Information Management Using Web 2.0 Technology

By

Juliet Duffy

Bachelors of Science in Chemical Engineering, University of Colorado at Boulder, 2002

Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
AND
Master of Science in Chemical Engineering

In conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology
June 2009

Signature of Author _____
May 8, 2009
Chemical Engineering, Sloan School of Management

Certified by _____
Kristala Jones Prather, Thesis Supervisor
Assistant Professor Kristala Jones Prather, Department of Chemical Engineering

Certified by _____
Roy Welsch, Thesis Supervisor
Professor of Statistics and Management Studies, Sloan School of Management

Accepted by _____
William Deen, Professor of Chemical Engineering
Chairman, Committee for Graduate Students Chemical Engineering

Accepted by _____
Debbie Berechman
Executive Director of MBA Program, MIT Sloan School of Management

This page has been intentionally left blank

Information Management Using Web 2.0 Technology

By
Juliet Duffy

Submitted to the Sloan School of Management and the Department of Chemical Engineering Division on May 8th, 2008 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Chemical Engineering

ABSTRACT

Web 2.0, the ultimate platform for tacit based knowledge work has finally arrived. User driven, collaborative platform based tools including wikis, web mash-ups, discussion boards, linkage based search engines, and tagging have the potential to vastly change how information is managed and how knowledge work is captured. This thesis investigates how the new paradigms and tools of Web 2.0 can be applied to the Pharmaceutical Industry and assist with information management at The Novartis Institute for BioMedical Research (NIBR). Applying Web 2.0 tools to NIBR's chemical compounds, targets, assays, people, and projects in a well thought out framework has the potential to yield tremendous productivity improvements in the drug discovery process.

Effectively harnessing the collective intelligence of thousands of scientists within Novartis's worldwide research network will enable a paradigm shift. A large, extremely knowledgeable user community can more effectively annotate metadata, hyperlink to important content, establish tags, and collectively author content. Such activities will not only improve the search ability of information but also allow important scientific connections to emerge linking biology to chemistry and furthering Novartis's understanding of disease.

Thesis Supervisor: Kristala Jones Prather
Title: Assistant Professor of Chemical Engineering

Thesis Supervisor: Roy Welsch
Title: Professor of Statistics and Management Sciences

Acknowledgements

Firstly, I would like acknowledge the Leaders for Manufacturing Program for its support of this work. Secondly, I would like to thank all of the individuals at Novartis Institutes of BioMedical Research for their support. My thesis would not have been possible without all of you. Greg Paris, Sepp Scheiber, Rajiv Chopra, Rishi Jain, Ivan Cornella Taracido, Jeremy Jenkins, Steve Cleaver, Marek Nowakowski, Jay Knowles, Detlev Biniszkiewicz and everyone that I interviewed was extremely generous of their time, provided tremendous insight, and was and willing to patiently explain the drug discovery process. Many thanks to Bill Egerton, Lazslo Urban, and Dmitri Mikhailov for teaching me about compound discontinuations and helping me put together case studies. A special shout out to my officemate Charles Snow, for showing me the ropes, answering all of my silly questions, assisting with information gathering and keeping me company on the quiet sixth floor.

Special thanks to my supervisor Dejan Bojanic for his belief in me and endless support throughout this project. I admire your visionary leadership and perseverance in introducing Web 2.0 into the organization.

Thank you to my thesis advisors, Kristala Jones Prather and Roy Welsch for all their insight and guidance. I was honored that you always made time for me in your schedule, no matter how busy things were. I especially appreciated your support throughout the thesis publishing process. Additionally, I would like to thank MIT and Harvard faculty members Tom Allen, Erik Brynjolfsson, Alex Pentland, and Andrew McAfee for taking the time to sit down with me and share their insights on social networking, collective intelligence, and Web 2.0.

Lastly, I'd like to thank the most important person in my life Brede Wegener who so very patiently encouraged me to keep working on this thesis, despite the excitement of our recent engagement.

This page has been intentionally left blank

Table of Contents

Table of Figures

This page has been intentionally left blank

Chapter 1: Thesis Overview

1.1 Overview

Pharmaceutical drug discovery is a very risky and expensive endeavor. The R&D process for bringing a new drug to market can take ten to fifteen years to complete and cost up to USD 1.0 billion/drug. On average, only one in 10,000 originally synthesized compounds will become a commercially available drug.[1] Furthermore, the industry is struggling under looming patent expirations, price controls imposed by governments and insurance companies, and rapidly rising drug discovery costs. Given these current challenges, it is imperative that the industry examines ways to increase the efficiency of the drug discovery process and reduce costs associated with discovering new chemical entities.

1.2 Problem Statement

In a 2004 white paper addressing industry productivity challenges the U.S. Food and Drug Administration commented "if biomedical science is to deliver on its promise, scientific creativity and effort must also focus on improving the medical product development process itself."[2] Information management is one aspect of the product development process that is often neglected. Indeed, effective sharing of data has been specifically identified by the FDA as one of the actions companies can take to improve product development.[3] Information management is a tremendous challenge for pharmaceutical research organizations. Researchers are not only exploring and generating massive amounts of data about the vast realms of the human genome (20,000-25,000 genes), they are also trying to map that information to the chemical space, typically screening chemical libraries containing anywhere from1-5million compounds on a regular basis. The sheer volume and diversity of information that is generated is often overwhelming.

Information management challenges are not only constrained to the pharmaceutical industry. The internet and advent of digital storage has exponentially driven worldwide data proliferation. It is estimated that 4.0 exabytes ($4.0*10^{19}$) of unique information will be generated this year, more than

---

[1] Novartis Company Website, *About Novartis- Research and Development, n.d.* Web. Accessed 14 April 2009 http://www.novartis.com/about-novartis/corporate-citizenship/business-conduct/research-developements.shtml

[2] "Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products." *Food and Drug Administration (FDA)* (2004). Whitepaper.

[3] Stephens, Susie. *Enabling Semantic Web Inferencing with Oracle Technology: Applications in Life Sciences* SpringerLink Book Series Lecture Notes in Computer Science, Springer Berlin/ Heidelberg 3791 8-16 (2005)

in the previous 5,000 years. Additionally, the amount of new technical information is doubling every two years." [4]

How is this excessive data generation being managed on a worldwide scale? Interestingly, a new paradigm in information generation and management has arisen. Often referred to as "next generation web" and dubbed "Web 2.0" this powerful paradigm shift has profound implications for how internet users, companies, and individuals alike generate and manage data.

1.3 Thesis Overview

This thesis investigates how the new paradigms and tools of Web 2.0 can be applied to the Pharmaceutical Industry and assist with information management at The Novartis Institute for BioMedical Research (NIBR). Specifically, the project explores the concept of using web-based Compound Homepages, similar to an internal Wikipedia, to share information about chemical compounds.

The thesis begins in Chapter 2 by providing background information on Novartis and explores the productivity crisis faced by the pharmaceutical industry. The next section, Chapter 3, provides an overview of tacit (i.e. knowledge based) work and explains how firms that successfully manage tacit interactions can gain a significant productivity advantage over competitors. An overview of various Web 2.0 tools (wikis, blogs, tagging, etc.) is provided in Chapter 4 and specific case study examples of Web 2.0 usage within the life sciences industry are explored in Chapter 5.

Chapters 6 and 7 present the study methodology and results from user interviews conducted at Novartis. The user research explores employee attitudes toward Web 2.0 and provides an assessment of the current Novartis data environment. Chapter 8 provides a three-lens analysis and explores strategic, cultural, and political barriers the firm should be aware of while implementing change.

Chapter 9 highlights how Web 2.0 has the potential to vastly improve internal search, break down organizational silos by connecting employees, and classify and capture compound information through user-generated tags. Chapter 10 provides a specific web framework for capturing information related to a specific research project, while Chapter 11 briefly examines how Novartis

---

[4] Fisch, Karl *Living in Exponential Times*. The Kim Komando Show, n.d. Web Video. Accessed 17 March 2009 http://videos.komando.com/2008/12/22/living-in-exponential-times/ Original video source data: http://www.lps.k12.co.us/schools/arapahoe/fisch/didyouknow/sourcesfordidyouknow.pdf

captures information regarding compound terminations and addresses how the organization can learn from project failures.

Lastly, Chapter 12 explores three different web-based functionalities for a Compound Homepage web system: 1.) Web Portals – to unify diverse sources of information in one easy to find place 2.) Tagging- to allow the user community to classify compounds and explore scientific relationships between chemistry and biology, and 3.) Wiki environment – to establish a user community for sharing and capturing scientific information.

Chapter 2: Industry and Company Background

2.1 The Value of Pharmaceuticals

Pharmaceutical products are a major weapon in treating disease and add tremendous value to the health care system. According to Dan Vasella, the CEO of Novartis, the purpose of the pharmaceutical industry is to reduce the incidence of disease, reduce death associated with disease, improve quality of life, and reduce suffering. [5]

Indeed, the introduction of new and novel drugs for treating all types of diseases has had a dramatic impact on human mortality. From 1965 – 1999, pharmaceuticals have significantly lowered the death rates for individuals with rheumatic fever, hypertensive heart disease, and Ischemic heart disease (Figure 1).

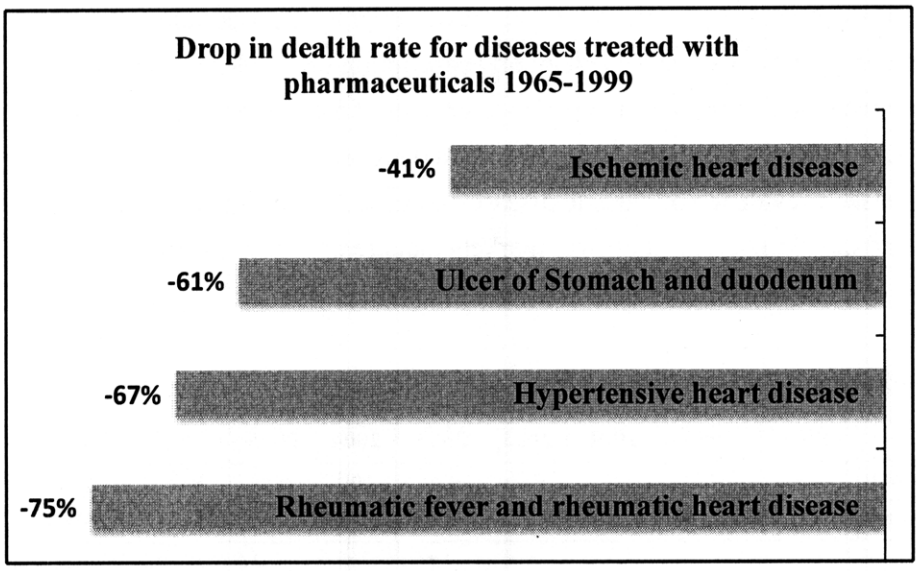


Figure 1: Medicinal Impact on Mortality (1965-199)[6]

Often pharmaceuticals are viewed as the most expensive component of healthcare and are unfairly blamed as the primary reason behind rising healthcare costs. A 2007 Price Waterhouse Coopers study showed that roughly 2/3 of consumers estimated that prescription drugs account for between

[5] Daniel Vasella, CEO & Chairman of Novartis, presentation to CEO Perspectives Class, MIT Sloan, Cambridge, MA 4 Feb 09. Lecture

[6] Ibid. (5) & *EFPIA 1999 – 2002*

40% - 79% of U.S. healthcare costs. This consumer perception was far from the truth. In reality, prescription drugs actually account for <10% of US healthcare expenditures.[7]

Dollar for dollar prescription medicines actually help to mitigate other healthcare expenditures. They decrease the need for hospitalizations, surgeries, nursing home admission and other costly healthcare alternatives. A 2005 study found that a 20% increase in patient adherence to taking prescribed pharmaceuticals yielded substantial savings. Every $1 spent on diabetes, cholesterol, and blood pressure medicines resulted in overall healthcare savings of $7.10, $5.10, and $4.00 respectively.[8]

2.2 The Pharmaceutical Industry

The pharmaceutical industry is a multi-billion dollar industry with total global sales of $712B USD in 2007[9]. The industry has experienced significant global growth over the past decade (Table 1). According to IMS Health, an agency that tracks 95% of all prescription drug sales in over 80 countries, the global pharmaceutical market will experience growth of 5-6% next year for total expected global sales of U.S. $735-745 billion.[10] Megatrends driving industry growth include an aging population, an increased prevalence of chronic disease associated with unhealthy lifestyles, scientific advancement, and developing world growth markets that are now able to afford pharmaceutical care. [11]

Table 1: Global Pharmaceutical Sales, 1999-2007[12]

	1999	2000	2001	2002	2003	2004	2005	2006	2007
Global sales (US$ billions)	334	365	392	428	499	560	605	649	712
Growth over previous year	14.5%	11.5%	11.8%	9.5%	10.3%	8.0%	7.3%	7.1%	6.4%

[7] "Putting Patients First – Meeting the Drug Discovery Challenge" *Novartis Institute for Biomedical Research, New Employee Orientation*. Cambridge, MA. June 2008 & PriceWaterhouseCoopers 9 Jan 07

[8] M.C. Sokol et al. "Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost." *Medical Care* 43.6: 521-530 (2005). Print.

[9] *Parexel's Bio/Pharmacuetical R&D Statistical Sourcebook,* PAREXEL International Corporation (2008/2009). Print.

[10] Ibid. (9)

[11] Daniel Vasella, CEO & Chairman of Novartis, presentation to CEO Perspectives Class, MIT Sloan, Cambridge, MA 4 Feb 09. Lecture.

[12] Ibid. (9)

The industry is relatively concentrated among a handful of large players. In 2007, the top ten pharmaceutical firms accounted for 51.3% of global prescription sales, with the largest firm, Pfizer, capturing 7.5% of the market (Table 2).

Table 2: Top 10 Companies by Worldwide Prescription Drug Sales in 2007[13]

Rank	Company	Rx Sales (2007) US$ billions	% Worldwide Market Share
1	Pfizer	42.6	7.50%
2	Sanofi-Aventis	38.8	6.80%
3	GlaxoSmithKline	37.8	6.60%
4	**Novartis**	**31.6**	**5.50%**
5	Roche	30.7	5.40%
6	AstraZeneca	28.4	5.00%
7	Merck & Co	25.8	4.50%
8	Johnson & Johnson	24.0	4.20%
9	Eli Lilly	16.9	3.00%
10	Abbott Laboratories	16.1	2.80%

2.3 Novartis Overview

Novartis is the 4th largest pharmaceutical company in the world, ranked by 2007 prescription drug sales (Table 2). The company was created by the 1996 merger of pharmaceutical companies Ciba-Geigy & Sandoz. The $27 billion deal was, at the time, the biggest industrial merger in history. The resulting combined company was named Novartis based on the Latin "novae artes", meaning "new skills".[14] In 2008 Novartis products treated over 850 million patients worldwide[15].

Novartis operations span beyond prescription pharmaceuticals with businesses in various health care areas. Daniela Vasella, the Chairman and CEO of Novartis describes the company's strategy as fulfilling four different customer needs: newer better drugs, affordable quality drugs, self-treatment without prescription, and disease prevention (figure below)

---

[13] *Parexel's Bio/Pharmacuetical R&D Statistical Sourcebook,* PAREXEL International Corporation (2008/2009). Print.

[14] "Putting Patients First – Meeting the Drug Discovery Challenge" *Novartis Institute for Biomedical Research, New Employee Orientation*. Cambridge, MA. June 2008

[15] Novartis Company Website, *n.d.* Web. Accessed 3 Feb 2008 http://www.novartis.com/

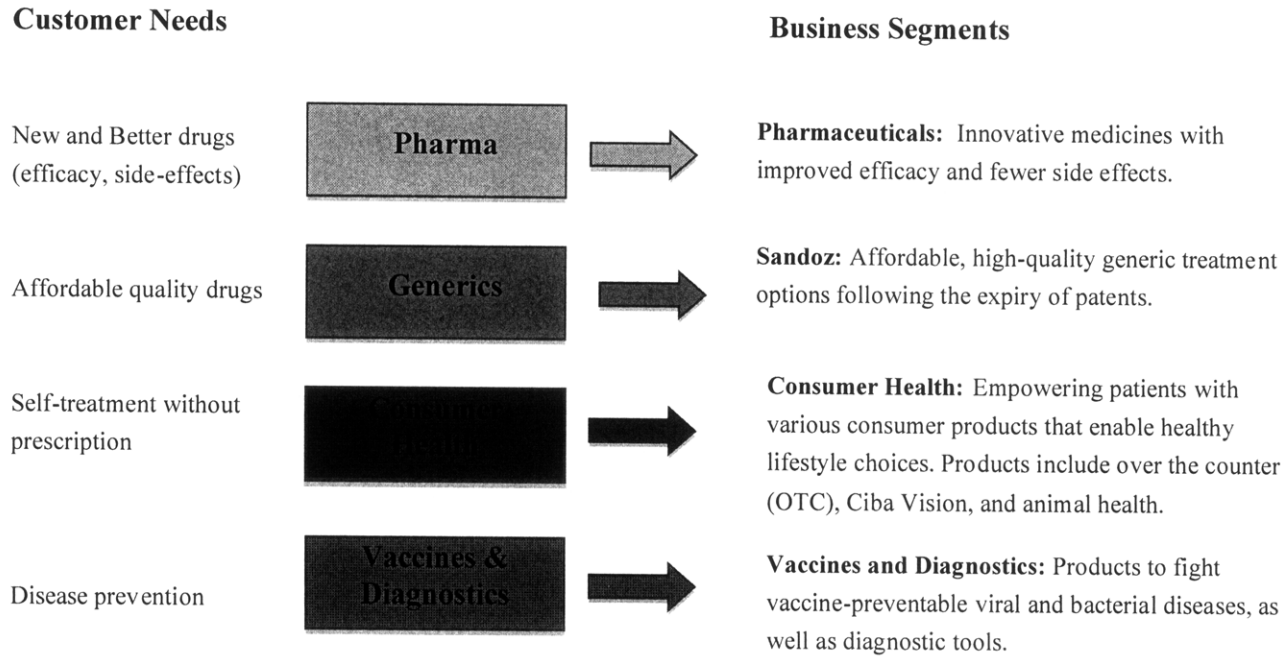


Figure 2: Novartis Business Divisions focused on Understanding Customer Needs[16]

2.4 Novartis Institutes for BioMedical Research (NIBR) Overview

The research for this thesis project took place solely within the Novartis Institutes for BioMedical Research (NIBR). NIBR is the research organization of Novartis and focuses on early stage drug discovery through pre-clinical development.

Research is De-coupled from Commercial Activities

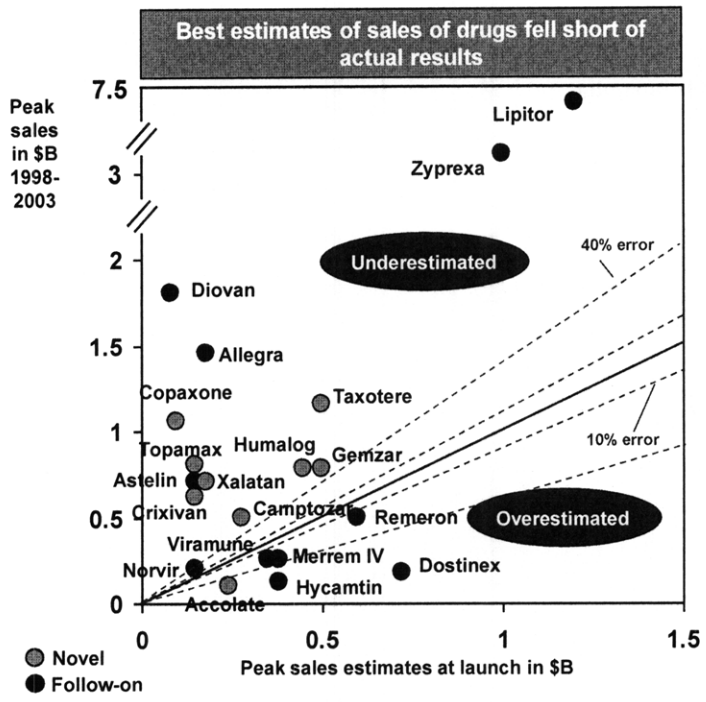
Prior to 2003, the Novartis research and discovery group was headquartered in Basel, Switzerland and reported directly into the Pharmaceutical Business Division. In 2003 Novartis made a deliberate strategic decision to separate the R&D entity from the commercial side of the pharmaceutical organization thus forming Novartis Institutes for BioMedical Research (NIBR).

The unique separation of the research division from its commercial counterpart gave the new research organization several advantages. Research headquarters were moved from Basel to Cambridge, MA in order to take advantage of geographic proximity to other global healthcare research centers. The geographic shift allowed the R&D center to tap into additional drug

[16] Daniel Vasella, CEO & Chairman of Novartis, presentation to CEO Perspectives Class, MIT Sloan, Cambridge, MA 4 Feb 09. Lecture.

discovery talent, establish collaborations with other intuitions, and gain biologics know-how for advancing its biotechnology portfolio.[17]

Secondly and most importantly, the separation gave the research division greater autonomy in portfolio prioritization decision making process. A study of the pharmaceutical industry by the Boston Consulting Group showed that commercial predictions of peak sales at the time of drug launch were often vastly inaccurate with little to no correlation to actual market sales (Figure 3). Additionally, actual commercial value was vastly under predicted for many novel chemical compounds. Revenue from Copaxone, a novel multiple sclerosis therapy, was under predicted by 10x, revenue from Topamax, a migraine medication, was under predicted by 5x (Figure 3). Novartis realized that commercial predictions were often inaccurate and felt that criteria beyond estimated commercial value should drive decision making within the R&D process.



Forecasting often missed by significant multiple

Drug	Error Multiple
Diovan	24x
Copaxone	10x
Allegra	8x
Lipitor	6x
Topamax	5x
Astelin	5x
Crixivan	4x
Xalatan	4x
Zyprexa	3x
Taxotere	2x
Camptozar	2x
Gemzar	2x

Figure 3: Forecasting Not Applicable for Portfolio Decisions[18]

[17] Elton, Jeff, NIBR COO speech to MIT Leader's for Manufacturing Class. *Novartis Institute for Biomedical Research.* Cambridge, MA. Aug. 2008

[18] Biniszkiewicz, Detlev - Portfolio Prioritization Research. *Boston Consulting Group*, 17 Feb 09

Therefore, NIBR established a new model for advancing research projects, one in which drug candidates are prioritized and advanced through the pipeline based on unmet medical need and an understanding of the underlying disease biology (mechanistic understanding). An ideal Novartis development candidate would rank both highly in unmet medical need and disease understanding (Figure 4).

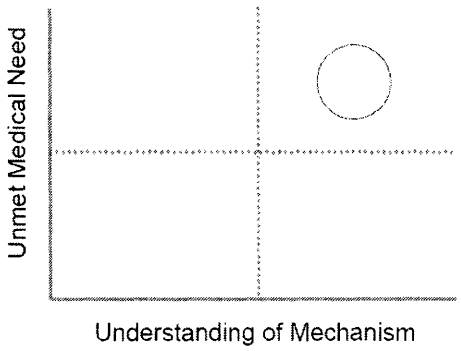


Figure 4: Novartis Model for Prioritizing R&D Portfolio

2.5 Industry Productivity Crisis

According to the U.S. Congressional budget office, the pharmaceutical industry is one of the most research intensive industries in the United States. "Pharmaceutical companies invest as much as five times more in research and development, relative to their sales, than the average U.S. manufacturing firm". Indeed, the 2007 average industry for R&D expenditures as a percentage of sales was 16.4%and R&D investment reached a record $58.8B USD."[19] Research and development costs are not only substantial; they have significantly ballooned over the past two decades. In 1987 PhRMA members spent $5.5B USD in R&D; today that spending has increased by 8x (Figure 5).

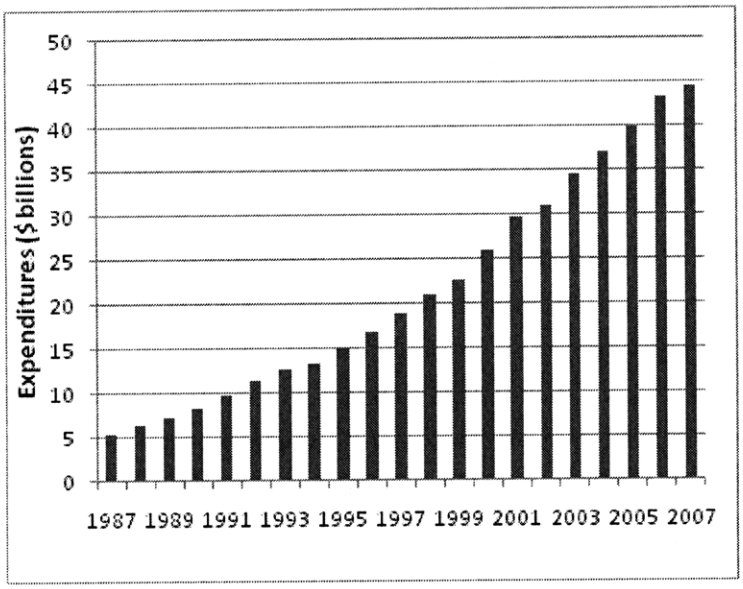


Figure 5: Total R&D Investment reported by PhRMA Member Companies1987-2007[20]

As research costs have risen dramatically, the number of new U.S. Food and Drug Administration new drug approvals (NDAs) has been relatively flat (Figure 6). The average number of NDA approvals per year over the last two decades (1987-2007) was 83.4. New drug approvals in 2001, 2002, 2003, 2005 and 2007 were below that average. Additionally, little productivity increase has occurred with new molecular entity approvals (NMEs), a subclass of NDA's representing drugs with a novel chemical structure. In 2005- 2007 NME approvals were below levels approved in the late 1980's (Figure 6).

