cBioPortal

Business Case 2015Q3

<table>
<thead>
<tr>
<th>Author (s)</th>
<th>Julia E. Rudolph – edited by Richard Holland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>0.1</td>
</tr>
<tr>
<td>Date</td>
<td>6th August 2015</td>
</tr>
<tr>
<td>Summary</td>
<td>Edited version of business case for public consumption – all confidential information and references to named individuals removed</td>
</tr>
</tbody>
</table>

info@pistoiaalliance.org  @pistoiaalliance  www.pistoiaalliance.org
## Contents

- cBioPortal .................................................................................................................. 1
- Contents .................................................................................................................... 2
- Executive Summary .................................................................................................. 3
- Project Aims ............................................................................................................... 4
  - Deliverables and milestones .................................................................................. 5
- Competitor analysis .................................................................................................. 5
  - Advantages of cBioPortal ...................................................................................... 6
  - How will we compete or collaborate with the competition? .................................. 6
- Market assessment .................................................................................................... 6
- Benefits and Cost ....................................................................................................... 7
  - Establishing an Advisory Committee model ....................................................... 7
  - Benefits of joining the advisory committee .......................................................... 7
  - Benefits to end-users ............................................................................................. 7
- Membership Structure ............................................................................................... 8
- Risks .......................................................................................................................... 8
- Sustainability plan ..................................................................................................... 9
  - Role of the Advisory Committee ......................................................................... 9
- Communications ........................................................................................................ 9
- Pistoia Alliance Conclusions ..................................................................................... 10
Executive Summary

The cBioPortal for Cancer Genomics is a web-based software system developed at Memorial Sloan Kettering Cancer Center (MSK). The software stores genomic data from large scale, integrated cancer genomic data sets, allows explorative data analysis, and provides simple download of subsets of data. The cBioPortal software was made available under the Affero GPL license in March of 2015, and all code is now available on GitHub at: http://github.com/cBioPortal/cbioportal. The software development project is now managed jointly as an open source software project by MSK and the Dana Farber Cancer Institute (DFCI).

With this transition towards an open source software project, the cBioPortal team recognizes the need to nurture a community of interest and assure the long term stability of the product. They have turned to the Pistoia Alliance for assistance in this transition, given its domain expertise with open source development, community building, and sustainability planning.

Several Pistoia Alliance Core members are either using cBioPortal or considering its use internally (the IP3 entry has 4 Core members in the interest community, as well as a Participating and a Contributing member). In addition, the Pistoia Alliance had another idea submitted through IP3 that focuses on PDX data, and suggests cBioPortal as the tool of choice. This business case was undertaken by the Pistoia Alliance in collaboration with MSK to articulate the MSK product and community plans for our members, and to identify whether or not to advance this idea as a possible Active Pistoia Alliance Project either around a Phase 1 for continued product development and/or around a Phase 2, for identifying and executing on a long-term sustainability plan or to have the developers take the project forward under their sole direction.

The ultimate recommendation coming out of this business case is that MSK is in the best position to take this project forward at this time under its sole control, and that the Pistoia Alliance move this project idea to HOLD status. A full fledged Pistoia Alliance project at this time is not feasible, given that the MSK development requirements do not lend themselves to an open RFP process, and given that their sustainability planning is two years out. Since Pistoia Alliance members are interested in cBioPortal, the Pistoia Alliance will sponsor a webinar on the subject on July 22, 2015, to expose a wider audience to cBioPortal, the product, its transition to open source, the current roadmap, and the plans for a corporate advisory committee. The Pistoia Alliance has also made relevant introductions, and provided high level outline for the long term sustainability planning that will be required. And finally, it is also, in a separate workstream, investigating the feasibility of the PDX project, which has dependencies on cBioPortal. If, over the next several months, MSK decides that they want to pursue further future development through an open RFP process, this business case could be reopened and reconsidered.
Project Aims

The cBioPortal now enjoys a large following in the cancer research community, with more than 1,000 unique users visiting the public website (http://cbioportal.org/) every day. In addition to the public site, the cBioPortal software is now running in at least 30 academic research institutes or hospitals to host private research or clinical sequencing data.

