct is a service that answers your questions about the European Union.
You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: http://europa.eu/contact

Finding information about the EU

ONLINE
Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU PUBLICATIONS
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information center (see http://europa.eu/contact)

EU LAW AND RELATED DOCUMENTS
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU
The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

This case study presents the Pistoia Alliance, an example of a global pre-competitive collaboration between competitors and academia that work together in pre-competitive research. The main objective of the Alliance is to streamline what companies have agreed to consider "non-competitive elements" of the R&D workflow by the specification of common standards, in an attempt to lower the barriers to R&D innovation effectively. By collectively working in the definition of industry standards and processes and providing open access to the resulting outputs of the collaboration and web services following an open source framework, the case inspires other potential actions that seek to facilitate a common framework for collaborations between corporates and the broader research community.

Studies and reports