---

[19] Pharmaceutical Industry Profile. *Pharmaceutical Research and Manufacturers of America (PhRMA)* Washington, DC (2008). Print.

[20] *Parexel's Bio/Pharmacuetical R&D Statistical Sourcebook,* PAREXEL International Corporation (2008/2009). Print.

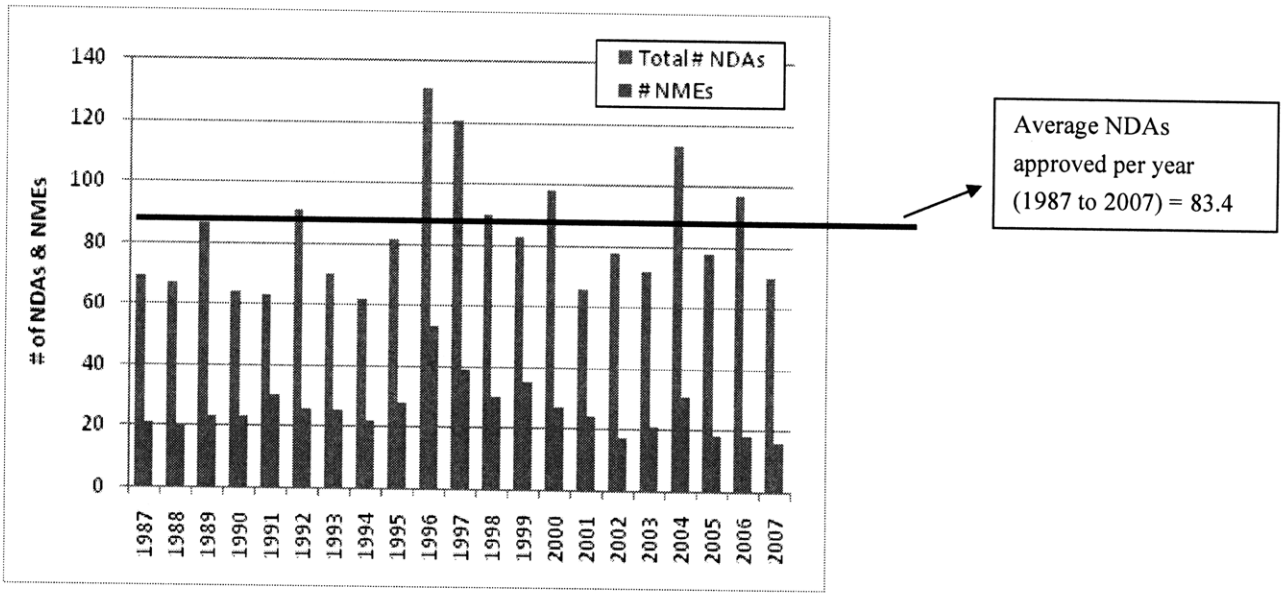


Figure 6: Number of NDAs and NMEs Approved by Year, 1987 - 2007[21]

Reasons for the decline in productivity have been examined extensively within the pharmaceutical community. A higher bar for regulatory scrutiny, increased expectation and cost of running clinical trials, and difficulty of translating highly academic research into effective therapies have all been cited. Regardless of the rationale, the industry is suffering from a productivity crisis. The crisis not only poses ramifications for the industry, but also healthcare systems around the world. Decreasing productivity means that the research cost associated with each successful drug approval continues to rise. The Tufts University Center for Drug Development estimates the cost of bringing a single new pharmaceutical drug to market at roughly $1.3billion (in year 2005 dollars). Furthermore, only 2 out of every 10 marketing drugs are able to produce revenues that match or exceed research and development costs. [22]

One of the factors contributing toward enormous R&D expenditures is the very low probability that a chemical compound ever reaches the marketplace. For every 5,000 – 10,000 compounds tested, 5 of those will reach clinical trials and only one will become an FDA approved drug (Figure 7). Furthermore, the R& D process is lengthy taking anywhere from 10 to 15 years to develop a single new drug. [23]

---

[21] *Parexel's Bio/Pharmacuetical R&D Statistical Sourcebook,* PAREXEL International Corporation (2008/2009). Print.

[22] Vernon, John, Joseph Golec, and Joseph DiMasi, "Drug Develoment Costs when Financial Risk is Measured Using the Fama-French Three Factor Model", (2008). Unpublished working paper. Submitted to the Journal of Health Economics.

[23] Pharmaceutical Industry Profile. *Pharmaceutical Research and Manufacturers of America (PhRMA)* Washington, DC (2008). Print.

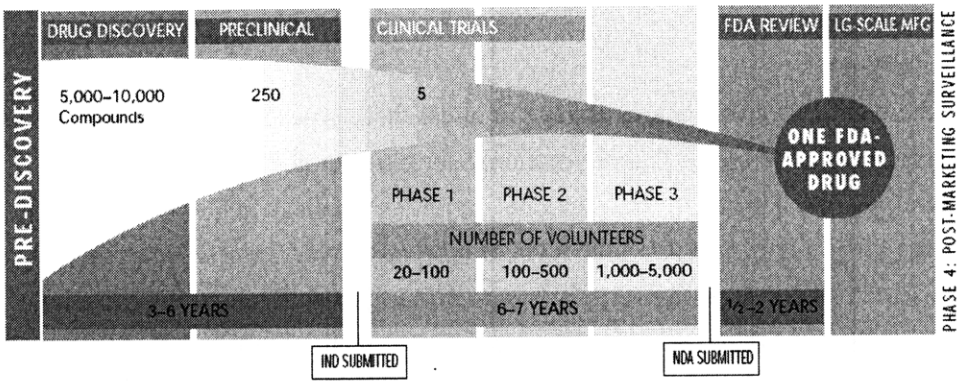


Figure 7: Probability of Success for FDA Approved Pharmaceutical[24]

2.6 Improving Information Management to help solve the Productivity Crisis

What can be done to reverse the industry productivity trend and help to jump start R&D productivity engines? A 2004 Food and Drug Administration (FDA) whitepaper focused on this issue. The paper points to importance of the product development process and making process improvements for bringing new drugs through the development pipeline. "If biomedical science is to deliver on its promise, scientific creativity and effort must also focus on improving the medical product development process itself."[25]

One aspect of the medical product development process is the effective storing and sharing of organizational information. Indeed, information management is one aspect of the product development process that is often neglected. Information management is a tremendous challenge for pharmaceutical research organizations. Researchers are not only exploring and generating massive amounts of data about the vast realms of the human genome (20,000-25,000 genes), they are also trying to map that information to the chemical space, typically screening chemical libraries containing anywhere from1-5 million compounds on a regular basis. The sheer volume and diversity of information that is generated is often overwhelming. Improvement to IT systems and data management would not only allow Novartis to be more competitive, it would allow knowledge to be more accessible throughout the organization having profound implications for solving the productivity crisis.

[24] Ibid. (23)

[25] "Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products." *Food and Drug Administration (FDA)* (2004). Whitepaper.

Chapter 3: Tacit Organizational Work

3.1 Definition of Tacit Work

McKinsey classifies the work that corporations engage in under three types of activities: transformational, transactional, and tacit. Transformational activities include the extraction or conversion or raw materials into a product. Transactional activities are classified as being routine and repetitive (think working on a production line). The third type of work, classified as tacit activities, is more complex. These activities rely on continuous collaboration with organizational colleagues, complex problem solving, making judgments, drawing on forms of information, etc. Essentially, tacit activities are those of so-called "knowledge workers."

Traditionally, firms have been engaged primarily in transformational and transactional activities. As these activities lend themselves to standardization and automation, historical gains in firm efficiencies have been focused on optimizing and improving these types of activities. The trend in work has now reversed with a large portion of work in developed countries now focusing on knowledge based tacit activities (Figure 8).

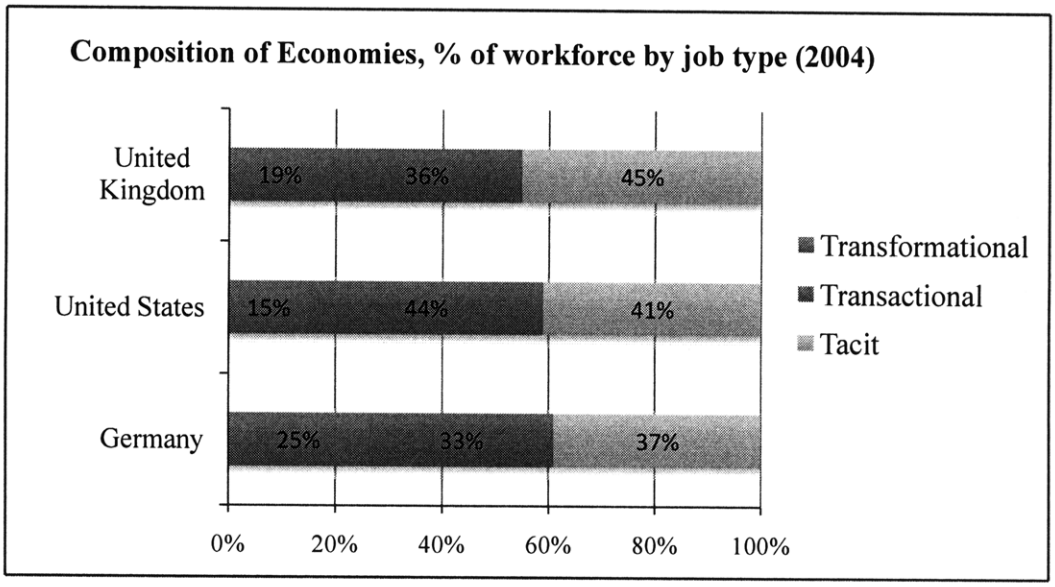


Figure 8: More Jobs Require Tacit Interactions[26]

[26] Beardsley, Scott, Bradford Johnson and James Manyika "Competitive advantage from better interactions" *McKinsey Quarterly* 2 (2006). Print

3.2 Tacit Work is Harder to Understand and Optimize

It is extremely important for companies to both understand the tacit knowledge work that takes place as well know how to boost the productivity of these knowledge workers. As improving tacit productivity is not as straightforward as standardizing a production line or leaning out a manufacturing facility, this is easier said than done. McKinsey's study of 8,000 US companies found that the performance of companies in heavily tacit sectors varied far more than that of other companies[27]. This suggests that companies and industries engaged in highly tacit activities have "significant competitive headroom"[28] for improving productivity in knowledge workers. Companies that master this activity will be able to gain a significant advantage over rival firms in the industry. According to McKinsey, "executives will have to learn how to compete, innovate and mange in an era when tacit interactions dominate and drive firm performance".[29]

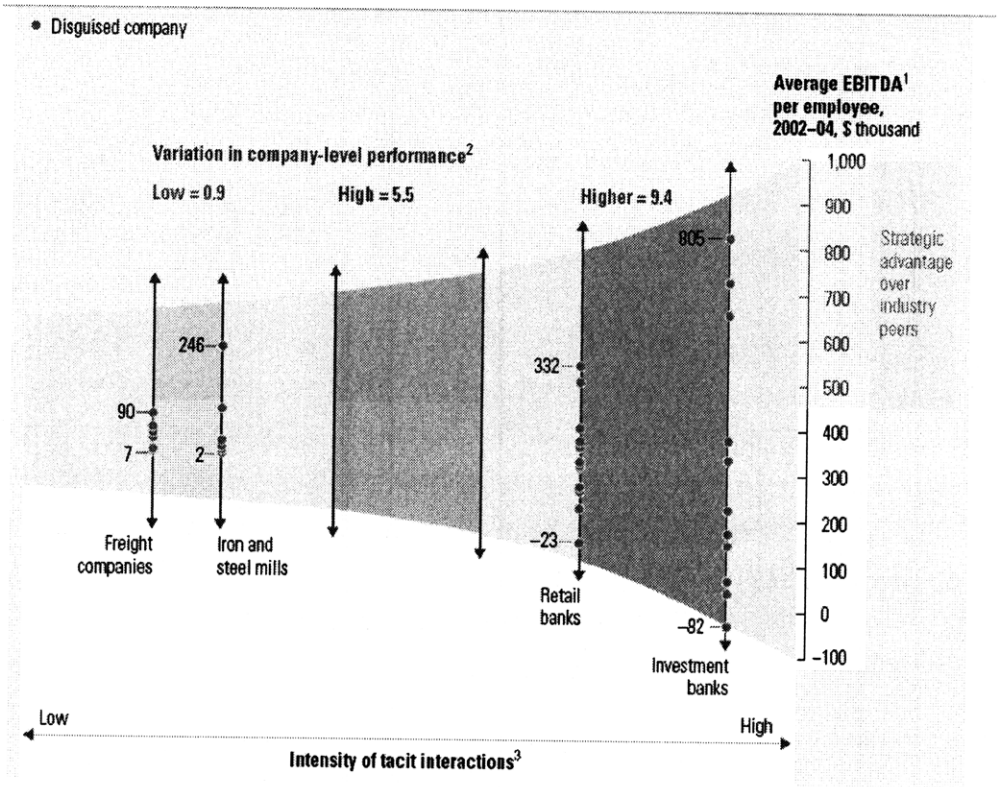


Figure 9: Performance Varies for the highly tacit[30]

---

[27] Beardsley, Scott, Bradford Johnson and James Manyika "Competitive advantage from better interactions" *McKinsey Quarterly* 2 (2006). Print

[28] Ibid. (27)

[29] Ibid. (27)

[30] Ibid. (27)

3.3 Leveraging Information Technology

As knowledge work is very dependent upon the sharing of information, collaboration, making judgments based on available information, tools that help to improve those activities can boost productivity of tacit organizations. Information technology (IT) is commonly cited as such a tool. In McKinsey's study, tacit-dominated organizations in the top 25% of labor productivity invested five times more in IT for their employees than firms in the bottom quartile.[31] Increasingly, companies are exploring a new generation of "Web 2.0" tools that encourage employee collaboration and a decentralized approach to information capture.

[31] Beardsley, Scott, Bradford Johnson and James Manyika "Competitive advantage from better interactions" *McKinsey Quarterly* 2 (2006). Print

Chapter 4: Web 2.0

4.1 What is "Web 2.0"?

"Do we finally have the right technology for knowledge work? Wikis, blog, group-messaging software and the like can make a corporate intranet into a constantly changing structure built by distributed, autonomous peers – a collaborative platform that reflects the way work really gets done." ~ Andrew McAfee, Web 2.0 Expert, Harvard Business School Professor

Appropriately, we turn to Wikipedia for a definition of Web 2.0:"Web 2.0 is the second generation of web development and design that aims to facilitate interconnectivity and interactivity of web-delivered content." [32] Web 2.0 technologies are essentially digital platforms that allow for content generation, sharing, and refining information over the internet.[33] Unlike the first generation of web development, often referred to as "Web 1.0", Web 2.0 focuses on increased flexibility and decentralization. A broad community of users is responsible for creating website content, rather than a single top-town source. Key differences between web generations are highlighted below (Table 3).

Table 3: Comparison of 1st and 2nd Generation Web Technologies

	Web 1.0	Web 2.0
Style	Centralized	De-centralized
Content generation	Created by a single source. "Top-down" flow of information	Content is created by community users. "Bottom-up" flow of information
Web page editing	Static web-pages that are non-user editable	Community constantly changes and updates pages

[32] "Web 2.0" *Wikipedia*, n.d. Web. Accessed 27 Feb 09 http://en.wikipedia.org/wiki/Web_2.0

[33] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

4.2 Examples of Web 2.0 Tools

In order to fully understand Web 2.0, it is useful to define some of the different types of Web 2.0 platform tools and explore specific examples of these tools that are currently being used over the World Wide Web.

Blog or Weblog- A website that is usually maintained by an individual author where posts and responses to those posts accumulate over time. Entries are usually posted in reverse-chronological order. Blogger.com describes blogs as:

> *"A blog is a personal diary. A daily pulpit. A collaborative space. A breaking-news outlet. A collection of links. Your own private thoughts. Memos to the world. Your blog is whatever you want it to be. There are millions of them, in all shapes and sizes, and there are no real rules".*[34]

Blog Examples:

- A Personal Blog: Harvard Business School Professor Andrew McAfee blog on The Business Impact of IT. http://andrewmcafee.org/blog/
- Technology Blog :Medgadget, Focused on emerging medical technologies: http://www.medgadget.com/
- Other Blogs - 2007 Weblog winners: http://2007.weblogawards.org/

Tagging - Simple descriptions called "tags" are used to identify and classify content. Items can be tagged with multiple tag descriptors. Categorization emerges over time as the result of user tagging.

Tagging Examples:

- Delicious– A social web bookmarking site where you categorize your bookmarks by tagging them with descriptors. You can share these categorizations and explore tag collections created by other users. http://delicious.com/
- Flickr, a photo sharing website, allows users to create tags in order to categorize content. Searching for "red dog" on Flicker yielded the photo below:

---

[34] "What's a Blog" *Blogger*, n.d. Web. Accessed 27Feb 09 http://www.blogger.com/tour_start.g

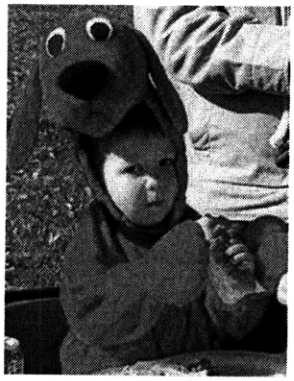


Photo was also tagged with the following "tag" descriptors:

- Red
- Halloween
- Hotdog
- Costume
- Matt

Figure 10: Photo obtained by searching Flickr for "red dog"

Wikis -"A wiki is a page, or collection of web pages, designed to enable anyone who accesses it to contribute or modify content."[35] Contrary to blogs, where contributions are usually by one author, wiki's are authored by multiple individuals within the community. The content is highly interactive with a constant re-write and editing process.

Wiki Examples:

- Wikipedia -(http://www.wikipedia.org/) an online, collaborative encyclopedia written by volunteer authors around the world. Wikipedia is one of the largest reference websites attracting over 684 million visitors in 2008. There are more than 75,000 active contributors working on more than 10,000,000 articles in more than 260 languages.[36]

RSS Feeds - Really simple syndication (RSS) allows users to keep track of changes on particular websites by "pushing" relevant information rather than "pulling" information (i.e. relying on a user to constantly revisit a website to see if new information has been posted). Software called "aggregators" periodically visit the subscribed to sites, and download new headlines. These headlines are then displayed together on the aggregator with links back to the website.

RSS Example:
Subscribe to any website that has an RSS icon ("")

---

[35] "Wiki" *Wikipedia*, n.d. Web. Accessed 27 Feb 09 http://en.wikipedia.org/wiki/Wiki

[36] "Wikipedia: About" *Wikipedia,* n.d. Web. Accessed 27 Feb 09 http://en.wikipedia.org/wiki/Wikipedia:About

Social networking software (SNS) - Social networking allows individuals to connect together in online communities. Connections can be based on shared activities and interests, professional associations, friendships, etc.

SNS Examples:

- Facebook- started as a limited community for university students to allow students to reach out and connect with others in their university. It has now grown into a global community with over 200 million active users[37], roughly the population of Brazil.[38] www.facebook.com
- LinkedIn- social network focused on establishing and maintaining professional relationships. Members can recommend the work of other individuals, keep track of current and former work colleagues, post resumes, etc. LinkedIn has over 36 million members in 200 countries. A new member joins LinkedIn approximately every second. [39]http://www.linkedin.com/ .

Mash-up- A webpage that aggregates information from a variety of different sources. For example, a mash-up page might pull information about a singer/songwriter artist from Wikipedia, a video of that artist from YouTube and a map on how to get to the latest concert from GoogleMaps.

Mash-up Example:

- Foxy Tunes -A music site from Yahoo Music. The Foxy Tunes mash up of the artist Bjork includes song lyrics pulled from Yahoo!, photos from Flickr, videos of Bjork on You Tube. You can even view ratings of the artists albums and buy them through an Amazon window.http://www.foxytunes.com/artist/bjork

---

[37] "Press Room Statistics" *Facebook,* n.d. Web. Accessed 2 May 09 http://www.facebook.com/press/info.php?statistics

[38] "List of Countries by Population" *Wikipedia*, n.d. Web. Accessed 21 April 09 http://en.wikipedia.org/wiki/List_of_countries_by_population

[39] "About Us" *LinkedIn*, n.d. Web. Accessed 27 Feb 09 http://press.linkedin.com/about

4.3 Web 2.0 in Business

Once solely the domain of the public internet, Web 2.0 is gaining widespread acceptance and usage within the business community. Companies are increasingly turning to Web 2.0 tools to help manage knowledge work. In June of 2008, McKinsey & Company surveyed two thousand companies around the world and asked about their Web 2.0 usage. 34% reported using blogs, 33% RSS feeds, 32% wikis, and 28% social networking tools (Figure 11).

Usage had not only increased from the previous year (2007), but companies reporting using a wider range of diverse tools. 2007 survey respondents said their companies had adopted just over two Web 2.0 tools on average. 2008 respondents indicated that they adopted 2.5 tools from the same list and more than three from an expanded one.[40]

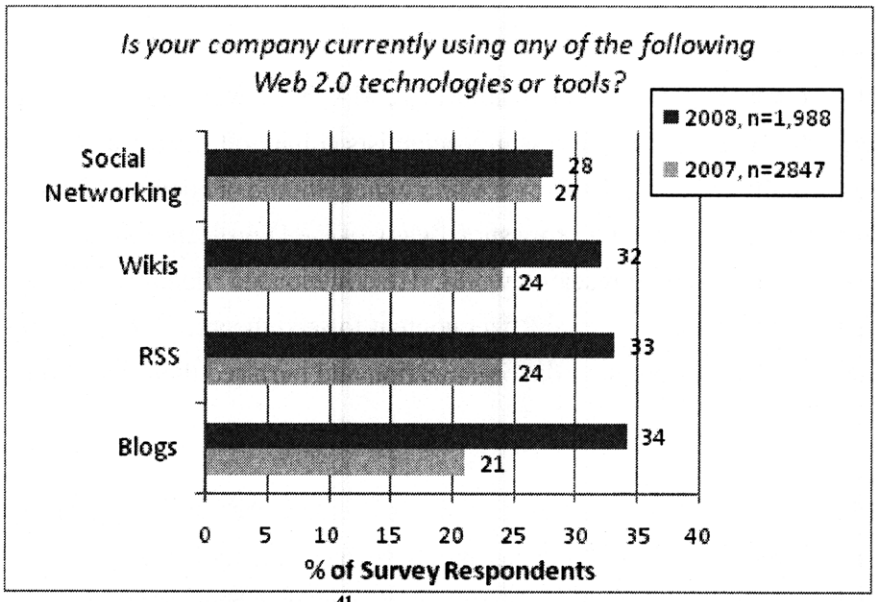


Figure 11: Web 2.0 Tool Usage[41]

In addition to just reporting the use of Web 2.0 technologies, some companies are reporting that the technology has started to transform the way that they organize and manage themselves. 16% of the McKinsey survey respondents stated that Web 2.0 has "created major new roles or function in our organization" while 14% reported that "it has changed the way our organization is structured."[42]

---

[40] "McKinsey Global Survey Results: Building the Web 2.0 Enterprise" *The McKinsey Quarterly* (2008). Print.

[41] Ibid. (40)

[42] Ibid. (40)

Chapter 5: Case Studies of Web 2.0 use in Life Sciences

Although examples of Web 2.0 use are plentiful in many industries and business settings, it is most relevant to examine specific uses within life sciences. Three specific case studies outlined below provide evidence that:

- Web 2.0 technologies are being adopted within the life sciences industry, in many cases by direct competitors to Novartis.
- User driven organizational and collaborative principles are providing value for gene annotation.
- Public data consolidation for chemical compounds via mash-ups, data portals, and hyperlinks has successfully allowed multiple data sources to be presented within one single system providing a one-stop shop for chemical information.