With the move to an open source project, the cBioPortal developers want to establish a collaborative software development community. Going forward the cBioPortal software development will be tightly coordinated between at least three academic sites:

1. Memorial Sloan Kettering Cancer Center: Nikolaus Schultz, Benjamin Gross, JJ Gao
2. Dana Farber Cancer Institute: Ethan Cerami
3. Princess Margaret Hospital: Trevor Pugh, Stuart Watt

This group of developers is currently planning the details of future development, which will focus on the main areas: 1) software architecture improvements, to ensure software stability, performance, and ease of modification; 2) new features for the analysis of cancer genomics cohorts and visualization of genomic profiles from individual patients for use in clinical care; and 3) building and maintaining of data pipelines and manual curation of relevant cancer genomics data sets.

In order to achieve these tasks, which will enable more insights from the rapidly growing body of cancer genomics data and enable precision cancer medicine, the cBioPortal team requires additional funding to hire more full time developers (Phase 1) and must determine a long term sustainability plan (Phase 2).

MSK wishes to maintain oversight over the initial Phase 1 development to assure the continued utility of the program for its own and collaborator’s research efforts.

Because of the above requirements, we have concluded that a Pistoia Alliance project is not feasible at this time, as the above requirements do not allow for an open RFP process. Instead, during Phase 1, the cBioPortal developers are proposing through this business case to establish an Advisory Committee to prioritize enhancements and further development of the cBioPortal, with particular focus on commercial pharma needs. The cBioPortal development team is inviting members of the pharmaceutical industry to join the Advisory Committee for a period of two to three years. An annual membership (fees outlined below in Table 1) will allow the sponsoring companies a seat on this committee, which will shape and prioritize the ongoing and future cBioPortal roadmap, as well as give them unrestricted access to full download of all data available in the public instance of cBioPortal (cbioportal.org). The membership monies will be used to fund additional developers of the cBioPortal dedicated to improvements outlined in the Roadmap, help develop new data pipelines, and allow for prioritization of future enhancements to the cBioPortal. There is also a proposal to include a data-only tier.
Phase 2 involves defining an infrastructure that will support cBioPortal to ensure long term sustainability and continued growth. There are several possible directions the cBioPortal could take such as creating a stand alone foundation, merging with an existing foundation or organization, or continuing with an Advisory Board, adding additional members in a tiered membership structure. The cBioPortal team will look to the Advisory Committee established in Phase 1 to make recommendations on the best position of the cBioPortal for continued growth, development and use for research within academia and industry. The work to implement the resulting plan could be awarded to an outside entity through an RFP process or it could be managed internally by MSK, DFCI, etc.

**Deliverables and milestones**

*Mid-July 2015*: Finalize business plan and roadmap in advance of webinar
*July 22, 2015*: Present Pistoia Alliance-sponsored Webinar
*End of July 2015*: Solicit feedback from webinar attendees
*August 2015*: Post Pistoia Alliance version of business plan on IP3.
*September 2015*: MSK to solicit membership to the commercial advisory committee, providing copies of the business case and developmental roadmap to support its request.
*January 2016*: Launch commercial advisory team
*Summer 2016*: Identify possible sustainability partners and build a preliminary long-term sustainability model for review by the advisory committee. Fall 2016: Finalize and implement sustainability plan.

**Competitor analysis**

There are several public and private tools available on the market that aim to answer similar questions to those answered by cBioPortal. Private efforts include NextBio, OmicsSoft, and Oncomine. Oncomine and NextBio have freely accessible components, however, their better analysis tools are only available for a fee. Users must pay a fee for using any aspect of software products offered by OmicsSoft.

There are many public tools that do bits and pieces of what cBioPortal can do: IGV, UCSC Cancer Genomics Browser, Regulome Explorer, GeneSpot, IntoGen, Tumor Portal, but none seem to have same spectrum of tools and visualizations, ease of use, or popularity of cBioPortal.

Among freely available tools, the TranSMART knowledge management platform probably comes closest to the functionalities of cBioPortal. While the data layer of cBioPortal is not as sophisticated as that of TranSMART, its focus is on visualization and ease of use.
Advantages of cBioPortal

- Long track record (7+ years)
- Many users in academia and industry
- Many local private installations
- Developed by researchers for researchers
- Focus on visualization and instant results (biologist friendly)
- Easy to use, existing users can train new users

How will we compete or collaborate with the competition?

There is no obvious way to collaborate with the developers of the competing commercially available products. However, there are many opportunities for collaboration with the academic developers of the freely available tools listed above. For example, cBioPortal already utilizes the IGV software for the visualization of more detailed sequencing and copy-number data. We also have plans to explore a potential integration with the tranSMART Foundation down the road in Phase 2 of this project when we are exploring long term sustainability plans for the cBioPortal.

Market assessment

The public cBioPortal site is accessed by more than 5,000 unique users per week, coming from academic institutions and biotech and pharma corporations from across the world.