5.1 Case Study #1: Pfizerpedia (Wiki)

Pfizer's Research Technology Center (RTC) located in Cambridge, MA, has developed an internal encyclopedia dubbed "Pfizerpedia" as a tool to help research scientists collaborate, share and find information. Pfizerpedia was initially conceptualized by a Christopher Bouton, a computational biology group leader at Pfizer. Frustrated with a narrow siloed view of R&D projects, Chris wanted to get a larger system level overview of research world. He downloaded Media Wiki, the software used by Wikipedia, and posted some internal Pfizer content to seed the site.[43] Leadership subsequently helped promote the project throughout the organization and nurtured the Pfizer Web 2.0 phenomenon.

Available metrics point to Pfizerpedia's immense success. The site was launched at the beginning of 2006 and "within one year was receiving 12,000 hits per month by users across Pfizer worldwide."[44] The internal wiki currently has 10,000 different articles and has been adopted throughout the organization to the extent that "even Pfizer's cautious regulatory affairs group is using the wiki."[45]

John Castledine, the Director of Learning and Development for Pfizer's Global Research and Development (PGRD) Division comments "For organizations that need to create and nurture an innovative culture, the development of an internal 'wiki' site can be an important element. It is

---

[43] Mullin, Rick "Seeing The Forest At Pfizer" *Chemical Engineering News* 85.36 (2007):29. Print

[44] Ibid. (43)

[45] "Enterprise 2.0 Attendees Talk Strategy" *Information Week*, 14 June 2008, Web. Accessed 18 Feb. 2009 http://www.informationweek.com/news/internet/social_network/showArticle.jhtml?articleID=208403742

certainly the case at Pfizer, where increasing evidence points towards the usefulness of Pfizerpedia in enabling our employees to share and access knowledge more quickly than before. The ability to publish freely attributed information on line can help overcome any tendencies that may exist towards 'silo protectionism' or a bureaucratic approval process." [46]

Pfizerpedia works in conjunction with a number of other Pfizer IT tools, rather than replacing them. Scientific data repositories and documents still reside on document management systems with security authorization, the wiki just provides an easy way to link to these. Furthermore, Pfizerpedia has been an important tool in helping to direct intranet traffic to other Pfizer discussion groups and blogs where more detailed scientific collaboration and discussion takes place on specific topics.

The Pfizer case study is interesting in that it points to several things:

- **Researchers are interested in and willing to share information**-This is especially apparent given that Pfizerpedia is run off of Media Wiki software. Often the use of wiki-markup language is cited as barrier for wiki adoption. Pfizer scientists were so interested or enticed to share information that they were willing to overcome the Wiki-markup language barrier and invest the time to learn it.
- **The tool is useful**- With over 12,000 hits per month Pfizer employees are visiting the Pfizerpedia site because it is directly providing value to them.
- **Quality issues are minimal**- Pfizer has official "curators" that maintain the website and remove inappropriate entries. According to Pfizer employee and wiki creator Chris Bouton, "a very minimal amount of that work is done. It just isn't needed".[47] The community has been able to successfully establish norms and rules of behavior and encourage appropriate behavior.

[46] "Pfizerpedia: knowledge repository at Pfizer" *Chartered Institute of Personnel and Development, n.d.* Web. Accessed 21 Feb 09 http://www.cipd.co.uk/helpingpeoplelearn/_pfzrpd.htm

[47] Mullin, Rick "Seeing The Forest At Pfizer" *Chemical Engineering News* 85.36 (2007):29. Print

5.2 Case Study # 2: Gene stub page generation on Wikipedia (Wiki)

In 2008 five scientists, supported by the Genomics Institute of the Novartis Research Foundation (GNF), undertook an effort to engage the scientific community to create a worldwide wiki community, specifically for annotating gene function. Rather than create a new wiki site, the scientists decided to leverage off of the existing Wikipedia platform and add new Wikipedia pages for genes not already represented.

Although established gene portals and model organism databases (e.g. Entrez Gene, Ensemble, Mouse Genome Database) contained useful tools for referencing gene function, the scientists felt it was important to create a format that allowed for flexible, freely edited articles.[48] Furthermore, project organizers felt it would be beneficial to leverage the power of the collective community, relying on small contributions from countless different users around the world.

Rather than wait for Wikipedia pages on various genes to emerge over time, project organizers hypothesized that they could accelerate the process by constructing "stub" pages (Figure 12) as building block templates and a starting place for further annotation. Stub pages contained a sidebar with external gene identifies, links to other primary source databases, and gene expression patterns. Relevant literature references were also harvested from Entrez Gene). After a basic stub page format was established, stub pages were constructed automatically on the Wikipedia site using an automated Java based computer program.

[48] Huss III, Jon, et al. "A Gene Wiki for Community Annotation of Gene Function" *PLoS Biology, Public Library of Science* 6.7:e175 (2008). Print.

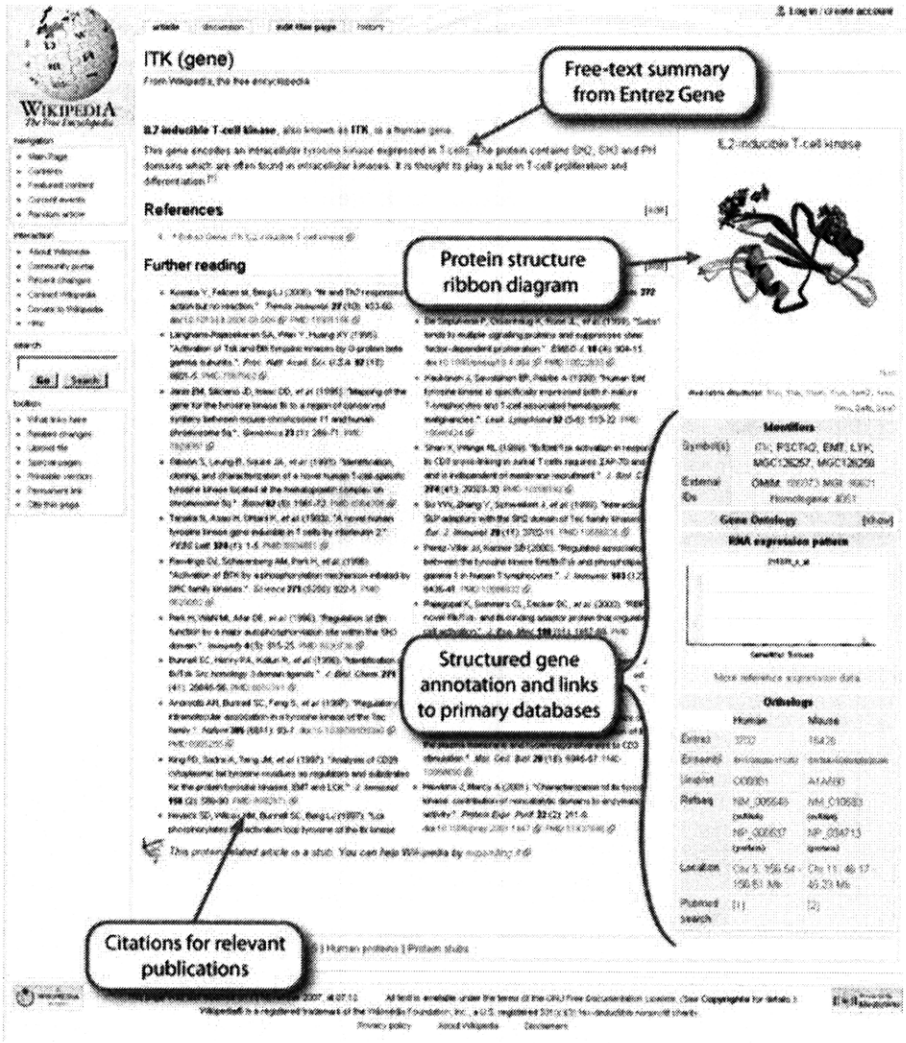


Figure 12: Example of Wikipedia Gene page stub[49]

As of February 2008, 7,500 new gene stubs were created and 650 existing Wikipedia pages were amended to include additional gene content. In the 15 weeks following template addition, Wikipedia page logs were tracked to determine if the stub page additions had in fact accelerated content creation. In terms of absolute numbers of edits, gene annotation activity on Wikipedia doubled. [50]Furthermore, Google search engine rankings of genes were examined both before and after stub page addition (Figure 13). The dramatic distribution shift toward higher page rankings for individual genes suggests that World Wide Web users are already starting to link to the

[49] Huss III, Jon, et al. "A Gene Wiki for Community Annotation of Gene Function" *PLoS Biology, Public Library of Science* 6.7:e175 (2008). Print.

[50] Ibid. (49)

Wikipedia gene pages feeding a cycle of additional viewing exposure and enforcing a positive feedback loop of content generation (Figure 14).

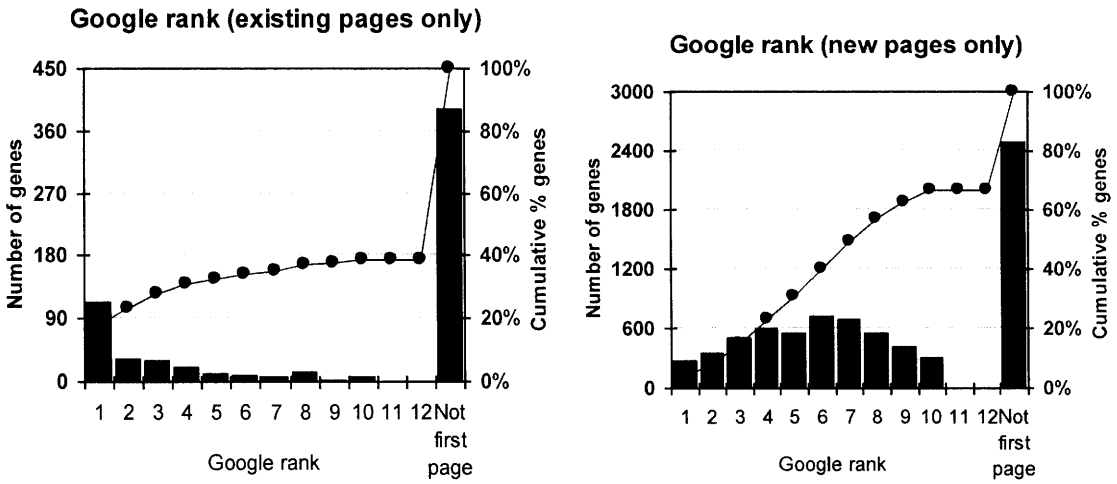


Figure 13: Google page ranking before and after creation of additional stub pages[51]

Figure 14: Additional viewing exposure reinforces positive feedback loop

[51] Huss III, Jon, et al. "A Gene Wiki for Community Annotation of Gene Function" *PLoS Biology, Public Library of Science* 6.7:e175 (2008). Print.

The Gene stub page generation case study highlights several important points:

- **Mash-up's successfully aggregate scientific information.** The gene pages on Wikipedia are similar to "mash-ups" aggregating information from a variety of sources (e.g. Entrez Gene, Protein Data Bank, etc.)The wiki platform works synergistically with these existing reference sources rather than replacing them.
- **Pre-existing structure is important.** Individuals are more likely to add information if a wiki page already exists. The activation energy of creating a new page is much higher than adding a simple edit.
- **Positive feedback loops are non-trivial.** The ability to leverage positive feedback loops will play a big role in obtaining additional exposure to the user community and help ensure the gene project's success.

5.3 Case #3: ChemSpider (Mash-up)

ChemSpider is a chemistry search engine that contains compiled data on more than 20 million compounds. The website serves as an aggregator, pulling chemical structure information from a variety of sources and compiling it in one easy to search data portal. ChemSpider is part Web2.0 mash-up and part information portal. External information is both integrated directly into the website (mash-up) as well as simply linked back to the original data source (portal). The ChemSpider database is open to the public at no charge.

The inherent value of ChemSpider lies within its ability to aggregate many diverse sources of information. Hundreds, if not thousands, of chemical databases exist containing information on chemical compounds. Examples include: literature data, chemical vendor catalogs, toxicity data, analytical data, regulatory postings, patent information, etc. Both the wide variety of available information and sheer volume of different data sources make it nearly impossible for an individual user to determine the availability of data for a single chemical compound. ChemSpider changes that fact and answers the commonly asked question: "is there specific information about my chemical?"[52]

As this thesis investigates an internal Novartis web system specifically for chemical compounds, it is extremely useful to explore the functionality of ChemSpider in detail. Very similar, and in some cases identical, functionality is later suggested for the Novartis Compound Homepage web tool.

Searching for Chemicals

The user can search ChemSpider by typing in a one of several chemical identifier names (Figure 15). Chemical compounds can have a variety of different names so it is important that the search interface recognizes and associates these different identifiers to the correct chemical structure.

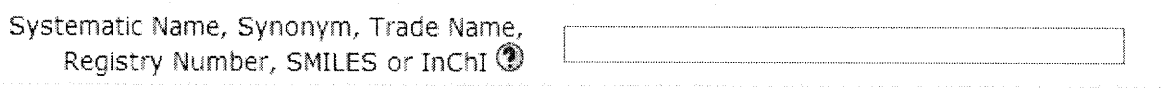


Figure 15: Search Interface for ChemSpider[53]

In addition to searching by names, scientists can also search by chemical structure. This is very powerful because it provides a way for chemist to think and search in structural terms rather than relying on textual based searches. In its discussion of ChemSpider, the Chemistry International

[52] About ChemSpider, *ChemSpider*, n.d. Web. Accessed 05 Mar 09 http://www.chemspider.com/About.aspx

[53] ChemSpider Search, *ChemSpider*, n.d. Web. Accessed 05 Mar 09 http://www.chemspider.com/Search.aspx

magazine points out the importance of this feature. "Chemists' natural affinity for communication via chemical structures demands the need to perform searches in a natural (i.e. structure-based) language"[54] ChemSpider users can use a chemical structure editor (Figure 16) to either draw a molecule from scratch or import and make changes to a molecule by specifying it via another identifier (IUPAC name, InChi, Smiles, etc). Searching the web by chemical structure was previously very difficult. ChemSpider changes that by opening the door to web-based structure searches.

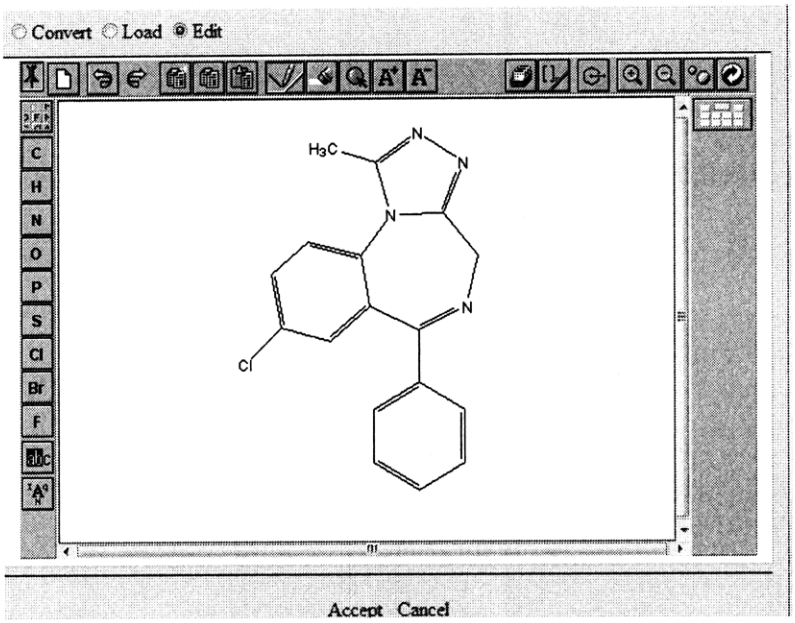


Figure 16: ChemSpider Chemical Editor Interface: Used to Search by Compound Structure[55]

Properties of the chemical such as molecular weight, mass, etc. are displayed after a user executes a search (Figure 17). Chemical identifiers such as SMILES string, InChi and InChIKeylink directly into Google so that the user can execute structural based searches with the Google search engine.

Additionally, a section on names and synonyms imports a list of all known names for a particular chemical compound. In the case of an approved drug, such as the example below of Xanax, the list of names can be quite extensive including many drug and trade names in different countries (Figure 18).

---

[54] Williams, Antony "ChemSpider and Its Expanding Web" *Chemistry International* 30.1 (2008). Print

[55] ChemSpider Structure Search. *ChemSpider*, n.d. Web. Accessed 05 Mar 09 http://www.chemspider.com/StructureSearch.aspx

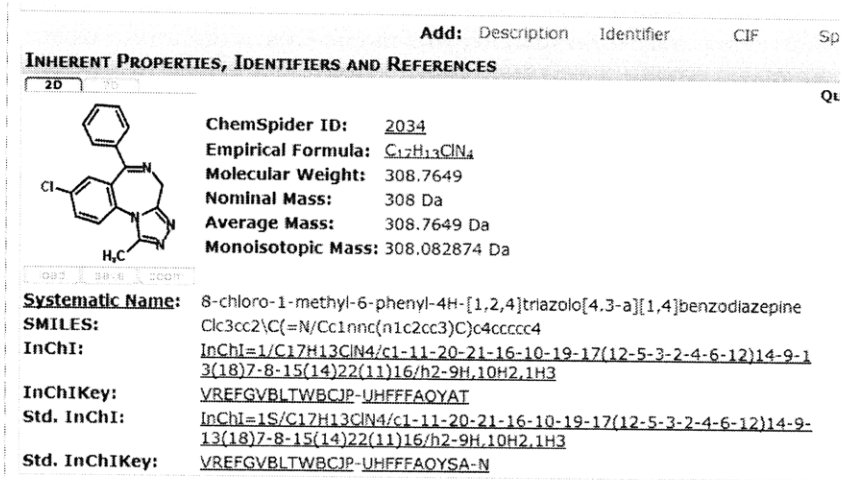


Figure 17: ChemSpider Chemical Properties, Identifiers and References[56]

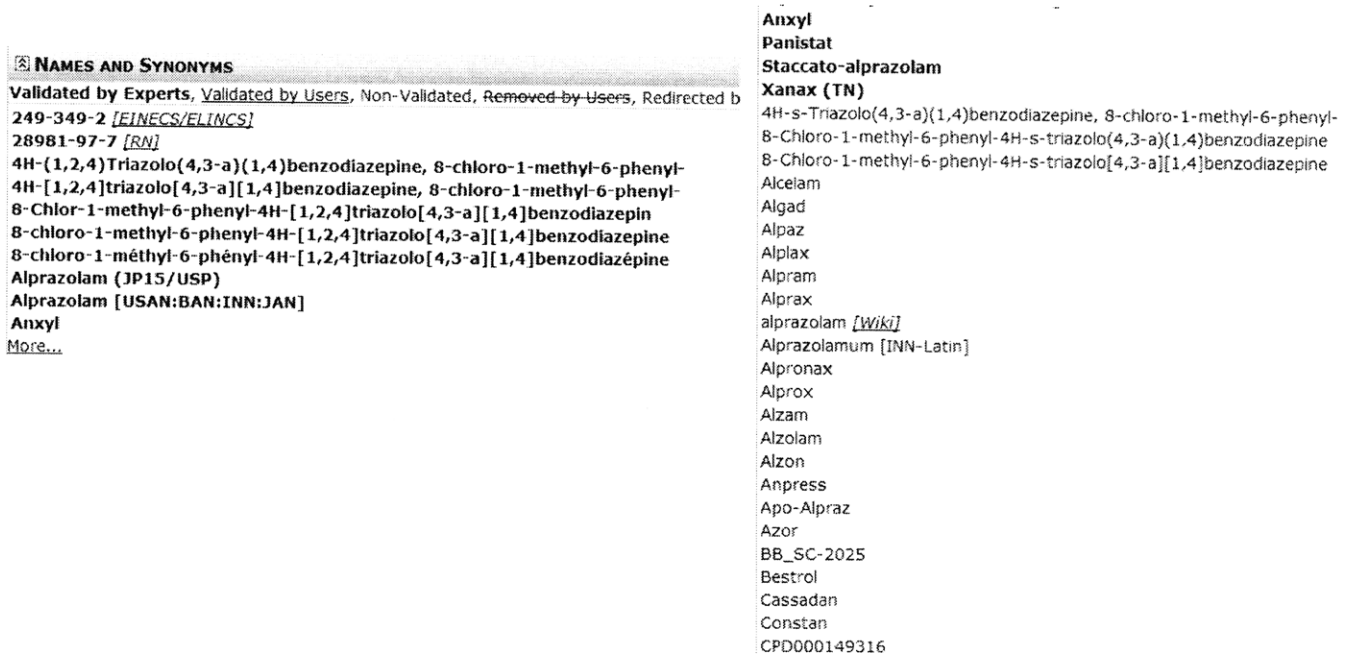
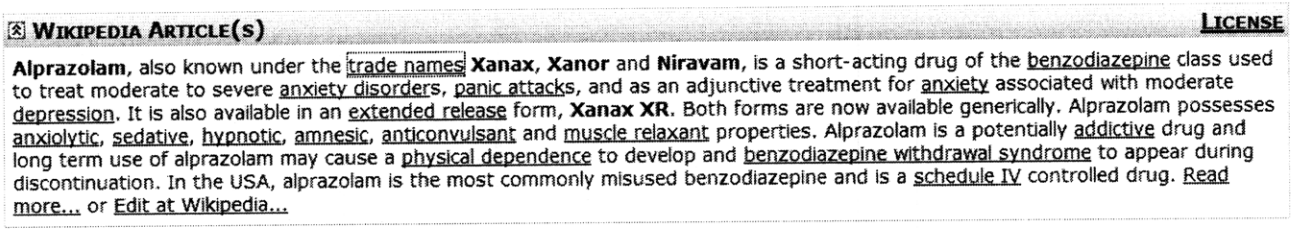


Figure 18: ChemSpider List of Names and Synonyms for FDA Approved Drug, Xanax[57]

[56] ChemSpider Search. *ChemSpider*, n.d. Web. Accessed 05 Mar 09 http://www.chemspider.com/Search.aspx

[57] Ibid. (56)

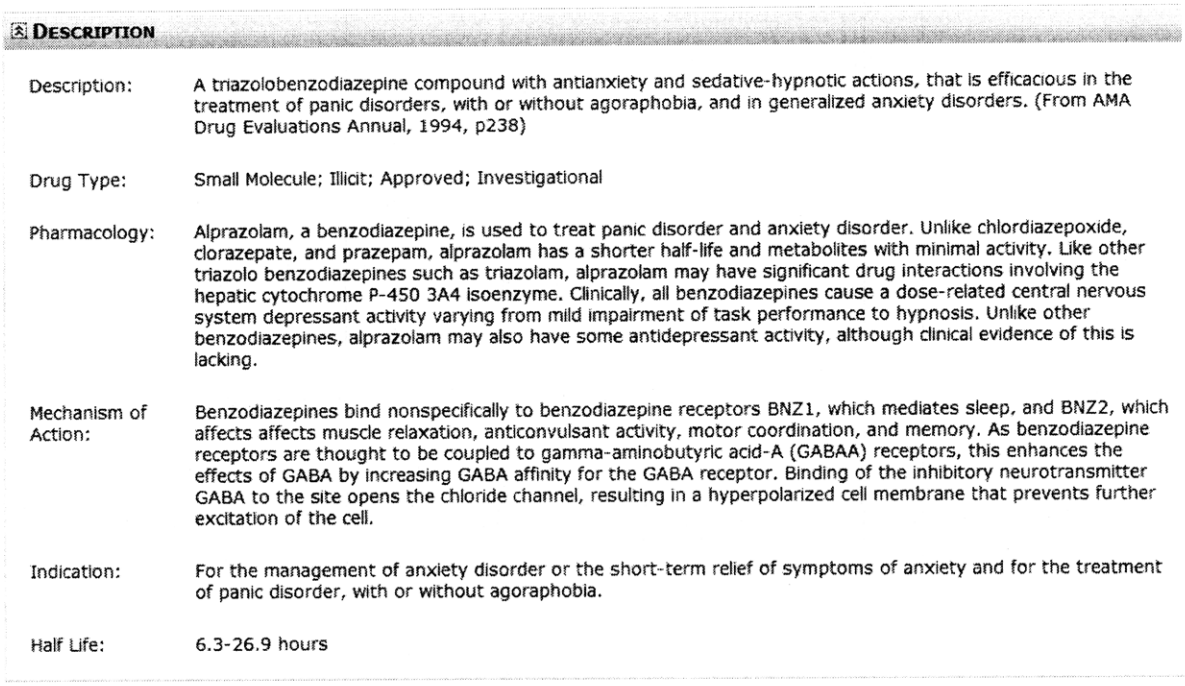
Mash-up applications in ChemSpider include the integration of Wikipedia articles related to the particular drug (Figure 19) as well as well as a description detailing the pharmacology, mechanism of action, indication, etc. (Figure 20).



WIKIPEDIA ARTICLE(S) LICENSE

Alprazolam, also known under the trade names **Xanax**, **Xanor** and **Niravam**, is a short-acting drug of the benzodiazepine class used to treat moderate to severe anxiety disorders, panic attacks, and as an adjunctive treatment for anxiety associated with moderate depression. It is also available in an extended release form, **Xanax XR**. Both forms are now available generically. Alprazolam possesses anxiolytic, sedative, hypnotic, amnesic, anticonvulsant and muscle relaxant properties. Alprazolam is a potentially addictive drug and long term use of alprazolam may cause a physical dependence to develop and benzodiazepine withdrawal syndrome to appear during discontinuation. In the USA, alprazolam is the most commonly misused benzodiazepine and is a schedule IV controlled drug. Read more... or Edit at Wikipedia...

Figure 19: Mash-up of Wikipedia Article on ChemSpider Website



DESCRIPTION

Description:	A triazolobenzodiazepine compound with antianxiety and sedative-hypnotic actions, that is efficacious in the treatment of panic disorders, with or without agoraphobia, and in generalized anxiety disorders. (From AMA Drug Evaluations Annual, 1994, p238)
Drug Type:	Small Molecule; Illicit; Approved; Investigational
Pharmacology:	Alprazolam, a benzodiazepine, is used to treat panic disorder and anxiety disorder. Unlike chlordiazepoxide, clorazepate, and prazepam, alprazolam has a shorter half-life and metabolites with minimal activity. Like other triazolo benzodiazepines such as triazolam, alprazolam may have significant drug interactions involving the hepatic cytochrome P-450 3A4 isoenzyme. Clinically, all benzodiazepines cause a dose-related central nervous system depressant activity varying from mild impairment of task performance to hypnosis. Unlike other benzodiazepines, alprazolam may also have some antidepressant activity, although clinical evidence of this is lacking.
Mechanism of Action:	Benzodiazepines bind nonspecifically to benzodiazepine receptors BNZ1, which mediates sleep, and BNZ2, which affects affects muscle relaxation, anticonvulsant activity, motor coordination, and memory. As benzodiazepine receptors are thought to be coupled to gamma-aminobutyric acid-A (GABAA) receptors, this enhances the effects of GABA by increasing GABA affinity for the GABA receptor. Binding of the inhibitory neurotransmitter GABA to the site opens the chloride channel, resulting in a hyperpolarized cell membrane that prevents further excitation of the cell.
Indication:	For the management of anxiety disorder or the short-term relief of symptoms of anxiety and for the treatment of panic disorder, with or without agoraphobia.
Half Life:	6.3-26.9 hours

Figure 20: ChemSpider Description of Pharmaceutical Xanax

In addition to providing users a window view directly into data sources such as Wikipedia (Figure 19), ChemSpider uses pointers, via hyperlinks, that allow the user to navigate directly back to the original data sources. In some cases this information is publically available free of charge (e.g. PubChem, NIST). In other cases the data is housed within a commercial database where a user needs a subscription to access information (e.g. Journal of Heterocyclic Chemistry). External data sources are identified by a mouse-over pop-up window that gives the user additional information about a particular database (Figure 21).

Associated Data Sources and Commercial Suppliers

Data Source	External ID(s)
ChemDB	3965382
ChemIDplus	028981977
ChemSpiderman	N/A, N/A, N/A, N/A, N/A, N/A, N/A, N/A, N/A, N/A
ChemZoo	100000724
DailyMed	Alprazolam
DiscoveryGate	2118
DrugBank	2118, APRD00280
EINECS	N/A
EPA DSSTox	38_FDAMDD_v3b
FDA	18276
Journal of Heterocyclic Chemistry	19800575_1, 19780161_6B
KEGG	C06817, D00225
LeadScope	LS-156344
MCISB	18675
MeSH	Alprazolam
MLSMR	MLS000559000
Molecule of the Day	Alprazolam
Nature Chemical Biology	nchembio747-comp35
NIH Clinical Collection	SAM001246696
NIST	979635492
NIST Chemistry WebBook	979635492
Prous Science Drugs of the Future	91399
PubChem	2118
PubMed	6140317, 17244765, 6141159, 10631626, 17058100, 10646701100

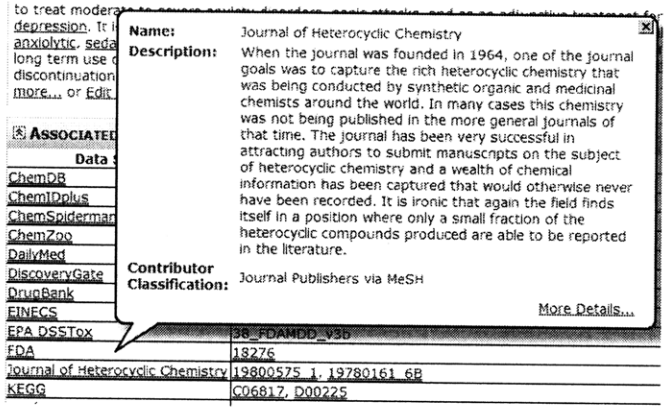


Figure 21: ChemSpider Associated Data Sources for Xanax

Another useful section within the ChemSpider website is the predicted properties table (Figure 22). *In Silico* properties of chemical compounds are calculated based on algorithms provided as part of collaboration with Advanced Chemistry Development (ACD) labs.[58] Chemists routinely use LogP, polar surface area, Lipinski's rule of 5, etc. in order to assess the drug like properties of a particular compound.

Predicted Properties

LogP:	ACD/LogP: 2.50 XLogP: 4.90	**# of Rule of 5 Violations:**	0
ACD/LogD (pH 5.5):	2.5	**ACD/LogD (pH 7.4):**	2.5
ACD/BCF (pH 5.5):	46.67	**ACD/BCF (pH 7.4):**	46.71
ACD/KOC (pH 5.5):	544.74	**ACD/KOC (pH 7.4):**	545.23
#H bond acceptors:	4	**#H bond donors:**	0
#Freely Rotating Bonds:	1	**Polar Surface Area:**	43.07 $Å^2$
Index of Refraction:	1.71	**Molar Refractivity:**	88.22 cm^3
Molar Volume:	225.5 cm^3	**Polarizability:**	34.97 $10^{-24}cm^3$
Surface Tension:	52.1 dyne/cm	**Density:**	1.36 g/cm^3
Flash Point:	261.6 °C	**Enthalpy of Vaporization:**	77.94 kJ/mol
Boiling Point:	509 °C at 760 mmHg	**Vapour Pressure:**	1.77E-10 mmHg at 25°C

Figure 22: ChemSpider Predicted Properties Table for Xanax

[58] ChemSpider User Guide. *ChemSpider*, n.d. Web. Accessed 06 Mar 09 http://www.chemspider.com/docs/chemspider_manual.pdf

Lastly, in addition to serving as a mash-up aggregator and linking portal, ChemSpider fosters a sense of community among its users. The site includes a user forum where members can view and edit posts related to chemistry discussions and a blog published by the sites founder. Furthermore, site users are encouraged to be both data depositors as well as curators. In order to help maintain the quality of the site, registration is required to deposit information and file changes/additions are flagged internally to be reviewed by the ChemSpider staff. [59]

ChemSpider has demonstrated its usefulness by attracting a large number of community users. According to a May 2008 article from Nature, over 5,000 users per day are currently accessing the site. [60] ChemSpider is an excellent example of how the new realities of the Web 2.0 are improving access to scientific information, creating unique user communities, and possibly changing the face of science by speeding up discovery.

ChemSpider case study highlights several important points:

- **The amount of information available for a specific chemical is tremendous-** This suggests that mash-ups are a useful tool in aggregating this different information. The amount of information available will only grow as scientific discovery pushes the frontier.
- **Chemists need to be able to search by structure as well as by name-**Structural based searching is very powerful, especially given the multitude of different compound naming conventions. Chemists think in structural terms and prefer to search by structure.
- **Pointers, via hyperlinks, that take users back to root databases are helpful-**Pointers allow scientists to not only explore content in other databases, but validate the trustworthiness of that data. They also vastly expand the scope of information available within the "single click of a mouse key."[61]

[59] "Watch out for explosive Mouse on ChemSpider", *ChemSpider Blog,* 04 Mar 09. Web Accessed 04 Feb 09 http://www.chemspider.com/blog/

[60] Brumfiel Geoff, "Chemists spin a web of data" *Nature* 453 (2008). Print.

[61] Dejan Bojanic, The Novartis Institute for BioMedical Research, Center for Proteomic Chemistry

Chapter 6: NIBR User Research Methodology

6.1 Overview of Research Methodology

Extensive user research was conducted as a first step in evaluating the feasibility of "Compound Homepage" a NIBR wide, Web 2.0based system for keeping track of chemical compounds. The user research not only evaluated the current Novartis data environment, but also explored NIBR specific cultural beliefs and attitudes about information sharing. Additionally, users were asked about their awareness of, and attitudes toward Web 2.0 technologies including wiki usage, social networking, etc. Forty five individuals within various groups of the NIBR organization were interviewed (Table 4).Although interviews were conducted in many different groups, a majority of stakeholders interviewed were part of the Center for Proteomic Chemistry (CPC) and Global Discovery Chemistry (GDC) organizations. These two organizations are very chemistry centric and would thus be the main users of the Compound Homepage. Interviews were conducted in a 1:1 setting and typically lasted anywhere from one to two hours each.

Table 4: Departmental Breakdown of NIBR Individuals Interviewed

Platform Groups	# People Interviewed
Analytical Sciences	**1**
Biologics Center	0
Center for Proteomic Chemistry	**13**
Developmental & Molecular Pathways	**4**
Epigenetics	0
Global Discovery Chemistry	**14**
Metabolism and PK (MAP)	**1**

Disease Areas	# People Interviewed
Autoimmunity and Transplantation	0
Cardiovascular & Metabolism	0
Gastrointestinal	0
Infectious Disease	0
Musculoskeletal	0
Neuroscience	0
Oncology	**1**
Ophthalmology	**1**
Respiratory Disease	0

Service Organizations	# People Interviewed
Communications Office	0
Education Office	**1**
Finance	0
Global Security	0
Human Resources	0
IT and Automation (NITAS)	**4**
Legal / NIBR Patents	0
Novartis Knowledge Center	0
Program Office	**3**
Research Facility Operations	0
Strategic Alliances	**2**

6.2 NIBR Organizational Structure

Novartis Institutes for BioMedical Research (NIBR) is organized in a matrix function by platform groups, disease areas, and service organizations. Disease areas focus only on one particular disease and have a high concentration of highly trained biology specialists. Conversely, platform groups work across all disease areas. For example, members of the Center for Proteomic Chemistry support high throughput screening activities for a variety of drug candidates in many different disease areas. Similarly, chemists in the Global Discovery Chemistry platform group might work on an oncology project at one time, but then later support a cardiovascular drug. Platform groups were the focus of the user interviews as they tended to have a higher percentage of chemists.

6.3 Interview Question Themes

A wide variety of question themes were explored with interview participants. Samplings of interview questions, grouped by themes are presented below. Please refer to Appendix A and Appendix B, for the complete list of all interview questions.

Theme 1: Understanding job function

In order to understand information usage of a particular individual it was important to gain a context surrounding the job function, organizational responsibilities, etc.

Q: What organization are you in? How long have you been at Novartis?

Q: What is your job function? What does your group do?

Q: What is your educational background? (biologist, chemist, biochemist, etc.)

Theme 2: Internal Data Generation and Usage

The second category of questions helped to explore both data generation and data usage. What type of compound specific data were scientists generating? What format was this information stored in and where was this stored?

Q: What information about particular compound(s) do you **generate** on a regular basis? Where is this information that you generate currently stored? (ex: reports, databases, etc)

Q: What internal information about particular compound(s) do you **use/consume** on a regular basis? Where is that information stored? How often do you access each of these databases/info sources? Which ones are most important? (please rank priority)

Q: What are the key pieces of info you need to make a decision(s) in your role/job?

Q: Do you deal with compounds on a 1:1 basis or more as a set, using/viewing multiple compounds at one time?

Theme 3: External Data Sources

Scientists were asked about their external (i.e. non-Novartis generated) data usage. The goal was to identify websites and data services individuals were using so that frequently used applications could be linked into Compound Homepage.

Q: What external data sources do you use? (ex: PubMed). What information do you typically extract from these sources? How often do you access each of these databases/info sources (daily, weekly, etc)? Which ones are most important?

Q: Do you think you would reference external databases more frequently if these databases were available in one easy to find/click portal?

Q: How often do you go outside of Novartis to access information in an external database? (daily, weekly, once a month)

Theme 4: Benchmarking the NIBR Information Environment

In order to evaluate Novartis's current environment for data management, interviewees were asked how easy it was to find information they needed, how much time they spent trying to track down data, etc. Password usage and resolution of access requests were also explored.

Q: How easy it is to find access to the internal information you are looking for?

Q: What information is easy to find? What information is hard to find? What makes it hard to find/access?

Q: How much of your time do you spend looking for information? (avg. time per week) Any specific examples?

Q: If you don't have access to a certain internal database how long does it take for you to get access? Any examples?

Q: If you've had a job elsewhere in Pharma/biotech how did the data environment compare to what we have at NIBR?

Theme 5: Web 2.0 Awareness

Interviewee familiarity with social networking sites and other external wiki's, such as Wikipedia, was assessed. Additionally, questions probed beyond simple usage and investigated if individuals had experience in actually editing or creating wiki's.

Q: Do you have any experience with using Wiki's (such as Wikipedia)? How much?

Q: Have you ever edited or helped to create a Wiki? If no, why not?

Q: Have you ever used Facebook? (yes/no), Linked-in? (yes/no)

Theme 6: Wiki Usage

Wiki usage questions explored the likelihood that the interviewee would actually contribute to an internal Novartis compound wiki. Interviewees were also asked if they used any of the other Novartis wikis (HTS pipeline wiki, safety profiling wiki, Target wiki).

Q: What wikis (if any) are you aware of at Novartis? Which of these do you currently use on a regular basis?

Q: If the interface allowed for user input would you be interested in adding notes/text to keep the Wiki up to date?

Q: Would you add content? How often (weekly, daily, monthly) Why or why not?

Q: How could we encourage people to add content to the wiki and avoid the "blank page syndrome"?

Theme 7: Novartis Culture

Web 2.0 experts have previously noted that organizational culture is important for Web 2.0 adoption. [62] In addition to asking users directly about the culture, Novartis interviewees were probed more informally on how they felt about sharing information, the likelihood of scientists to want to post scientific information on an internal website, etc.

Q: What is the Novartis culture like surrounding documentation of work? Do people check that things get documented appropriately? Do they care? Does this come up in year-end reviews?

Q: How do people feel about sharing information? Would you be comfortable posting information to a NIBR wide website? Any concerns?

Q: Should wiki users be allowed to add content and links? To what extent? Who should input info? Should this be open to anyone in Novartis? Project team members only?

Theme 8: Exploration of Potential Compound Homepage Functionality

Lastly, specific functionality of Compound Homepage was explored. Users were asked about the usefulness of linking; tagging, and wiki functionality

Q: How useful is the linking functionality? (The ability to go to the compound homepage and have links to sources of information in Novartis on that particular compound).

Q: Do you think the "friends" applications within Facebook could have relevance for compound Wiki? If so, suggest some useful "friends/network" groups of compounds.

Q: Would the friend/association groups actually be useful or just interesting to have?

Q: Would a compound wiki section, where you could post comments and write about specific compounds be useful? Overall, would you use this tool (yes/no/maybe)?

Q: Is there any specific information or functionality that you'd like to see included in Wiki?

[62] McAfee, Andrew. Personal Interview by Juliet Duffy. Aug 08.

Chapter 7: NIBR User Research Results

Interview results were analyzed by grouping answers to specific thematic questions as well as reoccurring comments together. In some cases, participants were directly asked to numerically rate or comment on specific items. Where applicable, these answers are presented in graphical distributions.

Methodology Disclaimer:
Interviewees were not selected from a random sample; rather they were identified on the basis of recommendations from other interviewees. For example, at the completion of each interview, individuals were asked "Who else would be good to talk with? or "Do you know any individuals in department xxx that I could interview?"

Not all interview participants chose to answer each question, and some individuals provided information about topics that were not necessarily asked about in the standardized interview questions. In these cases, statistics specifying the percentage of individuals that felt or believed a certain way are not available. Where possible, the author has attempted to designate if a majority of individuals felt a certain way or if the expressed opinion was just of one particular individual.

All quotations presented are anonymous in order to protect the identity of individual researchers/interviewees. Twenty two different individuals contributed quotations for this thesis. All names referenced throughout the thesis, including in other chapters, have been changed and do not represent the names of any employees within the Novartis organization.

7.1 Theme #1: Understanding job function

As expected, researchers at Novartis routinely engage in highly tacit knowledge based activities: performing experimentation, analyzing data, evaluating multiple drug candidates at various stages in the discovery pipeline, etc. Work is very information intensive requiring access to and interpretation of a wide range of data. Novartis employees are highly trained within their scientific specialties. The overwhelming majority of candidates that were interviewed hold PhD's in either a subspecialty of chemistry or biology.

7.2 Theme #2: Internal Data Generation and Usage

Both the amount of data, as well as the diversity of the types of data surrounding chemical compounds within NIBR is tremendous. Many types of data, as well as functional activities of each organization can be grouped according to drug discovery phase. It is useful to highlight the various stages of drug discovery at NIBR.

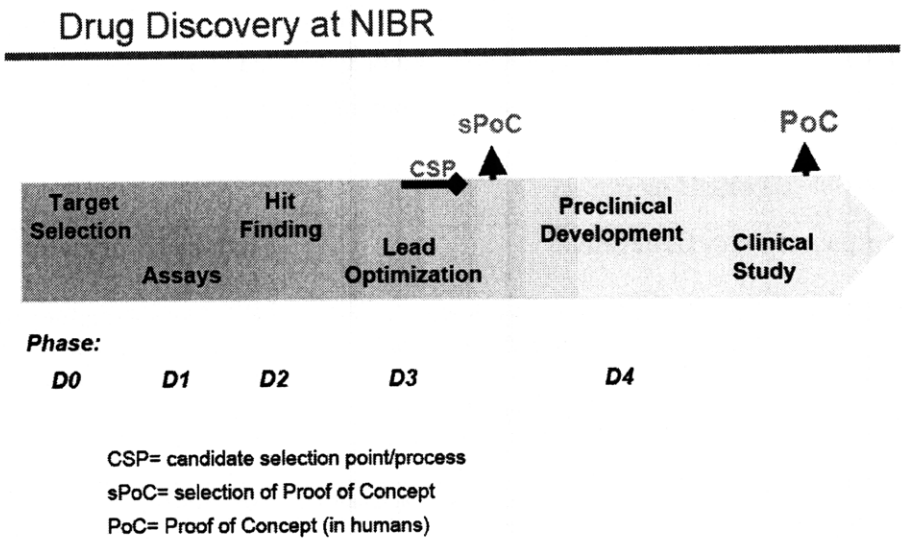


Figure 23: Drug Discovery Phases at NIBR[63]

Drug Discovery Phases

- **D_0 / Early Stage Discovery** - This stage of drug discovery focuses on understanding the biology of the disease. Specific biological "targets" within the disease pathway are identified. Activating or inhibiting these particular targets change the phenotypic or biological response. For example, inhibiting a specific target in a cancer pathway may lead to tumor shrinkage/tumor death.

- **D_1 / Assay Characterization** – Specific tests or "assays" are developed in order to effectively measure the biological response to a specific drug. These assays will be subsequently used to test different chemical compounds for their ability to either activate or inhibit a specific biological target.

- **D_2/ Hit Finding** – Millions of chemical compounds are screened within a specific assay in order to identify chemicals that interact with the specific biological target. This activity is referred to as "high throughput screening."

- **D_3/Lead Optimization** - Specific promising lead compounds identified from the high throughput screening are further refined. The compounds are optimized to create suitable drug like properties. The safety profile, metabolites, pharmacokinetic properties are studied and optimized for a particular chemical series.

[63] "Putting Patients First – Meeting the Drug Discovery Challenge" *Novartis Institute for Biomedical Research, New Employee Orientation*. Cambridge, MA. June 2008

Data usage and generation surrounding chemical compounds can be broken down into various drug discovery phases. Although certain pieces of information are used or generated at all stages, the majority of information tends to be pretty aligned with the development process (Figure 5).

Table 5: Examples of Data Related to Chemical Compounds

Discovery Stage	Examples of Data Generated*	Examples of Data Usage*
Early Stage Target Discovery & Assay Development (D_0, D_1)	• Most data relates to targets and not to Chemical Compounds (i.e. proteins, disease pathways, genetic expressions, etc) • Affinity Proteomics (3-D grid matching compound identification with cell type & protein pulled out)	• Compound history both internal & external • Compound origination & what library? • Patent data
Hit Finding/High throughput Screening (D_2)	• Compound Identification Information. Analytical data such as liquid chromatography & high res mass spec • Synthesis route • Compound registration information • Primary –biochemical assays • Secondary assays – counter screens to rule of false hits • Safety profiling results – looking for off-target effects	• Primary, secondary, safety profiling assay results • Compound purity & Identity information • Patent data
Lead Optimization (D_3)	• Metabolites – what kind of metabolites does this compound produce? • Pharmacology - In vivo (testing MTD - maximum tolerated doses (you can't tell any adverse effects), ATD - acute toxic dose. • Phenotypic studies in disease area. Example: mice models of the disease, did the tumor change size, etc.? • Solubility for formulation	• Bioavailability • Identity of metabolites • Results of toxicity studies in animal models • Commercial information • Formulation testing results

***Note: Examples are for illustrative purposes only and not inclusive of all activities**

Information Storage Formats

- **Specialized Formats** - Data throughout Novartis is stored in multiple formats. Many of these formats are very specific to the scientific application and require specialized software to both view and store information. For example, analytical compound identification tests such as NMR, Mass Spectroscopy, etc. are captured in a binary data format. Viewing and manipulating these data requires the appropriate software package such as Advanced Chemistry Development (ACD) or Top Spin. In some cases the data can be converted to an image or an Adobe Acrobat file. In other instances the data must remain in a specialized format (e.g. the specialized Spot Fire files used by the Global Discovery Chemistry Chemogenetics and Proteomics group).

- **Avalon Data Warehouse** – Assay results for a particular compound test are housed in a central database and accessed by users through a data warehouse interface called Avalon. Individuals throughout the NIBR research organization rely very heavily on Avalon. Over 90% of Novartis employees interviewed motioned using Avalon. Information retrieval from Avalon can be very difficult, especially if a user doesn't have specific metadata information such as the type of assay, exact assay name, individual that ran the assay, run date, etc. or if the user is trying to retrieve data for a very large number, 100+, compounds/tests.

- **Importance of PowerPoint**– Many data files and summary reports are stored in a PowerPoint format. NIBR teams rely so heavily on sharing information through PowerPoint that one particular individual dubbed PowerPoint as *"the official currency of Novartis."* Almost every individual interviewed mentioned using or storage some type of data/information within a PowerPoint file.

- **Science is visual** –Science is a very visual language. Chemists communicate primarily through drawings of chemical structures, assay results are often best viewed in a graphical format (IC50 curves), and biology is understood and communicated in pictures (DNA/protein bands on an electrophoresis gel, pictures of tumor shrinkage, etc). One reason that PowerPoint is heavily utilized is the ability to import and annotate many different types of visual images.

Team Centric Data Organization

The data environment at Novartis Research is highly organized around project teams. Information is shared among team members primarily via a team data storage site. Three team sites are

common: an internal shared storage drive (G:drive, L:Drive, etc), a Lotus Notes Team Space (similar to a shared drive) or an external collaboration platform (Groove) used when teams communicate with external collaborators. Specific file storage and organization guidelines do not exist on a global level. Instead, organization structure is highly dependent on each particular team. Some teams have very organized structured files in one particular location. Other teams store data randomly across multiple different platforms including email.

"How a project team captures information is very individual for the team. Things are stored multiple times, in many different places. For example, on my most recent project we primarily use our G: drive folder and Groove (shared collaboration software platform). However, some presentations, like those for decision board, will go to the decision board team space. It's pretty much up to the project team leader to decide how/where information is stored".

Restricted Access

Team file storage sites are heavily password protected and restricted solely to team members. As a result, the majority of data is not visible outside of the immediate team. Additionally, team sites and locations are often buried so it's very difficult for those outside of the team to even locate the shared drive, never-the-less obtain access to it. Multiple interviewees expressed frustration with their inability to access team data.

"Data from other projects that's not already in Avalon is impossible to find without help. To get access to other team's folders or G:drives you'd have to specifically ask the project team leader for permission"

Viewing Individual Compounds vs. Sets

Many times, scientists are not interested in one particular chemical compound, but rather a set or series of different compounds. Usually a particular drug scaffold is identified and individual chemical groups are modified on that core chemical scaffold in order to improve the properties of a compound. For example, replacing a highly hydrophilic (water loving) group with a functional hydrophobic (water hating) group will increase a compounds ability to pass through the lipid bi-layer membrane of a cell and access intracellular targets.