![Growth chart of unique weekly visitors to cbioportal.org](image)

**Figure 1: Growth chart of unique weekly visitors to cbioportal.org**

In addition, private installations of cBioPortal are used in at least 30 research institutes or hospitals to support research and clinical sequencing efforts (Weill Cornell, Columbia, NYU, NY Genome Center, British Columbia, University of Michigan, CHOP, Princess Margaret, SickKids, Vanderbilt, Emory, UNC, University of Pittsburgh, CRUK, EMBL, Charite Berlin, institutions in Japan, China).

We have been approached by several large pharma companies as well as smaller biotech or sequencing service companies with an interest in getting a local installation of the software. One key request by these companies is the availability of service and support for the installation and maintenance of the software. The Hyve, a bioinformatics service company with offices in the
Netherlands and in the USA, has recently begun to offer commercial support for cBioPortal software installation and data feeds, and have built some integrations between cBioPortal and tranSMART.

Benefits and Cost

Establishing an Advisory Committee model

Companies that want to analyze genomic data usually have two options: Build internal infrastructure or purchase software solutions from outside vendors. Both options can be costly and risky, since not all internal software is guaranteed to work well and outside vendors may change scope or pricing. The cBioPortal software can take the place of these two options. The software and data are freely available, and both can be modified and expanded to meet a company's needs. The cBioPortal system has been under development for seven years and has evolved into a streamlined and user-friendly system that is used by more than one thousand researchers every day (cbioportal.org, not counting local installations of the software). There are now approximately 8 full-time developers working on cBioPortal software code at the three institutions listed above, and an estimated 30 developer years have gone into the software over the years. With the funding from this effort, we expect to increase the number of development hours going into the project due to the involvement of others. We plan that all membership fees will go towards salary support for software developers at the three collaborating institutions.

Benefits of joining the advisory committee

Pharma companies that contribute to the project will have the ability to advise on future enhancements to the software by having a seat on the Advisory Committee. Their needs can be prioritized by supporting the development through 4-5 FTEs. In addition, Advisory Committee members get access to all data hosted in the public instance of cBioPortal and can make suggestions on which data sets to acquire / import next.

Benefits to end-users

The cBioPortal is attractive to users from industry and academia because of its easy visualization and analysis of private data. Broad support already exists for the system and by establishing an Advisory Committee we are creating an infrastructure for sustainability and the rapid development of new features outlined in the Roadmap. The cBioPortal has grown beyond the original developers' ability to keep up with user demand and this project will ensure that growth continues strategically and at a pace that meets our users' expectations.
Membership Structure

cBioPortal development is currently supported by federal grants and philanthropic funds, as well as institutional support. These funds support some level of software development, but as the cancer sequencing is steadily growing, they are increasingly used to build and support institutional data pipelines and database connections. To keep up the pace of development of the cBioPortal software, additional software developers are needed who can focus on software architecture and feature development.

We estimate that in order to achieve all deliverables outlined in the development roadmap document in the next two years, we will need to hire and fund an additional 4-5 full time software developers. This translates into ~$500,000. The cost for an annual developers meeting including travel expenses for all participants is estimated at $100,000. In addition, for this model to succeed, the developers will require strong project management to establish and run the Advisory Committee. That cost estimate is around $25,000. So total annual costs would be approximately $625,000. These funds would be solicited from the commercial members via the membership fees listed below.

<table>
<thead>
<tr>
<th>Member Type</th>
<th>Cost</th>
<th>Membership Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic membership</td>
<td>No fee</td>
<td>Access to all data</td>
</tr>
<tr>
<td>Commerical membership</td>
<td>$100,000 / year for 2-3 years</td>
<td>Advisory committee seat, access to all data, requests for data set curation</td>
</tr>
<tr>
<td>Commercial data subscription</td>
<td>$30,000 / year</td>
<td>Access to all data</td>
</tr>
</tbody>
</table>

Risks

A key risk is a potential for a lack of trust that the commercial funding of cBioPortal will lead to significant improvements of the software. The development working group will have to demonstrate a vision, a detailed plan of execution, an existing and operational team, as well as communication infrastructure (e.g., weekly developer conference calls) that ensures that the plan is executed.