Many teams classify and define lead chemical compounds by grouping them into particular collections based on structure similarity of core scaffold properties. An example of this is shown below (Figure 24).This particular project team classified their lead compounds in six different chemical series. Each series contained anywhere from 1-100+ compounds and were tracked via a

team excel spreadsheet. Specific examples of individual chemical compounds are noted by black dots associated with chemical compound code names (i.e. RAL122).

Note: all chemical code names, individual series, and milestones presented in the figure below are not accurate and have been altered significantly in order to preserve data confidentiality.

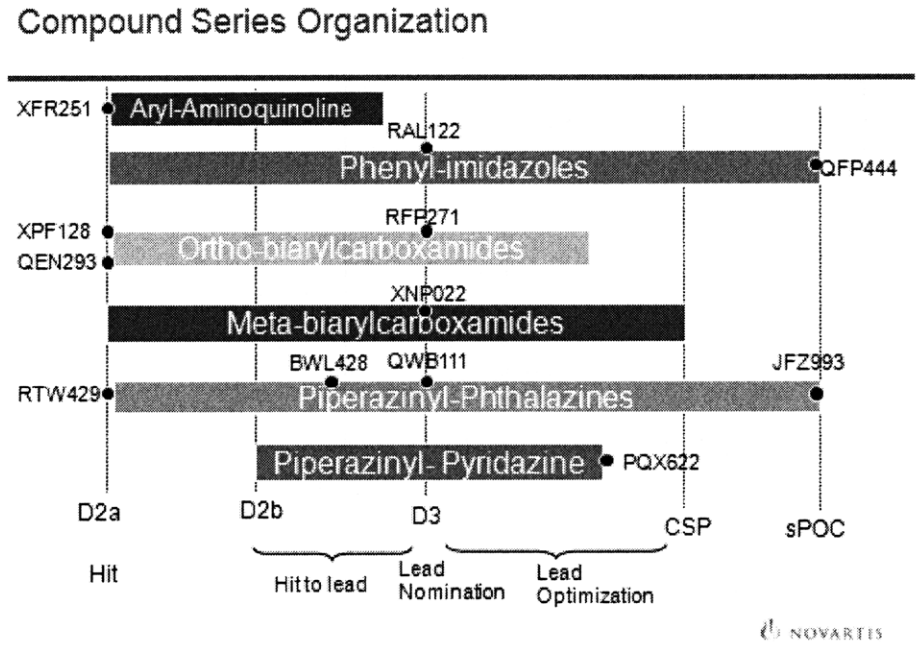


Figure 24: Example of Project Organization by Chemical Series*
***Note: all chemical code names, individual series and milestones have been altered in order to disguise data**

Individuals participating in the user interviews were specifically asked "Do you deal with compounds on a 1:1 basis or more as a set?" The majority, 56% percent of individuals surveyed indicated that they dealt with compounds on both an individual and a grouping basis.21% percent of survey respondents only examined compounds as a set, while just 12% of respondents reported only looking at compounds on a one-on-one basis (Figure 25).

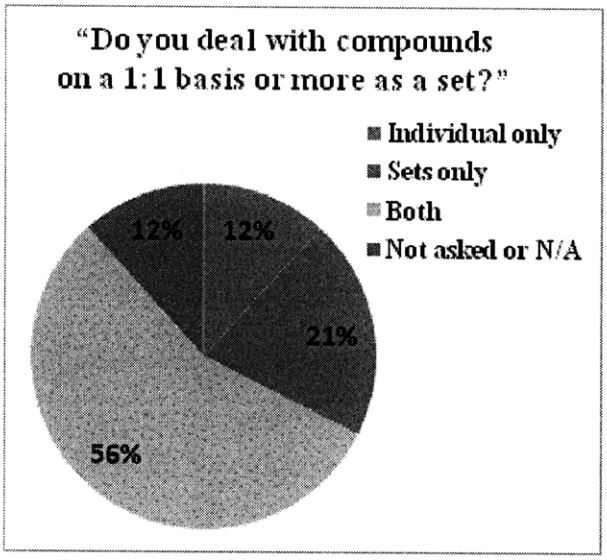


Figure 25: User Interview Results, "How do you deal with Compounds?"(n=34)

Q: Do you deal with compounds on a 1:1 basis or more as a set?

> *"It's important to examine compound series as a set to see if there are trends across the whole compound series"*
>
> *"We look at more than just one compound. More often we are interested in 1,000 structural analogs of this".*
>
> *"On an individual basis- I'm an organic chemist, I only synthesize one compound at a time."*
>
> *"Both extremes. Sometimes I want to know exactly about one drug. Sometimes you want to know about the entire class of compounds".*

Often the need to examine either individual compounds or sets of compounds was driven by the stage of the project. High throughput screening and the examination of structure activity relationships (SAR), changing specific compound functional groups and monitoring how biology activity changed with structural modifications, tended to warrant examination of large sets of compounds. Conversely, literature examination, refining the compound to eliminate a safety liability, or testing a specific compound in animal models demanded examination of an individual compound.

> *"I examine both sets of compounds and individual compounds. In the beginning of the project I tend to only look at one compound (who made compound, why was it made, etc?)*

as project progresses we know the structure activity relationship (SAR) we'll tend to look at more sets"

"Both. The outcome of a screen is a huge set. The challenge is to get rid of noise and find the diamonds. To do this I have to prioritize clusters (as a set). Other times I'll only want to look at a specific compound. This might be a compound inspired by literature, or I'll want to know very specific compound properties (permeability, solubility, etc)"

Size of Sets

In order to understand how users will want to view sets of compounds, and the implications for viewing compound groupings within the Compound Homepage system, it is important to understand the relative size of the chemical sets. Individuals that specified that they examine data in sets were asked specifically what size sets they typically examined. Apart from high throughput screening activities, which deal with hundreds of thousands of compounds at one time, answers tended to fall primarily within three size categories. The distribution of answers was roughly evenly split between the three sized groups indicating approximately equal usage of each sized group.

- Small sets (10-30 compounds)
- Medium Sized Sets (100-200 compounds)
- Larger sets (1,000-3,000 compounds)

7.3 Theme #3: External Data Sources

In addition to examining internal NIBR data sources, external data databases that contained information about chemical compounds were identified. The list below was obtained from the NIBR Information Technologies group (NITAS) and gives an idea of the large number of external data sources available with information on chemical compounds.

Pharmapendium	Medline PLUS
Patents	Various FDS/EMEA sources
CERES	PubChem
Feed	DrugBank
Prous Integrity	PDR
Pipeline Databases	KEGG
GVK	ChemDI
GeneGo	Vendor Databases
WDI	Wikipedia
ChemSpider	Human Metabolic Database
CAS sources	IUPAC Gold Book

Figure 26: NITAS list of Available External Databases

Scientists were asked about their external (i.e. non-Novartis generated) data usage related to chemical compounds. Every time an individual reported using a certain external database, that database was added to a list. The compiled list, ranked according to frequency of reported usage, is presented below (Table 6). Interestingly, Wikipedia was cited three times and Google was cited twice as data sources that were used on a regular basis.

> *"Wikipedia is a very, very useful tool. I use this daily. It's also amazing what one can find in Google without having to do a literature search."*

Additionally, the usage of a database portal was explored. Scientists were asked "Do you think you would reference external databases more frequently if these databases were available in one easy to find/click web portal?"

Table 6: Usage of External Data Sources (reported during user interviews)*

# of people reported using database	Database Name	Description of Database (if available)
8	Prous Integrity	Database of compounds that have demonstrated biological activity. Contains proposed mechanisms of action.
7	GVK	Contains 2M compounds with their known biological activities taken from patents & literature searches.
7	PubMed	"PubMed includes millions of citations from MEDLINE and other life science journals for biomedical articles. PubMed includes links to full text articles and other related resources"[64].
7	SciFinder	References from currently published journals and patents from more than 59 patent authorities[65].
4	Pharmapendium	Information on drugs & their established side effects
4	GeneGo	Bioinformatics pathways database but they recently starting linking in compounds.
3	World Drug Index (WDI)	What it a clinical candidate? Was it a drug? Was it ever withdrawn? etc.
3	Wikipedia	Online, collaborative encyclopedia written by volunteer authors around the world.
3	Patent Structure Searching	Markush structure as well as individual structures linked to Intellectual property information.
2	ChemBank	"ChemBank is a public, web-based informatics environment created by the Broad Institute's Chemical Biology Program. Includes freely available data derived from small molecules and small-molecule screens".[66]
2	NCBI databases (OMIM)	National Center for Biotechnology Information, databases of gene sequences.
2	Google	World Wide Web Search Engine. www.google.com

*Note: Databases with just one person reporting usage are not show in this table.

[64] PubMed, *NCBI*, n.d. Web. Accessed 2 May 09 http://www.ncbi.nlm.nih.gov/pubmed/

[65] SciFinder Website, *American Chemical Society,* n.d. Web. Accessed 2 May 09 http://www.cas.org/SCIFINDER/SCHOLAR/

[66] "What is ChemBank" *ChemBank,* n.d. Web. Accessed 2 May 09 http://chembank.broad.harvard.edu/

External Data Usage and Awareness

There was a fairly wide discrepancy of individual databases used by scientists. Prous Integrity, GVK, PubMed and SciFinder databases were mentioned routinely (Table 6) by scientists but over fifteen other databases were referenced by just one individual. Although certain databases are specific to only narrow job functions, in general many of the chemical databases could potentially be useful for a wider audience. It is hypothesized that many employees simply don't know about the existence of these databases.

Individuals within Strategic Alliances, the business development in-licensing group, talked about licensing the use of ChemNavigator for Novartis. ChemNavigator is a huge clearing house with over 12 million different compounds. A user can search for a specific compound and then order it through a catalog or "explode the hit by catalog". Ordering such chemicals could save chemists a significant amount of time in the synthesis of new chemical structures. Often, synthesis steps are complex and require a significant amount of research and time. The ability to order something close to the end product, further along in the synthesis pathway, would be a significant time and money saver. Despite the potential benefits of this tool, only one chemist that I spoke with had even heard of ChemNavigator. In an environment of information overload it is difficult to locate and keep track of potential resources.

Passing Through Compound Identification Information

Simply directing users to external databases via hyperlinks in a web portal would help to raise awareness but is not necessarily any more effective then ensuring scientists have bookmarked external web access points of interest. Indeed, when scientists were asked about their usage of external data sources over one-third of scientists resorted to opening up their web browsers to view their bookmarked favorite sites. One of those scientists commented:

"Why would I use this portal if I already have my favorite databases bookmarked? It's not going to help my life."

The real power of an external database portal would be enabling timely and effective search by passing the appropriate compound name automatically to the database of interest. This would save users a significant amount of time. Currently, every time the user enters a new database he/she is be required to re-enter the chemical name in order to search that particular database.

By instituting an information pass through, a user could navigate to a particular internal Compound Homepage and then execute the external search for a number of different databases directly from that page. One individual commented:

"A portal is useful only if I can pass through a compound identification and not have to re-enter into the search each time."

Intellectual Property Protection

Searching external databases for information on chemical compounds, with or without a chemical name pass through potentially puts the organization at risk for compromising data security. The fear is that another pharmaceutical company or competitive intelligence gathering entity might monitor searches of external databases originating from Novartis. By gathering the search criteria a competitor could deduce what specific chemical compounds are of interest, prior to the patent filings of those compounds. Novartis already has specific guidelines in place to limit the use of this practice. At least four individuals brought this up as a concern:

"Functionally an external search hub would be beautiful. Practically there are a lot of concerns about passing proprietary compounds over the firewall into external databases without having any level of security."

"Should we pass information to an external databases? If you pass say an inchikey identification string how strictly do you read Novartis guidelines, etc.? How much do you trust the websites you are passing to this? Basically, to be safe you'd have to wait until Novartis had filed a patent on that particular chemical compound. This would probably not occur until the late D3, lead optimization stage"

Eliminating the intellectual property risk would be possible and has already been done by several different individual departments. Novartis obtains licenses to a number of commercial external databases. These agreements can be structured to allow Novartis to import a specific database. Once an internal copy of the database is available within the internal firewall searching is safe and secure.

Searching for structural similarities

Eight interviewees mentioned the need to execute external searches not only by individual chemical compounds but also by a specific series or related group of compounds. Often a particular chemical compound will not have any public literature or any information associated with it, but a closely related structural analog will have published literature. Finding these close chemical analogs can be extremely powerful in both avoiding an already patented chemical space, as well as predicting biological functionality and properties by similar associations.

"I like to be able to execute a structural similarity search with all literature in the public domain for a particular chemical scaffold. For example, if I have a certain scaffold I'd like to know that information was published on a similar structure, say without a certain nitrogen group."

7.4 Theme#4: Benchmarking NIBR's Information Environment

In order to evaluate Novartis's current data management environment interviewees were asked how easy it was to find information they needed, how much time they spent trying to track down data, etc. Password usage and timely resolution of database access requests were also explored.

It is difficult to find certain types of information

In order to benchmark the current data environment, interviewees were asked "On a scale of 1-10 (1=extremely difficult), how hard is it to find information you are looking for?" The majority of interviewees reported that it was difficult to find information within NIBR's corporate environment. The distribution of answers ranged from 1 to 8, with an average score of 3.8 (Figure 27).

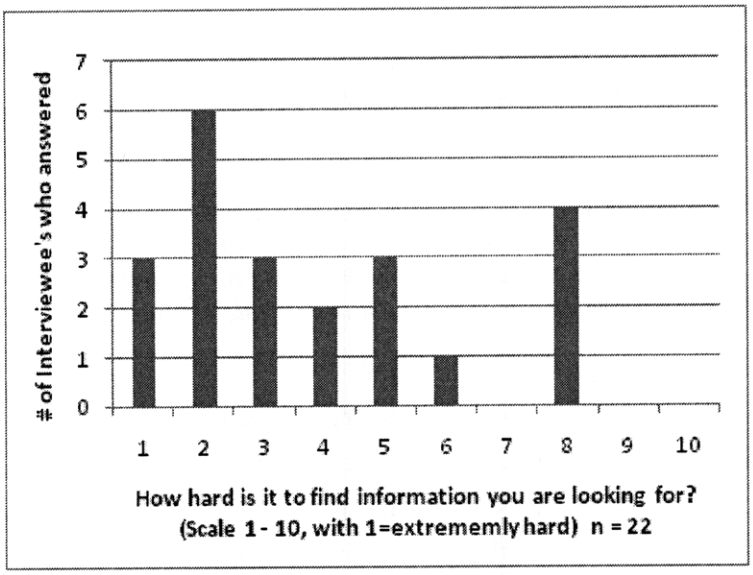


Figure 27: Distribution of Interviewee Responses

A large portion of interviewees indicated that their ability to find information was highly dependent on what type of information that they were searching for. Of the thirty two interview participants who we asked to rate their ability to find information, roughly 1/3 refused to answer the question on a numerical scale and simply stated *"It depends" or "It's incredibly mixed."*

"The ability to find information can really range from 1-10 depending what you are looking for"

Participants were therefore specifically asked to define examples of information that was hard to find as well as easy to find. A number of different themes emerged from this questioning and are presented below.

Current Data Environment is Not Intuitive

Interviewees consistently stated that one of the problems with the current data environment was that it was not intuitive to navigate. The author's experience at Novartis echoed this. Many different databases and information systems are present but it's difficult to navigate between these or even develop the detailed knowledge required to use a particular system. This is especially difficult if the data system was not used on a regular basis.

The author frequently had very detailed discussions with interviewees about Avalon, a particular chemistry data warehouse system. The author was surprised at the extremely wide-range of user expertise for this particular system. Some users were Avalon experts and could extract and manipulate a large amount of data from the system. Other, even routine users had only very basic knowledge of the system. Over ten interviewees made comments stating "Avalon is incapable of xxx". In reality, the system was capable of that functionality; the user just didn't have knowledge of how to perform that particular function. The wide-discrepancy of user expertise is most likely attributed to the fact that Avalon is not intuitive. Many of the features are not obvious and require detailed training to understand how to use the full range of functionality.

"Many people don't know how to use Avalon, I'd say 80% of my colleagues don't know how to use Avalon."

As the information environment is not intuitive and very fragmented, users are forced to climb up a very significant learning curve. As a result, productivity is substantially lower for new Novartis employees or those individuals in new job roles and positions. The problem is so significant that one particular Novartis department developed an informally run training session for all new employees. A seasoned departmental expert, someone who had been at the company for 20+years, routinely sits down with all new employees and reviews "how to go about finding things at Novartis". This particular individual was requested to teach this class so often, that he had even developed rough syllabus and supplemental information.

Improving the usability of data warehouse systems or providing information in a very easy to search web based environment could significantly lower the learning curve and improve productivity. Two individuals expressed their frustrations:

"How hard is it to find information? Using the official tools, with official training I would give this a "0": Becoming creative and using my network, I'd give it a "6"

"When I first started (at Novartis) it took quite a bit of time to find things. Now I know where to look."

Information that is easy to find

Information that was easy for individuals to find fell in approximately four different categories:

- **Data within project teams** - Employees expressed general satisfaction with being able to find data within their specific project teams. File storage organizational structure varied from team to team but in general project teams stored data within a Lotus Notes team folder or a particular shared hard drive space (example G:drive). Members of the team had access to this information and seemed to be able to navigate relatively easily throughout the shared drive/folder space. Teams were slowly adopting SharePoint, with some team files stored within the new SharePoint team spaces.
- **Current data**- Several individuals commented that they can easily find/access information that was created within the past year.
- **Information that is stored in current Novartis Wikis/Web Portals** – Several organizations within Novartis have recently created Wiki/Web Portals targeted to capture specific information generated by their department. The "Profiling Wiki" is a linking portal that integrates *in vitro* safety screening information for specific compounds into a one-page summary snapshot of safety information. This profiling wiki was mentioned repeatedly in user interviews as data that was easy to find. An example of a profiling report is presented in Appendix C (note: the compound is a publically available, non-proprietary Novartis compound).
- **Physical/Chemical Properties of Compounds** – Specific chemical properties such as solubility, purity, dose response curves (IC_{50} results), etc. are stored in a central NIBR wide-database and accessed through a data warehouse called Avalon. The general census was that this type of information was easy to find and retrieve if the user had access to specific metadata (i.e. the exact name of the assay, project code, etc.)

  "In Avalon - if you have a very specific query and knowledge how to get data it's easy"

Information that is difficult to find

Interviewees were specifically asked to indentify information that was difficult for them to find. Answers tended to align along six different themes. Firstly, information outside of an employee's immediate sphere of influence (e.g. project team or department) was often difficult to locate. In particular, *in vivo* data was specifically difficult to locate. Next, metadata, the contextual information surrounding experimental results, was infrequently captured. Additionally, accessing data for large sets of compounds and project status updates were difficult. Lastly, historical data including project decision rationale or information about project/compound failures was problematic. Additional specifics of the six categories are discussed in more detail below:

- **Information outside of my project** - The data environment at NIBR is highly organized around individual project teams. It is often very difficult for employees to access information outside of their immediate team environment. Five individuals specifically voiced their frustrations at the difficulty of finding information outside of their immediate teams:

  "I would rate the ability to find information at NIBR a 2 or 3 because it's hard to find information outside my project."

  "Data from other projects that's not already in Avalon is impossible to find without help. To get access to other team's folders or G:drives you'd have to identify and specifically ask the project team leader for permission."

  Avalon, the information warehouse for chemical compounds serves as the primary mechanism for providing cross-organization visibility outside of project teams. Avalon provides numerical testing results and is unable to capture the rich data required for drug discovery.

- **Animal and Human Clinical Data** - Information relating to compounds further along in the research development process was repeatedly mentioned as difficult to access. *In vitro* (lab/bench top testing results) were easily accessible but *in vivo* data (animal or human testing results) were almost impossible to locate for individuals outside of the immediate project team. When asked "What information is hard to find?" seven people specifically elaborated on the difficulties associated with gaining access to *in vivo* data.

  "In vivo information or data that was produced just with a program is impossible to find. You'd have to know the biologist or pharmacologist to get access to this data .Novartis is good at capturing early data (in vitro) but once compounds are tested in vivo we don't capture this systematically."

"Human PK data is impossible to get and we need to have this! I can't get a simple table of average oral bioavailability and dose for the drugs we put in phase I. I have no idea how these drugs fared in man. I don't even know what database this information is located in."

"What data is impossible to access? Animal studies that have been done with the compound."

"One of the key things I'd want in a wiki/data portal would be in vivo data generated in the disease area pharmacology groups and MAP. This data is just not accessible at the moment. It would even be useful to just know that this data exists."

"Clinical stuff is hard to access. Even if you have access it's hard to get anything useful out of this. You can't effectively search this data; it is structured by report study number. You don't know what study is for. The database is designed to please the FDA, not to help research scientists."

- **Metadata/Contextual information for assays -** Finding exact testing results, such as the numerical readout from particular assay, is easy. In contrast, finding the surrounding contextual information related to assay conditions is difficult if not impossible. Ten different individuals specifically commented on the difficulties associated with finding assay metadata such as: project and target association for each assay, information for how the assay was ran, assay relationships in the project (what was the primary, secondary/counter screen assay, etc.)

"Finding data is easy but finding the context of how it's generated is hard. An IC_{50} assay result only has value to me if I know how the assay was run, etc. I could go and read a 30 page research (RDS) document but I still don't understand the context. Why was the assay changed? This is not reported and it takes so much away from quality of information."

"Avalon gets cloudy with legacy information. You don't know how assay was run, etc. It would be nice to have easy access to these assay conditions."

"Sometimes you can't even track down how the assay was run. It would take me four days to find out how these five assays are different and how they cross compare".

- **Project Stage** –Often, information detailing what project development phase a particular compound is in is difficult to locate. In particular, two individuals mentioned difficulties with tracking project stage information for compounds that were in more advanced development stages.

"What stage is a particular compound in at any one point in time? In the later stages of development this info is hard to find. I've developed my own personal list to keep track of these. There should be a central place where this is made available to all."

- **Batch Data** - Information related to a single compound can easily be extracted from Avalon. However, extracting the same information for a series of compounds is often very difficult using the available tools and data infrastructure. Five individuals voiced specific frustration with their inability to extract batch data from Avalon.

  "Avalon is very easy to access if you only need information on ten compounds. It's impossible to access in batch way without the use of workaround tools."

  "My group is looking to identify macro level data trends. However, if you ask Avalon a research question on a large scale (1,000+ compounds) it falls apart. The system is too clunky"

- **Historical Data and Project Decision Rationale** –Historical project data is difficult, if not impossible to access. Fifteen different interviewees expressed frustration over not being able to locate rationale surrounding compound synthesis, project team decisions, or compound failures.

  "It's hard to find the rationale for the synthesis of a compound. Is this a byproduct of some synthesis route or the end goal of a problem that that the team was trying to resolve (i.e. is this a heart safety issue?) Knowing this type of information would be unbelievable!!

  "I have to go chasing chemists down and ask "why did you make this compound?"

  "It's hard to understand what the history of a compound was. You can guess from assay data or project team names. The research (RDS) reports get you a little bit farther but it's difficult to piece together the entire story of the compound. You really need to find the project team leader or someone that worked on this project."

  "Only the compounds going forward are captured – nothing documented for the ones that don't get selected and rationale for why those didn't get selected is simply lost".

Database Access and Password Protection

Interviewees were questioned about their ability to gain access to different databases within NIBR. Over 50% of users reported that data base access was not an issue and that it typically took only 1-2 days to gain access to the majority of NIBR databases. Clinical databases, outside of the NIBR organization were more difficult for individuals to access. Several interviewees reported waiting for 2+weeks to gain access to these databases. Others stated that they were simply unable to gain access to clinical data.

While database requests for access were not a major issue, password control projection was often cited as a frustration for users. Interviewees were not specifically questioned about the NIBR

password environment. None-the-less, seven users adamantly commented how often they were required to enter various passwords, multiple times, in order to gain access to data. Two interviewees went to extensive lengths to personally demonstrate the amount of password control. In the first case the interviewer attempted to access an *in silico* profiling tool by navigating through various bookmarked webpage links. Over fifteen minutes and five required password entries later, the site was finally accessed.

"You have to use different passwords to get into all these systems. Then you spend half your life typing in passwords and they all change at different times.

"Get rid of the password hell. Can we do away with this sometime? I want to be able to log in once, and not have to repeatedly enter passwords through each new web browser."

"Databases are nightmare - they ask me for password over and over."

7.5 Theme #5: Web 2.0 Awareness

Wiki Awareness

Interviewee familiarity with internal and external wikis was assessed as well as awareness of several social networking sites. Interviewees were asked if they had experience using wikis. 100% of interviewees reported having at least looked at or used a wiki for information, 61% specifically cited using Wikipedia.

Several internal Novartis wiki/web portals have been developed are currently in use within the organization (Table 7). The high throughput screening (HTS) and safety profiling wiki were developed within the Center for Proteomic Chemistry (CPC) group while the Target wiki was developed within the Developmental Molecular and Pathways, a group primarily responsible for identifying new disease targets. In addition to asking about outside internet based wikis, interviewees were polled about their familiarity and usage of these various internal NIBR Web 2.0 tools (Figure 28).

Table 7: NIBR Currently Existing Wikis

Name	Description / Functionality
High Throughput Screening (HTS) Wiki	Primarily used as a web portal. Provides HTS project updates and links to relevant project resources such as team members, relevant reports, etc.
Safety Profiling Wiki	Primarily used as a web portal. Links to safety and target information associated with safety profiling screening assays. The majority of the information is updated and provided by the CPC safety profiling group.
Target Wiki	Wiki organized around disease target. Drug discovery targets are organized by wiki page.

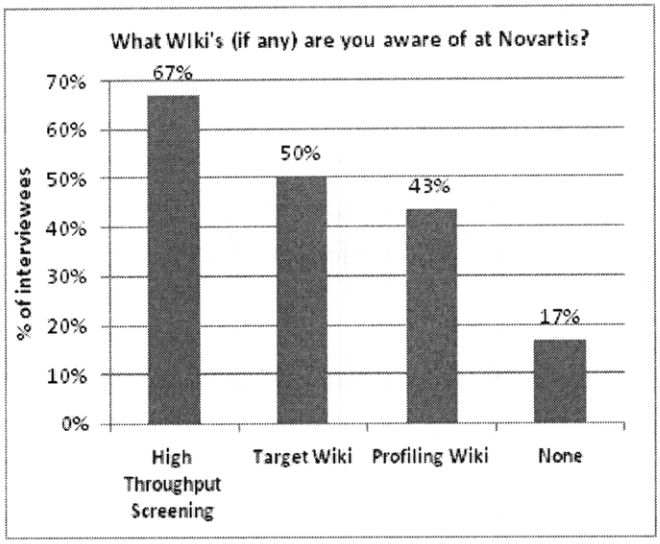


Figure 28: Interviewee Familiarity with Internal Novartis Wiki's/ Web Portals

Social Networking Tools

In addition to wiki usage, familiarity with several social networking tools was explored. Users were asked if they used Facebook (www.facebook.com) and Linked-In (www.linkedin.com), two popular networking tools. Only 15% of users reported that used Facebook while 67% of users had used Linked-In (Figure 29).

The mention of Facebook often tended to be polarizing and somewhat generational. Although the age of interviewees was not specifically asked for or recorded, younger Novartis employees reported routinely using Facebook while many older or middle-aged employees stated that they had "heard of Facebook" but choose not to use it. Two users had a very negative reaction to Facebook. Comments included: "No, I don't need something to keep track of all of my friends!!" and "No, I'm not into human networking."

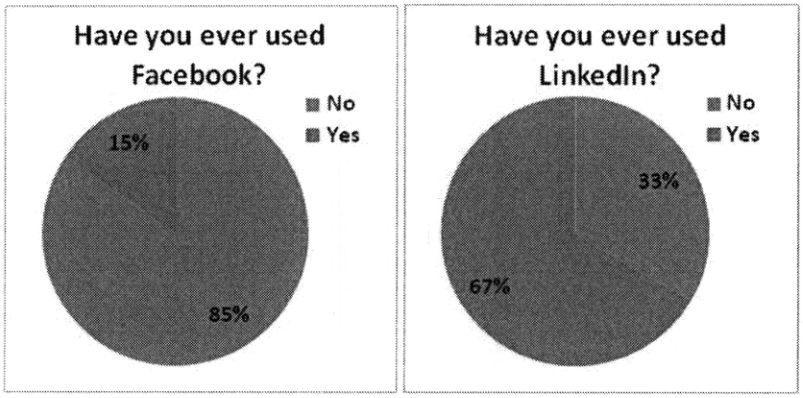


Figure 29: Facebook (n=26) and Linked-In Usage (n=24)

7.6 Theme #6: Wiki Usage and Motivations for Contributing

Motivations for Wiki Contribution

Motivation for user contributions to both general wikis as well as Novartis internal wikis was explored. Interviewees were asked if they had ever edited or helped to create a wiki. A large number of those surveyed, 44% of respondents had contributed to a wiki in some capacity. The majority of those contributions were to a Novartis internal wiki, specifically the High Throughput Screening wiki or the Safety Profiling wiki. Conversely, 56% of interviewees responded that they had never contributed to or helped to create a wiki either inside of Novartis or elsewhere. When asked why they had not contributed individuals cited a variety of reasons, the most common was that they simply didn't have the time (Table 8).

Table 8: Reasons Cited for Not Contributing to a Wiki

Reason	% of respondents (n=14)
I don't have the time, I'm too busy with other projects	29%
I don't see the value	14%
I'm not sure what I can offer	7%
It would be a conflict of interest to add knowledge to an external wiki that was associated with my work	7%
I'm behind the curve	7%
Reason was not identified	36%

Would individuals be more motivated to contribute to an internal Novartis wiki verses an outside internet wiki, like Wikipedia? Evidence of this is mixed. On the one hand, 81% of individuals reported that they would add content to a Novartis compound wiki tool if it was available. However, when asked how frequently they would contribute 60% of interviewees stated that they would contribute either monthly or less than once a month (Figure 30).

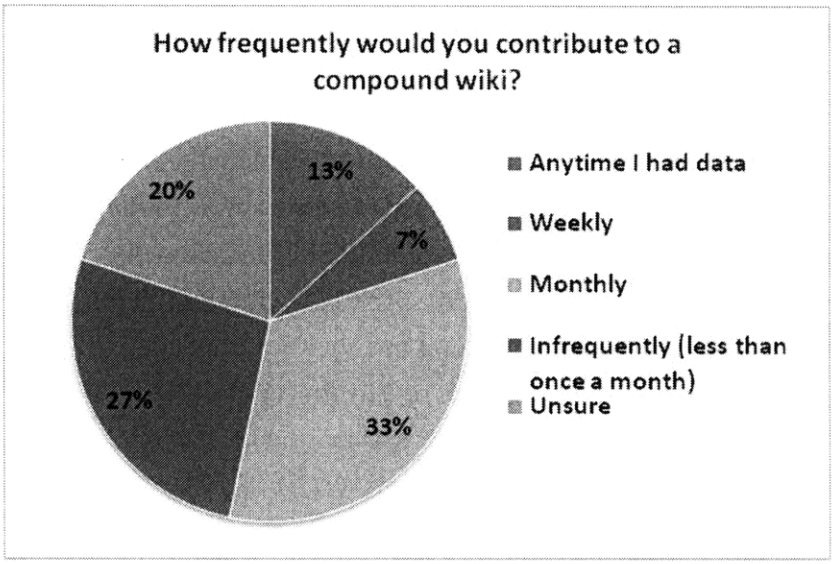


Figure 30: Anticipated contribution frequency for NIBR Compound Wiki

Critical Factors for Wiki Contribution

Several factors were identified from user comments as being critical toward obtaining adoption and usage of a compound wiki.

- **Ease of Adding Information**–Users consistently expressed their desire for a system that required minimal work to add content. Novartis employees typically already have large workloads and are consistently working under tight time deadlines. Anything that requires a significant amount of effort would face adoption hurdles.

  "To encourage adoption you should make this something that requires absolutely zero work "

  "The wiki must require minimal mouse movement to encourage people to add content. If I have to go to the wiki, find the correct compound page, then click on edit, enter a password before I can edit, forget it".

MediaWiki software, the platform used for Wikipedia, requires users to enter text in wiki syntax language (Table 9). Although not complicated, the syntax can be intimidating for users who are new to wiki editing. It also presents an additional learning curve. It is unclear that users would be willing to overcome this learning curve to contribute to a Novartis wiki. Several wiki software platforms are available that allow users to enter text in a MS Word like format for example, hitting a "bold text" button, rather than requiring user to enter **bold text**. The use of such a software editor will be critical in facilitating ease of use for data entry.

Table 9: Examples of Editing in Wiki Syntax Language [67]

What You Type	What You Get
bold text	**bold text**
__underline text__	underline text
normal^^superscript^^	normal^{superscript}

Employees expressed their desire to add comments and content in multiple ways, rather than being forced to open up a web browser and navigate to a compound homepage each time they want to make a contribution to a compound page. Requests for posting functionality included having the ability to directly post the contents of an email to a wiki page. One individual suggested having a macro included in either MS Word or PowerPoint that would allow a user to highlight specific content and post that content directly on the Compound Homepage.

"I'm already generating a lot of content about compounds in management reports etc. I'd like to have a macro in MS Word that would allow me right click on text and specify "release content to wiki." Otherwise entering in info becomes repetitive. I'm suddenly a high priced secretary putting information in different places."

- **Easily link to multiple files/formats** – Information content about compounds is currently being generated in a variety of MS PowerPoint files, Excel documents, and Word reports. Individuals will not re-create this information specifically for the wiki. Therefore, it will be imperative to create an environment that supports linking to and uploading external file content.

- **Web Portal Performance** – Page upload time will be critical for performance. Components of the Compound Homepage that extract and display information from other databases will need to load swiftly. Users will not wait for the system to pull real time information from multiple different databases. According to interviewee comments system speed and performance are highly preferred over real time, 100% up-to-date information.

- **Clear Definition of Data Sources** –Compound information at Novartis exists in a variety of databases and formats, often users rely on their knowledge of internal databases and sources to determine what information is the most accurate and up-to-date. Web portal sections of the Compound Homepage should identify the data source and if possible link back to the original

[67] Wiki Syntax, Inline Formatting, *Wikidot*, n.d. Web. Accessed 19 March 2009 http://www.wikidot.com/doc:wiki-syntax

data. An index page specifying the data source of each item would be helpful for this purpose. One individual expressed the importance of being able to identify source data:

"It's not always hard to find data but the issue can be getting "good" data that you know is accurate and reliable".

- **Integration into project work streams** – The most common theme that interviewees stressed for user adoption was the incorporation of the Compound Homepage into existing project work streams. As work within NIBR is primarily centers around project teams, a system that integrates Compound Homepage usage into the project team work flows will be the most effective for capturing compound information.

  "I'd add information about once a month about the compounds I'm working on. However, if I was actively using this for my project, I'd do it more often".

  Additionally, integrating the compound homepage into existing work streams will ensure that compound information and decisions are captured as they occur.

  "It would be difficult to make the compound and then later add comments. It would be easiest to capture at the moment you made the compound - some filed lab notebook that would feed into the Compound Homepage. People won't be willing to go back later when they've generated hundreds of compounds."

  "I've made thousands of molecules. I'm not going to go and comment on each one unless it's part of registering the compound. Otherwise, this would take too much time."

  Based on the author's observations, the existing Novartis culture often focuses on short term goals and deadlines. As such, the average scientist is not necessarily thinking about the value of capturing information for long term organization learning. Integrating compound information into the typical work flow would ensure that this information is both preserved and shared throughout the entire organization. One individual commented:

  "How should we motivate people to add content? This is the classic problem of having a rush to get to the next deadline. Everyone in the project team knows what's going on so they know why things are going on, what compounds have been made, etc. Putting info specifically into a wiki would really be for the outside world. We don't recognize the value or importance of that relative to our short term goals."

- **Ability to comment on a group of compounds**– Scientists at Novartis not only examine individual compounds, but also routinely analyze sets of compounds (Figure 25). Interviewees

consistently commented that they wanted to be able to add commentary that applied to an entire set of compounds.

"I'd like to have the ability to add the same comment to a specific set of compounds. I'm not going to individually go to each Compound Homepage and add this comment for each compound."

"On my project we are working with approximately 1,900 compounds. If you could divide list up that big list into smaller series that would be more useful. You could then have a discussion on a series, rather than an individual compound."

Potential Roadblocks for Wiki Contribution

- **Time** – Employees are consistently working under tight timelines and struggling to meet immediate deadlines. Taking the time to add commentary to a compound homepage was often perceived as a luxury. Several interviewees openly stated that adding comments to a wiki would fall lower on the priority list.

  "I love the idea of adding to a wiki but realistically I don't think people will actually have time to put information in here. It's very hard to take the time to go through and make a lot of commentary on things that have been done. We tried to do this ~ 2 years ago with Focus but this never got off the ground. It turned into more of an email exchange."

- **Compound Space is Infinite** – The Novartis compound library contains millions of chemical compounds. It will be unrealistic to expect that employees will contribute information on the majority of these compounds. The compound space is too vast in comparison to the potential number of contributors.

  "There is a good chance people won't add anything - especially for all library compounds."

- **Organizational Culture** – Organizational culture will be a significant driver for Compound Homepage adoption. The user interviews provided an extensive glimpse into the Novartis culture and attitudes about information sharing. A complete analysis of the currently Novartis culture and its implications for wiki adoption are presented in the next chapter.

Chapter 8: Novartis Culture and Three Lens Analysis

8.1 Three Lens Organizational Analysis Overview

Web 2.0 experts have previously noted that organizational culture is important for Web 2.0 adoption. [68]MIT Sloan Management Professor John Carroll advocates a three lens approach to analyzing company culture and how change will affect a given organization. He argues that it is important to consider an organization based on three distinct viewpoints or "lenses": strategic, cultural, and political. [69] This organizational analysis framework is taught within the MIT Sloan MBA management curriculum and has been applied to NIBR organization in order to analyze barriers of adoption for the Compound Homepage Web 2.0 tool. As part of the analysis specific interviewee comments are highlighted and discussed as they provide additional insight into the culture and challenges of managing organizational change.

8.2 Cultural Lens

The cultural lens component of the three lenses analysis explores organizational behavior based on cultural values within the organization. An organization's traditions, customs, accepted norms, and "the way we do things around here and why we do it" comprise organizational cultural.[70] Cultural norms and behaviors relating to data management and data sharing are explored below.

Continuation of Academic Culture

Novartis research has a culture driven by academic values and ideals. The majority of scientists are at a PhD level, meaning that they have spent 4-6 years within an academic laboratory setting getting aPhD, often followed by a 2-4 year academic post doctoral experience. As a result, the university academic lab culture carries directly over into the Novartis research organization and drives a number of cultural beliefs and behaviors.

<u>Independence</u>

Independence is highly valued within Novartis. Scientists are used to operating in the academic environment with nearly complete autonomy. They expect a similar environment in the work place. As a result, Novartis project teams operate as very independent units, and drive a project through the drug discovery path at their own discretion.

---

[68] McAfee, Andrew. Personal Interview by Juliet Duffy. Aug 08.

[69] Carroll John, "Introduction to Organizational Analysis: The Three Lenses" *MIT Sloan School of Management Working Paper.* 14, 1-13. Sloan Communications Office (2006).

[70] Ibid. (69)

Evidence of Independence:

- Novartis is starting to implement Microsoft SharePoint, a software package designed to encourage team collaboration, forums, blogging etc. SharePoint is relatively new and several groups within the Global Discovery Chemistry (GDC) organization as well as the IT group (NITAS) are attempting to provide a standardized format for the team workspace so that data search and file structures can be easily coordinated around the company. This common workspace has received push-back from several team leaders. They have strongly stated the desire to create their own unique team spaces, rather than comply with any type of pre-defined template.

General Implications:

- Independence is a double-edged sword. One on hand, it's a very necessary part of the scientific process. Some of the greatest scientific break throughs would have been impossible without independent thinking and experimentation. On the other hand, such independence hinders a coordinated effort and systematic approach to drug discovery.

<u>The Importance of the PhD and Subject Matter Experts</u>

PhD's are designated at Novartis by the title "Lab Head", a title that confers an elevated status within the research organization. PhD's make a majority of the decisions, tend to get preferential access to data, and are in charge of running each research lab. PhD's further distinguish themselves by becoming experts in certain scientific areas. Individuals have expressed hesitancy to believe data or even scientific hypotheses unless they were made by so called experts.

Evidence:

- Up until only a short time ago Avalon, the data warehouse containing information about research compounds, was only open by default to individuals with a PhD. In order for someone with BS or MS qualifications to be able to gain access to data, a PhD had to advocate on their behalf, often jumping through several organizational hoops to get their associate approval. Although this has recently been changed and many non-pHD's have access to Avalon, the mentality associated with restricted access still remains.

General Implications:

- It is the author's opinion that the organization has the potential to dismiss contributions from non-PhD associates or those without deemed appropriate level of scientific expertise.

Compound Homepage specific implications:

- How will the PhD hierarchy transfer to a Web 2.0 environment? Will individuals trust and believe data if they don't know if the post was made by a PhD or a project "expert"? Additionally, should non-experts be allowed to contribute to the user community? Will others believe and value the information that these non-credentialed employees post?

Interviewees in the user research study were specifically asked "Should the wiki be open anyone who wants to add content and links? Who should be allowed to add content and links?" Although 62% of NIBR employees believed that access should not be restricted, other individuals had options about restricting user access. 38% of individuals believed that the Compound Homepage should be restricted. Opinion was divided on if the system should just be restricted to contributions from Lab Heads (PhDs) or Project Team Heads directly in charge of a specific project (Figure 31).

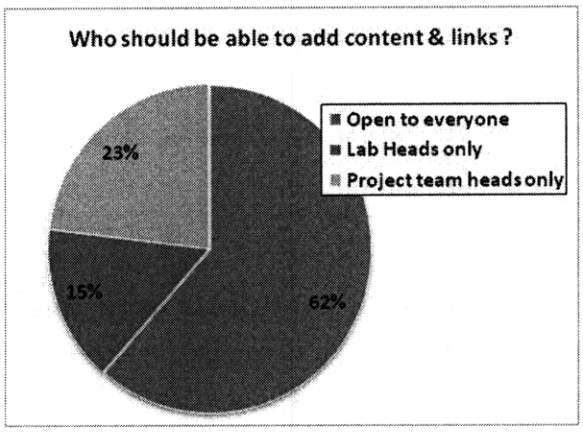


Figure 31: User Opinions on Access Restrictions for Compound Wiki

> *"The wiki should be open to everybody. People should not be afraid of adding something that is not true. You should add what knowledge you have at that time, even if it contradicts something already in the wiki."*

> *"This should be open to all lab heads and above. What about associates? Some of the associates don't know enough. You could open it NIBR wide but the chance of someone putting something stupid in here goes up astronomically. We need to have modest security. Restrict to Lab Head and above- basically all PhD's that have a reason for using it."*

Hesitancy to Share Data Prematurely/ Risk of Reputation

Academic culture fosters an aura of secrecy around data until it is ready for publication. In part, this behavior originates from the need to pre-empt other labs from publishing first or stealing information. In the academic world data is deemed ready to share when it is complete and has passed rigorous analysis and peer review. Results are then announced to the world through prestigious journal publications.

In the Academic world, sharing pre-mature data risks scientific reputation. As data is heavily scrutinized by the community, scientists are hesitant to announce conclusions to the entire world unless they feel that results are hole-proof, backed by solid research results.

Evidence:

- Several interviewees expressed their concerns about posting information that was visible to the entire organization. Fears surrounding judgment by others and reputation were specifically mentioned.

"People are reluctant to share data. They are simply afraid of sharing to a wider audience because this essentially becomes the equivalent of a publication – you'd want to run the experiment six to eight times to verify results before publishing. People want to be certain what they say is true and verified. So people won't share at all. This will inhibit free form discussion - you just can't share information very widely like that."

"I'd like to add information but it's very tricky to make it so that you're not judged, and that this is peer driven".

"User contributed part is going to be a huge hurdle. Some people are willing to contribute. Others will be worried about wiki discussion part. They don't want to look stupid, uniformed, inflammatory, etc. This is a cultural issue because contributions are not anonymous like Wikipedia. In people's minds they could be judged".

General Implications:

- Individuals spend a lot of time and effort making data ready for publication. If time does not allow for a complete analysis or report, results will often get buried on a personal hard drive and never released beyond the project team.

Compound Homepage Implications:

- The culture of data sharing must change within the organization for a Web 2.0/wiki environment to be successful. Specifically, management needs to address that it is expected for individuals to share data even if the data is incomplete or turns out to be flawed. Rewards and recognition for sharing ideas and information must be greater than negative ramifications of risk to reputation and judgment of ideas by others.
- Web 2.0 relies on user generated content and an openness to sharing data and information. Without a culture supporting this, a Web 2.0 initiative will fail. Specific measures that could ease individuals into sharing data need to be considered. For example, setting the default permissions on NIBR data folders access to open access, etc.

Data speaks volumes, but everyone is worried about interpretation

Scientists are uncomfortable releasing raw data without an extensive report or contextual annotation to explain results. In a culture where data reigns supreme, the fear is that results will be misinterpreted by those that may not have an understanding of the study or are missing the complete background and context.

Evidence:

- Throughout user interviews two individuals brought up the fear of others, specifically management, misinterpreting their data.

General Implications:

- A "data-open" environment suddenly becomes closed down for fear of misinterpretation and interference by others.

Compound Homepage Implications:

- There could be widespread resistance to sharing data on the basis that it will be misinterpreted, regardless of if this worry is legitimate or just a perceived risk.
- Web 2.0 might actually be a very useful vehicle for allowing scientists to explain their data. Those who fear misinterpretation could elaborate extensively on results or provide linkages and resources explaining how to appropriately analyze results.

Data Security and IP Protection

At first glance data security appears to be a strategic issue, a rational organizational behavior in response to threats from competitive intelligence, data leakage, patent protection, etc. Probing deeper reveals that extreme caution related to data security is sometimes part of the NIBR culture.

Evidence:

- Currently employees are required to complete training on NIBR data security. Although a necessary step in protecting Novartis data, the training makes some employees overly cautious about posting information. Many different interviewees during user research mentioned that sharing or openly documenting information in a Web 2.0 environment could be highly problematic due to security concerns.
- NIBR employees lock doors, hide documents, and restrict file access. Security concerns are cited as the rationale behind this behavior.
- The default file settings on document storage sites are "restricted access"
- Password protection is abundant

General Implications:

- Access is restricted, making it hard for individuals to find data that they need in order to perform their job function.

Compound Homepage Implications:

- Many employees will be hesitant to share data without very clear guidelines for legal/security concerns. Specific individuals have stated that they will "err on the side of caution" and simply refrain from posting information or contributing to the user community due to security concerns.

8.3 Strategic Design Lens

The strategic lens component of the three lens analysis explores organizational strategy based on rational analysis of opportunities and capabilities. The approach focuses on rational behavior and logic driven decision making.[71] Strategic opportunities associated with data management and the Compound Homepage project are addressed from a rational perspective.

Localized Data Management

Novartis managers often aim to empower scientists and try to ensure that bureaucracy and management does not get in the way of decision-making. This approach has given research scientists additional freedom to focus on what they deem important but has created localized cultures where standards for documentation and database management exist primarily at a departmental (vs. global) level. In the absence of a coordinated global guidance individual departments are developing their own ways of keeping track of data.

Evidence:

- Individuals and departments are determining rules and procedures for tracking data. Departmental databases abound as well as microcosms throughout the NIBR organization that operate very differently from one-another.

Table 10: Examples Departmental Specific Databases*

Department	Database
Developmental & Molecular Pathways	Affinity proteomics database.
Analytical Sciences	Database for storing high resolution mass spectroscopy testing results.
Clinical Development Group (TS)	Database for capturing results of clinical trials.
Program Office	Project management data.
Center for Proteomic Chemistry, PSU	Protein Crystal Structure Database.

*Note: Table is non-inclusive and meant to only provide examples

Five individuals elaborated on NIBR's localized data management:

"Everyone does their own thing, independently setting up information."

"NIBR also has an oncology group in Emeryville, CA. They maintain their own databases, own tools, etc."

[71] Carroll John, "Introduction to Organizational Analysis: The Three Lenses" *MIT Sloan School of Management Working Paper.* 14, 1-13. Sloan Communications Office (2006).

"There is no global day-to-day recognized standard for documentation. There are some standards within disease areas but these are not global."

"Novartis is comprised of many different sites and disease areas, each with its own different cultures. Mandates about documentation are pretty much driven by local cultures."

Occasionally during interviews, the topic of management involvement arose. Three individuals strongly expressed that management should play a minimal role in getting individuals to contribute to the Compound Homepage. Comments reflected that wiki contributions were not up to managers. Scientists need to determine what's useful to them and really define the wiki site and management shouldn't force this upon people.

> *"Don't have management drive the Compound Homepage. In my opinion this really needs to be a grassroots effort. Management would not be adding content so they wouldn't really have any credibility to drive this."*

General Implications:

- Novartis culture is being driven bottom-up without consistency across the organization. Multiple initiatives are being carried out by different departments without any type of centralized coordination or leadership. This leads to duplication of efforts.