Competition poses only a small risk at this point. While several companies operate in the space of cBioPortal, they don't currently have the full spectrum of features that cBioPortal has, nor do they have the intuitive user interface and broad user base. Existing sites run by academic institutions, such as tumorportal.org, probably don't have the manpower to improve and expand their features as quickly as needed in order to compete with cBioPortal. Furthermore, they are currently not available as open source software, making it impossible to run local versions with custom data.
A relatively small risk is the general lack of familiarity with the Affero GPL license that was chosen for the software. However, since this license is a modern variant of the frequently used GPL license (its main difference is that redistribution over a network also triggers code contribution requirements), this may not be an issue.

**Sustainability plan**

In the first phase of this project, we are trying to find a handful of commercial life sciences companies who will fund continued development at the official cBioPortal development sites for a period of 2 years. Once the Advisory Committee is established, we will define future sustainability options. Options here include the continuation of the existing advisory committee, merging with an external project or organization (e.g., tranSMART), or setting up a standalone foundation. We will work closely with the existing commercial members and other companies with an interest in joining to determine the direction and establish the infrastructure most likely to foster long term success.

If no further external funding materializes, we anticipate that the existing software will still be maintained through funds of academic institutions and hospitals that rely on it for research and patient care, such as MSK, Dana Farber, and Princess Margaret. Further software development will then happen at a slower pace but may be supported through grants or philanthropic sources.

**Role of the Advisory Committee**

The Advisory Committee will consist of one member from each of the commercial partners and the collaborating academic institutions. The role of the committee will be to give advice on the cBioPortal planning and software development, specifically:

- review and critique the features described in the roadmap
- suggest new features for development
- prioritize features
- prioritize data curation
- advise on long term sustainability

The Advisory Committee will have quarterly phone meetings.

**Communications**

cBioPortal related communication has to reach three different audiences: 1) Users, either of the public site at cbioportal.org or local installations at academic institutions or companies, 2) system
administrators and installers of local versions of cBioPortal in academic and commercial settings, and 3) software developers who contribute to the cBioPortal code that is hosted on GitHub. We reach these groups through several different mechanisms:

- **cBioPortal.org website.** Approximately 1,000 unique users visiting cBioPortal.org every day - this site will be used to announce new features, data, and the general availability of the software.
- **Google Group.** We have an active message board in Google Groups, which users can turn to for any questions about cBioPortal features and data on the public site.
- **Twitter.** The cBioPortal Twitter handle @cbioportal currently has over 300 followers.
- **Mailing list.** We have a mailing list with over 300 subscribers.
- **cBioPortal GitHub page.** We will use this page to communicate with cBioPortal software developers.
- **cBioPortal Slack channel.** We have recently set up a cBioPortal Slack channel, which is intended to be used by developers and system administrators to get quick answers to their questions.
- **Conferences.** cBioPortal is frequently mentioned at academic conferences - a large fraction of presentations at these meetings include graphs generated by cBioPortal. For example, at the recent ASCO Conference both MSK and Dana Farber showed data from their current clinical sequencing cohorts in cBioPortal.

**Pistoia Alliance Conclusions**

- While the established requirements do not lend themselves to creation of a Pistoia Alliance project, MSK is in a good position to move this project forward according to their specifications by the creation of an advisory committee, with each member paying annual fees for participation. This committee would manage the development roadmap for 2-3 years, during which time a strong open-source development community can be nurtured.
- A truly open source community will allow for broad external code contributions, as well as an infrastructure that can review and track bugs, etc, and define “official releases” of the cBioPortal software. A long term sustainability plan will need to be implemented to reach this goal.
- Given the 2-3 year horizon before implementation of the sustainability plan, Pistoia Alliance at this point is also not able to set up a project around establishment of such a plan, because the time lines are outside our project timeline requirements. However, we have made introductions to other organizations that could play a role in the longer term plan.
- Pistoia Alliance recommends that MSK strongly consider collaboration with an existing foundation when it seeks a long-term home for cBioPortal, because we see a likely consolidation of such foundations in the future, and difficult times ahead for niche, single-product organizations. As part of the sustainability planning, we recommend that MSK conduct a survey of possible like-minded organizations should be pursued as part of the sustainability planning process.
• Since Pistoia Alliance’s members are interested in cBioPortal, the Pistoia Alliance is pleased to sponsor a webinar on the subject on July 22, 2015, to expose a wider audience to cBioPortal, the product, its transition to open source, the current roadmap, and the plans for a corporate advisory committee.

• Pistoia Alliance will continue to investigate the feasibility of the PDX idea submitted to IP3, which has dependencies on cBioPortal.

• MSK has indicated some interest in the RFP development model down the road. We recommend that the project idea be put on Hold, allowing for reconsideration in this event.