Compound Homepage Implications:

- In order to realize the global scale advantages of Web 2.0, leadership must buy into the project on a global level. Economies of scale will not be achieved without organization wide user adoption and a coordinated global Web 2.0 environment.

Wasted time and resources due to data silos and lack of central coordination

Strategically, the lack of central coordination makes it difficult if not impossible to find data. This has very costly organizational consequences including a huge amount of time spent searching for data, employees developing work-arounds, lost data, and repetition of experiments. Three individuals elaborated on this point:

"Data from GNF, our sister group database is hard to access. I'd need to know to the database, have the password, etc. to get access to this information. Sometimes I might just request to have the data reproduced again at NIBR because sometimes it's easier."

Lost Economies of Scale

The organization is missing out on tremendous economies of scale that could be realized from sharing data across the entire organization. The inherent value within a large pharmaceutical company is the ability to synthesize and aggregate information from a variety of sources and allow all discovery areas to potentially benefit from that knowledge.

Costly to fix disparate systems and finally integrate data

The lack of coordination and central data storage policies has created a very tangled web of systems and databases containing information in different file formats, naming terminology, etc. It will be very costly to sort this out and integrate into one coherent system. An extensive data management initiative is already underway to begin these activities. However, it has taken the team an entire year just to start to untangle the web.

Legal ramifications regarding data sharing

There are a number of legal and regulatory issues to consider regarding documenting data and sharing information. Legally, Novartis must take into consideration what type of documentation exists around side effects or toxicity caused by different chemical structures. Documented comments openly stating that a chemical early in the research stage is "toxic" or causes "cell death" could later be misinterpreted and used against the company in a law suit if a marketing pharmaceutical product was found to have serious side effects (e.g.Vioxx). One employee explains:

"It's extremely important to capture the annotation associated with a particular comment regarding toxicity. A chemical is not just toxic in general – it's toxic in a specific cell line, under certain testing conditions, at a particular dose. The same compound under different conditions (i.e. lower dose, exposure time, etc) could be perfectly safe and very therapeutic."

Although documentation regarding safety and toxicity is a strategic concern, it also overlaps in the political fear/power arena. Individuals use the excuse of legal concerns to not document data or refrain from sharing information on compound discontinuations. Sometimes it's because they want to retain control over the data. Other times employees are not opposed to sharing information; it's just safer for them to err on the conservative side. As such, Novartis needs to set very clear guidelines regarding documentation of safety, toxicity, and side effects.

8.4 Political Lens

The political lens component of the three lens analysis explores who holds power within the organization. The political view of the organization is "a contested struggle for power among stakeholders with different goals and interests".[72] Political aspects associated with data management and the Compound Homepage project is addressed below.

Data is power

Novartis, like many highly scientific organizations, is driven by data. As such, control and distribution of data give an individual or a team power within the organization. In a completely open data environment all employees within the organization would have equal access to data. This would change the power balance.

Power Balance between IT and Science departments

The NIBR Information Technology CIO is relatively new within the organization. He has joined within the last two years and is refining the IT group's role within the greater research organization. As data associated with drug discovery is extremely scientifically complex, the IT organization must partner closely with scientific experts in order to deliver beneficial data solutions. A power struggle often arises between scientific organizations and the IT group due to unambiguous roles and responsibilities. Scientists often want (and have) developed their own programming, databases, and web solutions in the past. That said, they often lack the resources to train others, roll out solutions to the entire research organization, ensure that their specific data world integrates within the organization as a whole, etc. At the moment the power balance is pretty equal between the two departments. As the IT organization grows in employees and scope, the power balance may shift.

---

[72] Carroll John, "Introduction to Organizational Analysis: The Three Lenses" *MIT Sloan School of Management Working Paper.* 14, 1-13. Sloan Communications Office (2006).

Chapter 9: How Web 2.0 Can Help

Web 2.0 tools and technologies have the ability to significantly improve the current data environment at Novartis. This chapter explores how the use of wikis, social networking software and tagging could facilitate enabling internal Google like search, connecting people, and establishing important relationships between biological and chemical space. Secondly, the chapter enumerates how these various Web 2.0 technologies could provide Novartis with a significant competitive advantage and boost the productivity of the company's research organization.

9.1 Search

Individuals generally report a very positive search and navigation experience on the World Wide Web. Interestingly, this is usually not the case for an internal corporate intranet. Forrester Research conducted a survey on usability of the corporate intranets. Only 44% of respondents agreed that it was easy to find what they were looking for on their intranet.[73] Comments from the user study at NIBR, as well as the author's personal experience, also highlighted this view. The Novartis intranet was no different and suffered from similar navigation difficulties.

"I use Google about fifty times a day. I can find a relevant publication much faster through Google than through the Novartis internal tools."

"I would love to have an internal Novartis Google search. We really need this."

One interviewee was so adamant about better internal search that he viewed the installation of an internal Google "pointed at all of Novartis's data" as the ultimate solution to solving Novartis's data challenges.

"We have a Google algorithm problem- this is something that someone needs to solve. What is the Google application pointed at currently (within Novartis)? Could this be expanded? Can you give Google access to information that Avalon is pointed at?"

So why isn't the solution to NIBR's data management challenges to simply install an internal version of Google and allow this search engine to have access to all NIBR data? Wouldn't this instantly enable powerful search within the company firewall? Unfortunately, this is not the case. The key to Google's superior searching ability lies in its patented search algorithm called PageRank.

---

[73] Morris, Meredith, Tom Pohlmann and Oliver Young "How Do Users Feel About Technology?" *Forrester Research* (2005). Print.

Google's PageRank software algorithm gives every page on the internet a ranking of importance from 0-10. This overall page ranking is determined by analyzing the number of external links that lead to that particular website. Basically, each link is counted as a vote for that particular page. For example, a very popular website may be referenced thousands or even millions of times by other websites (via links), while an unpopular site will seldomly be referenced on other sites. The PageRank methodology goes beyond just absolute number of links and also analyses the importance of the website that casts a vote for a particular site. A vote (via a link) from a website with a higher Google PageRank will count more than a vote from a website with a low PageRank.[74]

In this manner the Google search engine capitalizes on the internet community's knowledge of what is important. People will link to content that they find useful. This creates a very powerful mechanism for filtering out good content from bad content. Harvard Business School Professor Andrew McAfee comments on this issue "if you let people author web content and you let them interlink, the cream is going to rise to the top in a sense; the most popular content will be evident very quickly."[75]Tom Malone, a professor at MIT Sloan, refers to this concept as "collective intelligence." Essentially millions of Web users have created an inherent structure in the Web by creating linkages to their web pages. It is this collective intelligence built into the linkage network that Google harnesses to create unparalleled search results.

In comparison to the World Wide Web, the Novartis corporate intranet doesn't have the same degree of cross-linking. The corporate intranet is built by relatively few individuals and has a very limited inherent linking structure. In order to change this structure many people have to be given the ability to build links.[76] One way to do this is to open the corporate intranet up to the entire Novartis community. In particular, user editable wiki's will allow employees to create web pages and link to content that they deem to be important. The creation of a large dynamic corporate internet by many different users will allow a powerful searchable structure to emerge and enable a powerful Novartis Google search engine.

---

[74] Richardson, Ben "How does Google Work? PageRank Explained" *Switch IT*, March 2005. Web. Accessed 3 March 2009 http://www.switchit.com/news/improve-pagerank.asp

[75] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

[76] Brynjolfsson, Erik and Andrew McAfee "Beyond Enterprise 2.0: The Future of the Web", *MIT Sloan Management Review* 48. 3 (2007). Print

9.2 Social Networking

The user research at Novartis showed that scientists relied very heavily on their social networks to find out information about a wide variety of topics. Roughly one third of all interviewees discussed their use of social networks in finding information.

"To find data/information I'm looking for I have to be creative and use my network. You have to call people, you have to know people. You've got to have a head start"

This suggests that the current data environment doesn't adequately support finding information. It is easier for individuals to find out information by talking to someone they know. That said, these observations also highlight the importance of face-to-face interactions or personal conversations via phone calls in order to adequately discuss problems in depth and exchange complex scientific information. MIT Professor Tom Allen, one of the first pioneering researchers of social network mapping and an expert on communication patterns in the workplace describes the importance of face-to-face interactions:

"In a workplace with a highly technical environment you need extreme bandwidth to convey highly complex information. Face-to-face interactions are the best at getting this type of information across. Other channels, such as email, simply don't have enough bandwidth." ~ MIT Professor, Tom Allen[77]

Pharmaceutical research and drug discovery is extremely complex. Scientists are constantly dealing with highly multifaceted, very involved research information. As a result, face-to-face interactions, and where this is not possible video conferences/telephone calls help to convey information most effectively.

How can a Web 2.0 site or Compound Homepage tool help scientists connect with one another? If individuals are encouraged to share information via website postings on a wiki site, isn't this counterproductive or a replacement for face-to-face interactions? Quite the contrary, simply incorporating Web 2.0 social networking tools within the Compound Homepage can dramatically improve the size and scope of an employee's network and therefore connections and access to different types of information within the company. Often, just providing a medium, online or otherwise, that connects informational experts with information seekers can facilitate further phone conversations and higher bandwidth interactions.

[77] Allen, Tom. Personal interview by Juliet Duffy. Sept 08.

For example, suppose a scientist was struggling to figure out how to solve formulation issues for a particular compound. It would be extremely powerful if they could instantaneously identify, via the Web 2.0 environment, other chemical compounds that had experienced similar formulation issues. They would then want to know the teams associated with those compounds, and contact names for scientists on those team. Given that information they would instantaneously be able to jump from reading a few comments on an online wiki to picking up the phone and having a meaningful conversation about formulation challenges and how best to tackle the specific issue.

Andree McAfee addresses the value of social networking within the business environment in his online Blog:

> *"Consider the prototypical knowledge worker inside a large, geographically distributed organization. She has a relatively small group of close collaborators; these are people with whom she has strong professional ties. Beyond this group, there's also a set that includes people she worked with on a project with in the past, coworkers who she interacts with periodically, colleagues she knows via an introduction, and the many other varieties of 'professional acquaintance.'*
>
> *Beyond this group there's a still-larger set of fellow employees who could be valuable to our prototypical knowledge worker if only she knew about them. These are people who could keep her from re-inventing the wheel, answer one of her pressing questions, point her to exactly the right resource, tell her about a really good vendor, consultant, or other external partner, let her know that they were working on a similar problem and had made some encouraging progress, or do any of the other scores of good things that come from a well-functioning tie." ~ Andrew McAfee Blog*[78]

Incorporating social networking aspects into the Compound Homepage tool would be extremely powerful. Several overlaps exist between the "compound" world and the "people" network at Novartis. These intersections should be fully exploited.

- **Identify synthesis origin** -Often, being able to track a particular compound back to the person who first made (synthesized) the compound yields a tremendous amount of information. It may provide not only the rationale for synthesis but also yield clues into what other chemical molecules were made in a particular chemical synthesis, stability and synthesis information, etc.

- **Provide project associations for compounds and link project associations to project team members.** Chemical compounds are made in hopes of creating a specific drug that

[78] McAfee, Andrew "How to Hit the Enterprise 2.0 Bulls eye" *McAfee Blog*, 3 Nov 07. Web. Access 5 March 09 http://andrewmcafee.org/blog/?p=470

will modulate the activity of a particular target. Projects teams will make hundreds, if not thousands, of compounds over the lifetime of the project. Associating particular compounds back to individual projects can yield rich information about why a project team was interested a certain compound, what scaffold groups were considered, what modifications were made to the compound to improve certain drugable properties, etc. Further linking this information back to the individuals on the particular project will facilitate this type of discussion.

- **Identify all individuals who post wiki comments on a compound page.** Adding an editable wiki page specifically for one compound allows individuals with knowledge about that particular compound to self identify themselves as experts. An electronic signature/name identification should be mandatory for all who leave comments. Linking the name of the individual back to the company directory will facilitate further interactions.

9.3 Categorization by Tagging Compounds

Implementing tagging, another powerful Web 2.0 tool will help to organize the vast amount of content and make the Novartis data world more searchable. To gain an understanding of tagging it is useful to explore a case study on tagging digital photos and then transfer this analogy to the compound world. The results demonstrate the powerful nature of tagging and how the ability to digitize information is vastly changing how the world organizes information.

The amount of data in the world is vastly expanding. A popular video featured by syndicated radio talk show host Kim Komando describes this trend of digital proliferation. "It is estimated that 4.0 exabytes ($4.0*10^{19}$) of unique information will be generated this year, more than the previous 5,000 years. The amount of new technical information is doubling every two years."[79]With so much new information how are people keeping track? In his book "Everything is Miscellaneous" David Weinberger explores this new digital age and examines how the messiness and miscellaneous nature of the digital world is leading to new ideas, efficiency, and social knowledge.[80]

David points to the organization of photographs as a prime example of information proliferation and how the ability to categorize items in many different ways is having transformative effects. Information from David's book is presented below and expanded upon in order to illustrate how tagging has been applied to sorting and sharing photographs and what revolutionary implications this could have for the organization of chemical compounds at Novartis.

Photo Sharing and Sorting

The advent of digital photography has made the incremental cost associated with snapping that additional photo virtually zero. Ten years ago, in the era of Kodak film an individual might take several rolls of film, roughly 48 photos during a typical vacation. Now, particularly for the photo enthusiast, it is not uncommon to capture an entire digital photo card, 2 GB which for an average picture resolution would be roughly 2,000 photos. Recently the author's fiancé, a photo enthusiast, succeeded in exceeding the counting feature on his digital camera. He has taken so many photos on his digital camera that the digital number scheme has cycled through the allotment of file naming numbers, (i.e. IMG_89732) and is now back to labeling photos IMG_00001, causing file conflicts when photos are uploaded to the same hard drive that previous photos name IMG_00001 were

---

[79] Fisch, *Karl Living in Exponential Times*. The Kim Komando Show, n.d. Web Video. Accessed 17 March 2009 http://videos.komando.com/2008/12/22/living-in-exponential-times/ Original Video Source data: http://www.lps.k12.co.us/schools/arapahoe/fisch/didyouknow/sourcesfordidyouknow.pdf

[80] Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder.* New York, Henry Holt and Company, (2007) Print.

stored on. Even non photo enthusiasts that don't exceed the digital numbering allotments on their cameras are inundated with literally thousands of photos on just a yearly basis. Organizing all of these photos, with names like IMG_89732.jpg, has become a royal headache even for the most compulsive organizers. As a result, the labeling of photographs has begun to evolve into a social process with others pitching in to help us organize.[81]

Flickr.com, a photo sharing website, is a perfect example of how tagging is being used on a community wide scale to assist with organization. Individuals post their photos on the site and apply descriptive labels called "tags". Not only can a user apply tags to their own photos, but community members can tag others' photos adding to the descriptive information. Essentially, the categorization is both defined by and outsourced to the user community. The user categorization scheme that emerges is referred to as *folksonomy,* a categorization developed over time by "folks". In contrast to *taxonomy*, the up-front organization scheme developed by experts, folksonomy is a bottom-up, user driven classification system that reflects the relationships that people actually use.[82]

Tagging benefits extent beyond simply outsourcing the work to the user community. Tagging breaks down the constraints of the physical world and allows information to be classified and stored in multiple places simultaneously. "Instead of items being placed in one particular area, or occasionally in two, they can be classified in every different category in which users might conceivable expect to find them."[83] This is very powerful because without an upfront, imposed first order structure individuals with different needs can search through information in different ways to most effectively meet their needs.

In order to demonstrate this concept, consider a hypothetical example of a collection of photographs taken during a student's time at MIT Sloan. Over the course of two years of study such a student might accumulate a multitude of photos relating to specific classes, student orientation, extracurricular activities, hanging out with friends, etc. While the entire photo collection is no doubt extensive, for the simplicity of this demonstration just six photographs have been chosen below (Figure 32).

---

[81] Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder.* New York, Henry Holt and Company, (2007) Print.

[82] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

[83] Ibid. (81)

Each photograph represents some aspect of student life at MIT and has been annotated with *metadata*, essentially information about the particular photo. This metadata is organized into tags, brief descriptions about each photo. For example, photo #1 is a picture of the "beer game" a game played by all incoming business students at the MIT Sloan student orientation. The game is a representation of a beer factory and demonstrates how small variations in demand are magnified throughout the distribution supply chain producing an uncontrollable, highly variable system. Photo #1 has been tagged "team work" because students play the game as a team. It is also tagged "Juliet Duffy", the name of the individual we see playing the game. The "Sloan orientation" tag refers to the fact that the beer game is played at during the Sloan student orientation. "Supply chain" and "bullwhip effect" demonstrate the operations concepts learned in the game. The other photographs within the collection have been tagged with similar descriptors providing additional information about each photo.

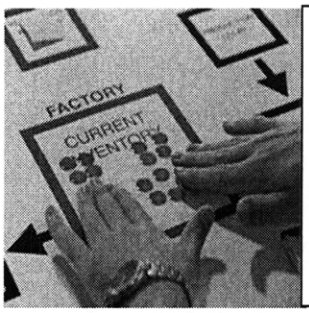


Photo #1
MIT
SloanOrientation
Team Work
Beer Game
Juliet Duffy
Bullwhip effect
Supply chain

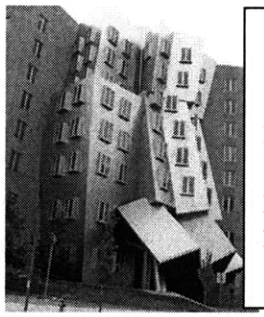


Photo #2
MIT
Stata Center
Buildings
Student Center
Frank Gehry
Kendall Square

Photo #3
The lab – Nerd Golf
MIT
Stata Center
Sam
Abe
Mini Golf
Cambridge

Photo #4
MIT
Sloan Turkey Trek
Spring Break
Istanbul
Juliet
Brede
Stacy

Photo #5
MIT
Sloan Turkey Trek
Spring Break
Istanbul
Constantinople
Buildings

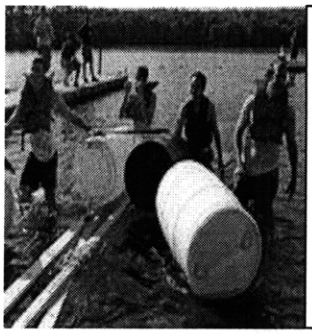


Photo #6
MIT
Sloan Orientation
Abe
Juliet
Jason
Team Work
Rafting
Water

Figure 32: MIT Sloan Photo Collection

Consider how one might organize these photographs in a photo album. In the physical world, you are forced to choose an organizational scheme. All of the photos relating to spring break might go into one physical album entitled "Spring Break 2009"while the other random MIT photos might be organized in a general album simply called "MIT photos." Each photo would be classified in a single album as it could only occupy one physical slot in a photo sleeve.

Now suppose that over the next semester the student decided to take a class on architectural themes and motifs. Suddenly, the photographs relating to buildings might be very valuable. Among thousands of photos it might be impossible to locate these without a significant amount of effort. However, digitizing the photos and annotating them to with tags related to their content would make such a search extremely fast and simple. One could type in "buildings" and instantly photo #2 and photo #5 would be displayed from the collection.

In the digital world these photos could simultaneously exist in many different categories. Rather than imposing an upfront organization scheme by placing the photos in a particular album, the photos would exist as a miscellaneous giant pile and could be instantly organized to suit an individual's particular needs of looking at the data. What albums would be useful from the above collection? Perhaps, an administrator from the school wants to use the photographs. They might be interested in assembling examples of team activities to showcase how MIT Sloan is highly focused around teamwork. On the other hand, they may want photographs depicting the 1st year orientation activities. A different stakeholder, a student named Abe, might just want to sort through the collection and identify photographs of himself. The possibilities are endless.

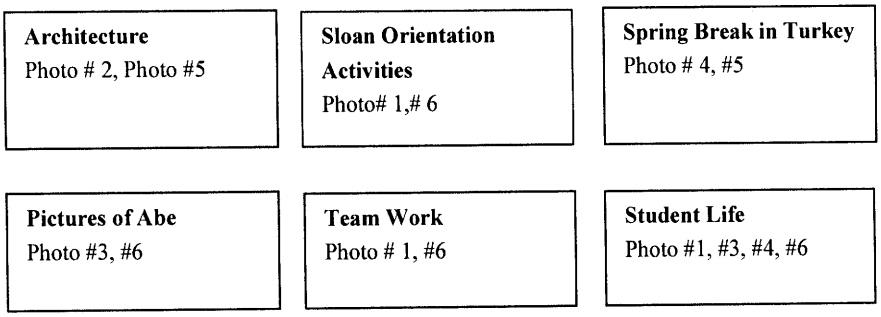


Figure 33: A Few Photo Album Possibilities

The ability to sort and instantly classify different photo albums is compounded in a community environment where millions of users come together to each add their photos and tags describing those collections. Flickr (http://www.flickr.com/) is a prime example of this phenomenon. The site has over 225 millions photos already uploaded with roughly 5,000 new photos uploaded every minute.[84] Community users are entirely responsible for the categorization of content via tagging. Remarkably, photos pertaining to any topic imaginable are easy to find.

"On Flickr it's remarkably easy to find photos on almost any topic and pull together collections of photos on these that mix and match those topics. Want to find photos of dogs wearing red clown noses? A search at Flickr finds twenty one of them. Researching car-crash art? Flickr finds two hundred and thirty three photos that may help your studies."[85]

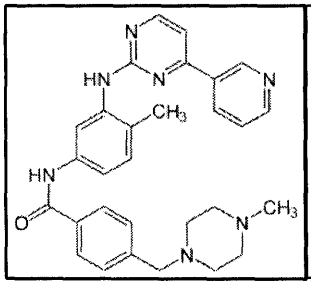
Compound Sharing and Sorting

So how are the lessons related to tagging and photo sharing on Flickr applicable to the pharmaceutical industry and Novartis's quest for discovering life saving medicines? The lessons in data organization are remarkably powerful when applied to chemical compounds. At the heart of the pharmaceutical industry are individual chemical compounds. Once a biological disease target has been discovered the quest begins to develop and find one specific chemical compound that will change the activity of this particular biological target while having minimal side effect profile. How big is Novartis's photo, or chemical compound collection? Currently, the Novartis screening library consists of over a million compounds. What if you could categorize and capture information about every single chemical compound in a way that is similar to the photo sharing on Flickr? Compounds could be tagged with not only scientific information, but also associations such as people that have made the compounds, projects that compounds were used in, etc.

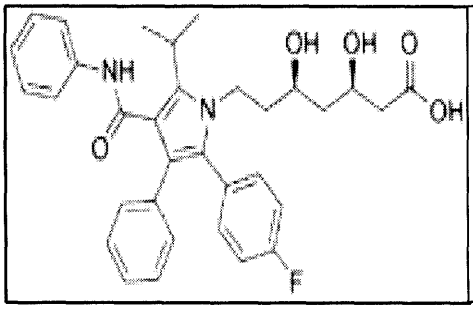
Six compounds are presented below with various tags (Figure 34). Similar to assembling different digital photo albums, the compounds could be instantaneously grouped according to user needs. It is useful to explore what specific "albums" could be assembled from the tagged information in the theoretical compound collection.

[84] Welcome to Flickr. *Flickr*, n.d. Web. Accessed 19 April 09. http://www.flickr.com/

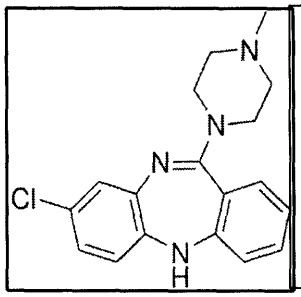
[85] Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder.* New York, Henry Holt and Company, (2007) Print.



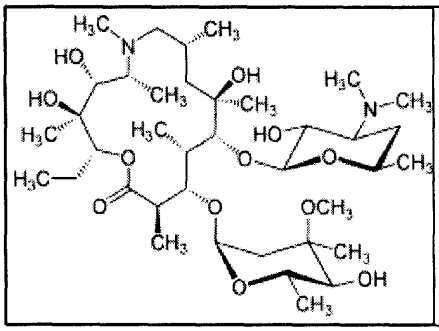
Compound #1
Project X2751(Series A)
Stage: CSP
Origin: NIBR (Basel)
Pathway: 243
Target: GRX
Metabolite: HGR



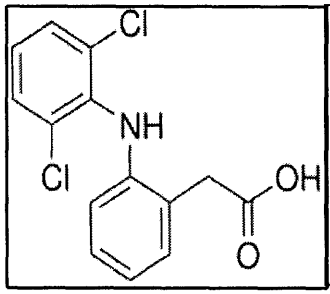
Compound #2
Project H6643 (Series C)
Stage: D2
Origin: NIBR (Basel)
Pathway: 478
Target: RRQ
Metabolite: MXA



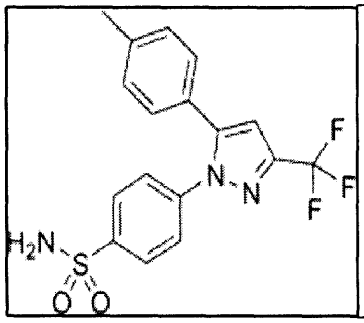
Compound #3
Project X2751 (Series B)
Stage: D3
Origin: Contract lab ABC
Pathway: 112
Target: XYZ
Metabolite: ABC



Compound #4
Project Q4729 (Series A)
Stage: CSP
Terminated, phototoxicity
Orgin: NIBR (Cambridge)
Pathway:579
Target: TRH
Metabolite: IWQ



Compound #5
Project R4551(Series B)
Stage: D3
Origin: Contract lab ABC
Pathway: X
Target: XYZ
Metabolite: ABC



Compound #6
Project X2751 (Series B)
Stage: D2
Cancer
Origin: NIBR (Basel)
Target: LKW
Pathway: X
Metabolite: RYT

Figure 34: Theoretical Novartis Chemical Compound Collection[86]

Note: the compound structures and resulting tag data are for illustrative purposes only. Data is completely arbitrary and does not in any way represent real relationships or chemical compounds in development at Novartis.

[86] Chemical structural pictures were taken from approved pharmaceutical compounds listed on Wikipedia. The associated tagging relational data is entirely arbitrary and made-up for concept demonstration purposes.

Project association –Project development teams routinely keep track of the chemical compounds that they have synthesized. Several teams within the user interviews indicated that they had assembled chemical lists of compounds relevant to the project team. These lists were further segregated into different chemical "series" based on chemical scaffold similarity (refer to Figure 24 for an illustration of how compounds are organized by chemical series). Although project teams have an intimate knowledge of the chemical compounds that they are working on, there is very little visibility for this in the organization at large. Tagging compounds with both the project name and the chemical series within the project would be useful to both project team members and scientists in the organization at large.

Album compiled:

Album #1 **Project X2751, Series B** Compound #3, #6

Potential Album users:

- A scientist within project team X2751 wants to examine all of the chemical compounds that the team has synthesized with core scaffold structure series B. He is compiling a PowerPoint presentation relating to some structural activity relationship (SAR) studies for an upcoming meeting and wants to make sure that he includes data for all relevant changes to the core scaffold.

- A scientist outside of the project team is analyzing high throughput screening results from her project. She notices that several of the compounds that had significant activity in her screening assays were extensively investigated as part of another project (X2751). She wants to identify all of the other compounds that were investigated as part of the X2751 project series and see how that data might relate to her specific project.

Development Stage –The amount of scientific information available for a specific compound is directly related to the compound's development phase. A molecule that has only been used in D_2 high throughput screening studies will have very little information compared to a molecule that made it into the candidate selection point (CSP) development phase. In order to identify compounds with a significant amount of information, it is useful for scientists to be able to view chemical structures associated by development stage.

Albums compiled:

Album #2 **Stage: CSP** Compound #1, #4

Potential Album users:

- A Scientist trying to narrow down hits for their particular project.

"I'd like to be able to see if a compound was selected for CSP. CSP is more of a highly refined compound - people have taken a close look at this in the past. If this was a hit I'd go and look at this again in our assays because there would be tons of quality information that I could use for this compound."

Terminated Compounds –The vast majority of chemical compounds never become a pharmaceutical drug. For every 5,000 – 10,000 compounds tested, 5 of those will reach clinical trials and only one will become an FDA approved drug (Figure 7).Compounds could be tagged if they were terminated in a project after reaching a particular development phase (D3 or CSP).Information about compound termination is currently tracked within a project management database but has poor visibility within the broader organization. Providing both global identification and analyzing this select group of compounds would yield a tremendous amount of information and benefit multiple individuals within the organization.

Albums compiled:

Album #3
Terminated in CSP, phototoxicity
Compound #4

Potential Album users:

- A scientist analyzing screening results and deciding what compounds to select out of the screen.

- A safety group, or quality team (i.e. quality plus) establishing additional tests and mechanisms to identify phototoxicity or other issues earlier in the development process.

- Groups investigating compound repurposing (business licensing department, a number of scientific organizations). Science develops and changes at an extremely rapid rate. Compounds are often discontinued for reasons or problems that either might not be relevant in other disease indications, or can be solved at a future data using technology that might not be foreseeable at the current time. Both internal organizations as well as the business licensing group often go through and "re-purpose" compounds, re-using these valuable assets for other internal projects or licensing them outside of the company.

Validated Target or Pathway Hits– A specific biological target is screened in high throughput assays against millions of chemical compounds in order to identify "hits," specific compounds that bind to and modulate activity of the particular biological target. A chemical compound often binds to many different targets. This has implications for side-effects and also potential use of the drug in other disease indications.

Albums compiled:

Album #4
Validated Target hit XYZ
Compound #3,#5

Album #5
Pathway hit: Pathway X
Compound #5, #6

Potential Album users:

- A scientist in the safety profiling group has just discovered that a new enzyme, Target XYZ, is responsible for producing a nausea related drug side-effect. They would like to investigate all other Novartis compounds that are known to hit this enzyme.

- An early stage biologist is trying to understand a specific biological pathway. Knowing all the compounds that change activity in this particular pathway would shed clues into pathway mechanisms and lead to increased disease understanding.

Metabolites - When the body processes chemical compounds, the chemical structures are often broken down into smaller components referred to as metabolites. Reactive metabolites can be toxic and cause a number of different side effects.

Albums compiled:

Album #6
Metabolite: ABC
Compound #3,#5

Potential Album users:

- A project team is struggling to modify a chemical compound so that it does not cause a reactive metabolite. The team would like to ask "what other compounds have resulted in this specific metabolite? How did the project teams of those compounds solve this issue?"

Associating Metadata via Tagging

As demonstrated, adopting a Flickr like tagging approach for chemical compounds could be tremendously valuable. Tagging categories would expand beyond those listed in the hypothetical

example above. In fact, the more information that Novartis can associate with each chemical compound, the more powerful the system will become.

David Weinberger elaborates on this point by bringing up the example of library catalog cards.[87] Before the digital age, library card catalogs held the metadata associated with individual library books. Each library book had an associated catalog card would list a book's author, subject, title, publisher, etc. Books could then be sorted by these various categories in different card catalogs. The amount of metadata that could fit on each library card was limited by the physical card space. Cards larger in size than a large index card would simply be unmanageable. In the digital world, the amount of metadata that can be associated with a given object is quite considerable. As storage has dramatically dropped in price, it is no longer expensive to capture large amounts of metadata.

In the digital world, the more information that can be associated with a particular chemical compound, the easier it will be to find and make associations about that particular compound. Indeed, "the solution to the overabundance of information is more information." [88] Our large unwieldy digital photo collections become suddenly manageable when we extensively tag the individual photos with many pieces of descriptive metadata.

Should all of the tagging information be added by the user community? On one hand, it makes sense to leverage the Novartis research community at large to help organize and categorize chemical compounds. Often a scientist will have very in depth insight into a handful of chemical compounds that he/she has worked on extensively. Additionally, scientists are already compiling lists of compounds that share certain properties. For example, a Developmental & Molecular Pathways scientist was introducing compounds to cells then observing how gene transcription changed as a result of the compound addition. He had compiled a rich network of clusters compounds by associating them with this gene transcription data. Capturing such information in tagging networks would be tremendously useful to the organization at large. User tagging is perfect for a large portion of knowledge related work that is often unpredictable, arises on an ad hoc basis, and demands flexibility.

In other instances data might be more static, already occur in a pre-existing database, or be more easily generated by a computer algorithm. For example, tagging compounds as validated hits against a specific target would make more sense to automate. It will be necessary to allow

---

[87] Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder.* New York, Henry Holt and Company, (2007) Print.

[88] Ibid. (87)

compounds to be tagged automatically by computer searches and algorithms as well as by individual users within the Novartis community.

Why is Tagging so Powerful?

Tagging in part is so powerful because it allows users to slice the data in every conceivable way possible. As demonstrated above, different users within the organization will want to answer very different scientific questions. The MIT admissions officer does not necessarily care about the photographs of various architectural buildings but is very interested in the pictures that depict the Sloan orientation activities. Rather than trying to anticipate user needs and impose an upfront organizational scheme, tagging makes all of the information accessible and lets users decide how to explore relationships.

Such an organizational structure supports random browsing. Consider the difference between exploring books on Amazon.com and within a conventionally organized book store. In a brick and mortar bookstore the books are carefully categorized into specific sections: History, Autobiographies, Cooking, Travel, Fiction, etc. A history buff interested in the US conflict in Vietnam might thoroughly explore the history section but would miss an equal book of interest, "Miss Saigon" a fictional musical about the cultural effects of Vietnam war, because it was classified in music and film section.[89] However searching for Vietnam books on Amazon Title on Amazon yields an amazing assortment of books. This is because books can be classified in multiple places simultaneously. The user might not only discover "Miss Saigon" but could come across a cooking book exploring Vietnamese cuisine or a travel book detailing a writers recent journey to historical points of interest. Furthermore, Amazon takes tagging one step further and automates tagging relationships through the use of extensions. The Amazon algorithm says to users "attention: purchasers of this book "x" also typically like book "y."Suddenly, a reader is enticed to browse and stumble upon new content. Similar "browsing" within the scientific environment could easily lead to the discovery of new powerful relationships between compounds, side-effects, and diseases that would not have originally been envisioned.

[89]Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder.* New York, Henry Holt and Company, (2007) Print.

How are Web 2.0 tagging approaches different from database searching?

Firstly, databases are highly structured and impose a pre-specified organizational structure. Wiki inventor Ward Cunningham elaborates "For questions like 'What's going on in the project?' we could design a database. But whatever fields we put in the database would turn out to be what's not important about what's going on in the project. What's important is often the stuff you don't anticipate." [90,91] Scientists are constantly exploring new relationships or have a desire to ask questions that might not be anticipated by the IT group. Furthermore, IT resources would be instantly overwhelmed if every scientific group within the organization required the setting up of a database to capture and use specific information. For example, during user interviews a scientist within the protein structure unit group made the following request:

"Our group also generates a lot of information about compounds that fail in some way. These are compounds that won't co-crystallize with the protein. I'd love to know if a crystallographer over in Basel looked at it, even if it failed. I could learn a lot from his experience."

Would a database get set-up to capture this information? This is highly unlikely. However, user initiated tagging or comments on a wiki would make this instantly easy to capture. The scientists could annotate the compound with "Attempted to crystallize protein X with this compound. Attempt failed." The information could be captured and instantaneously make available to the entire organization. No phone call to IT, or approval for a special database to capture this information would be required.

Secondly, the current NIBR database environment is very disjointed. Individual databases are often set up, run, and maintained by individual departments. It is currently impossible to link between these various databases. A tagging structure would allow for the exploration of relationships without necessarily having to link together these discrete databases.

For example, suppose a scientist was studying a disease that required the inhibition of a kinase. In order to visualize how different compounds bound to kinase inhibitors it might be very powerful to have a list of all compounds within the organization that were validated hits against kinase inhibitors and had also been crystallized. Finding this list of compounds in the current data structure would be impossible. Assay results would be stored in Pharon, a list of specific targets that were kinase inhibitors would be in another database (if it existed at all) and all crystallized

---

[90] Venners, Bill "Exploring with Wiki: A Conversation with Ward Cunningham, Part I" *Artima Developers*,(2003). Web. Accessed 12 April 09 http://www.artima.com/intv/wiki.html

[91] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

compounds would exist in yet a different database. However, consider the situation if the compounds had been tagged. Finding this list of compounds would be as simple as searching for the intersection of the tags "crystallized compound" and "validated hit against kinase inhibitor." The list of results compounds could then be further explored. The actual crystallography data would be available through the compound homepage linking page that would lead into the crystallography database.

Another example of data connections made possible by tagging is demonstrated by a scientist interested in improving the safety testing assays for cardiac irregularity (hERG activity). The individual would be interested in asking 'What compounds have tested positive in hERG*in vitro* assay A and have been evaluated in dog telemetry studies'? The current approach to answer such a question would require tedious searching between two unconnected databases. Tagging compounds "positive result in hERG assay A" and "dog telemetry study completed" would instantaneously allow the intersection of each group to emerge and encourage further exploration of the identified compounds.

9.4 Competitive Advantage of Web 2.0

Web 2.0 technologies have the ability to provide Novartis with a significant competitive advantage by vastly improving the data environment and boosting the productivity of the company's research organization.

Decentralizing Information - One of the benefits of Web 2.0 technologies is that they provide a common platform for knowledge work that is both visible and accessible to all areas of the organization. This can help break down organizational silos and give individuals access to relevant information regardless of where they are within the organizational structure.

"Hierarchies and organizational silos short-circuit tacit interactions: information moves up and down a hierarchy at defined management levels. By contrast, to stimulate interactions organizations want whatever information is relevant for solving a particular problem to be shared among team laterally, in real time, irrespective of reporting channels and silos." [92]*~McKinsey Quarterly*

Network Effects- Web 2.0 technologies have significant network effects. The more people that add linkages, author, tag, etc. the more powerful the network becomes. Historically, increasing an organization's size has made it more difficult to find information. Web 2.0 technologies act in the opposite direction and make larger organizations more searchable and navigable than smaller ones.[93] Suddenly, size is no longer a hindrance but a significant competitive advantage.

Collective Intelligence–Research talent is highly concentrated within NIBR. The majority of employees have PhD's and years of experience in highly concentrated scientific specialties. The collective intellectual capacity is on par with, if not exceeds, certain major academic research institutions. As most employees could single handedly publish subject matter text books, imagine how powerful and knowledge rich a Novartis compound Wikipedia could become if even a fraction of those individuals dedicated their time toward authoring content. Furthermore, what better expertise to establish a dense linking and tagging network based collective knowledge?

Scientific Relationships - A manager within the Global Discovery Chemistry organization powerfully summarized how discovering relationships across biology, chemistry, disease areas can create significant competitive advantage:

---

[92] Beardsley, Scott, Bradford Johnson and James Manyika "Competitive advantage from better interactions" *McKinsey Quarterly* 2 (2006). Print

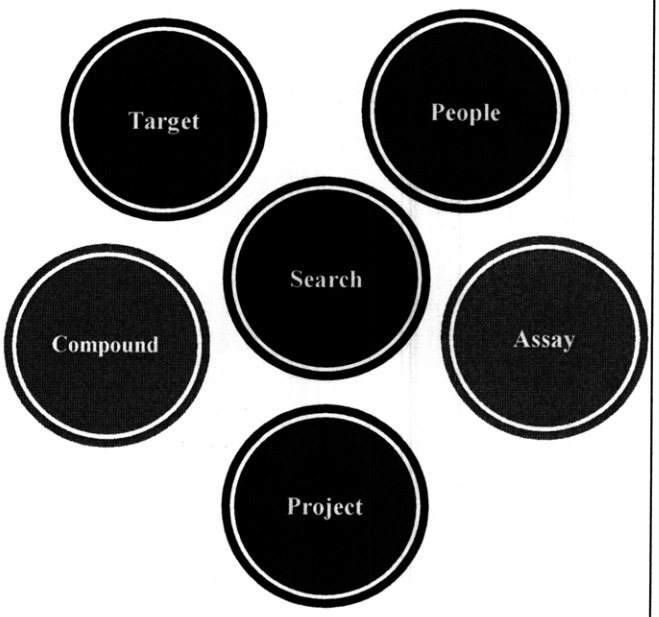
[93] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

"What is Novartis's competitive advantage? Pharmaceutical labs in India and China can synthesize chemical compounds by the thousands for a fraction of the cost. Our real advantage is our wealth of historical data, being able to look at the data as an aggregate and create linkages across different diseases, targets, etc. which is impossible to do without an organized, open data environment."

Chapter 10: Project Homepage

10.1 High Level Overview: Projects, Compounds, Assays, Targets, People

Although this thesis focuses on defining how Web 2.0 could be used to specifically organize chemical compounds, it is appropriate to step-back and present a quick high level overview of the envisioned Novartis wide scientific web environment. Information at Novartis tends to fall within several different broad categories or "bubbles". Each of these categories would serve as logical entry points into the Novartis Web 2.0 system. [94]



Target–As defined by Wikipedia, a target is a biopolymer such as a protein or nucleic acid whose activity can be modified by a pharmaceutical compound. Early in the development process project teams focus on identifying targets and pathways that are associated with diseases.

Compounds- The chemical molecules that modify the activity of biological targets. Once a target has been identified efforts turn to discovering (stage D_2) and optimizing (stage D_3) chemical compounds.

Assays- Assays are the specific tests where compound activity against a specific target or pathways is tested.

Projects- Projects refer to the inherent project teams working to develop a drug for a particular target.

People- Refers to the individual employees at Novartis

Figure 35: High level Novartis Web 2.0 Organizational Structure

Although the Novartis wiki/Web 2.0 system would essentially be one giant organizational network, it is useful to break down the system and examine how the environment should be structured based on the different categories defined in (Figure 35). The table below lists a proposed description of the organizational framework for each "bubble" and suggests individuals within the organization that would be the primary users.

[94] "Bubble" concept envisioned by Dejan Bojanic, The Novartis Institute for BioMedical Research, Center for Proteomic Chemistry

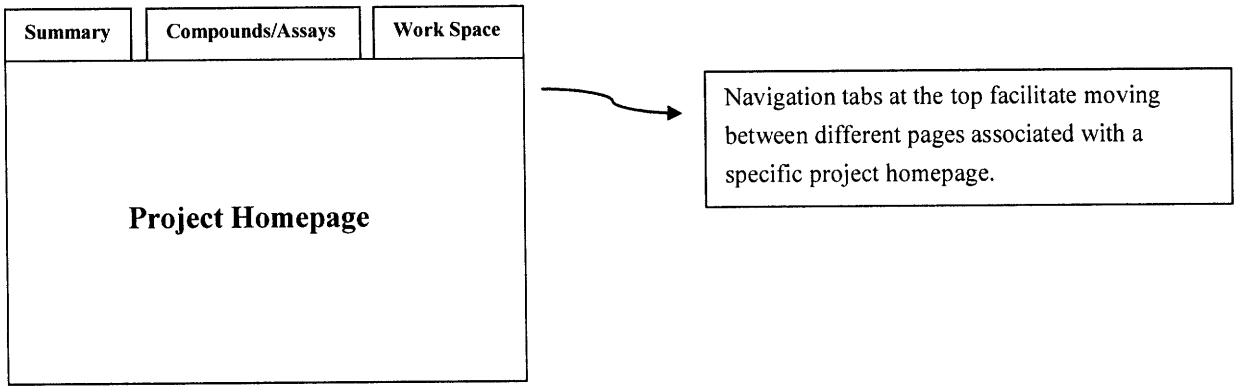
Table 11: Description of Proposed Novartis Web 2.0 Networks

Network /Entry Point	Description	Individuals that might use this entry point / framework
Target	Wiki pages structured around targets - each biological target would have a specific "homepage" containing & linking to information relevant for that particular target.	Biologists, early stage researchers
Compound	Wiki pages structured around compounds - Every chemical compound within the Novartis screening library would have a specific "homepage" containing &linking to information relevant for that particular compound.	Chemists, later stage researchers
Assay	Wiki pages structured around assays - Every assay within the Novartis would have a specific "homepage" containing & linking to information relevant for that particular assay.	All scientists
Project	Project team spaces – Every project team would have team space. In addition to capturing project updates, project status, listing team members, this wiki would be a virtual working environment for project team members.	Project teams, management
People	The Novartis internal directory. In addition for providing contact information for individuals such as phone, email, mailstop, etc. the hub could serve a social networking environment allowing scientists to connect with one another, tag associations "I've working on project x", "I have expertise in Y" etc.	Everyone in NIBR

Although different Web based "bubble" environments will be important for the Novartis organization to define and develop, only the Compound and Project Homepages were investigated as part of this thesis. Project Homepages are touched upon in relation to Compounds as the both of the "project" and "compound" web environments must be tightly integrated.

10.2 Project Homepage Overview

Each project team at Novartis should have a designated "Project Homepage". This would serve as a user editable knowledge management platform where the team could actively collaborate and capture their work. The project team page would be structured as a series of associated team web pages with navigation tabs at the top.



10.3 Project Summary Page (Tab#1)

This would be the 1st page /tab displayed when visitors navigate to the Project Homepage site. The Project summary page would give a brief overview of the project, target associated with that project, project phase, etc. Project team members would be responsible for updating all information. Webpage would serve as an interface for the new project management database.

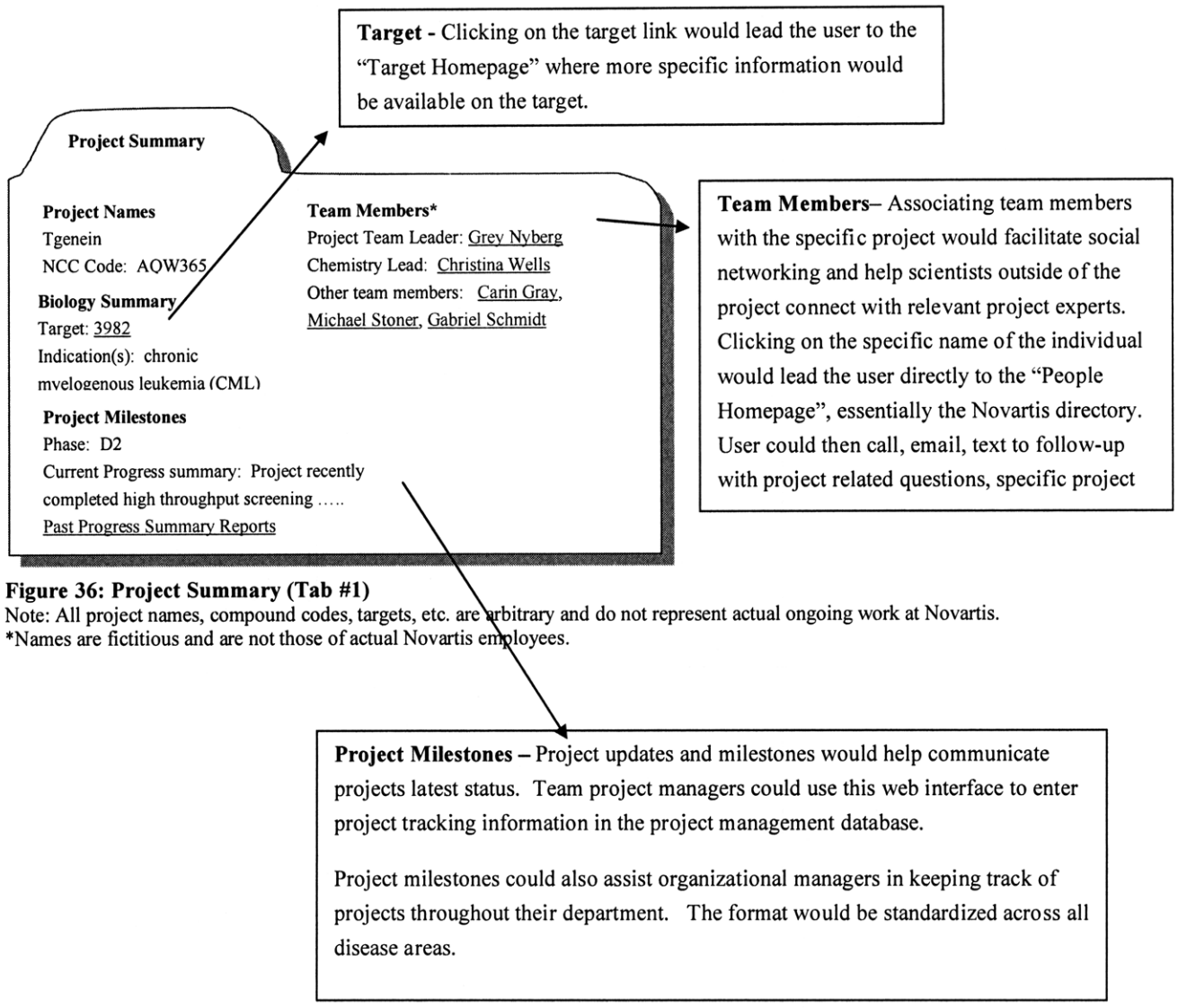


Figure 36: Project Summary (Tab #1)
Note: All project names, compound codes, targets, etc. are arbitrary and do not represent actual ongoing work at Novartis.
*Names are fictitious and are not those of actual Novartis employees.

10.4 Project Compound/Assay Page(Tab#2)

Central to the concept of the project homepage would be a mechanism for keeping track of all chemical compounds that are associated with the team project. [95]Although not shown in the figure, assays associated with the project could easily be tracked in a similar manner.

Tagging Compounds –The "Add Tag" button would allow the project team scientists to easily tag compounds with additional information. For example, the majority of team members in the user interviewers organized compounds by chemical series. Users could add the chemical series name (e.g. Ortho-biarylcarboxamides)and then simply sort the compound list by chemical series. Tags could be added to multiple compounds at one time simply by checking the box to the left of each compound.

Adding/Removing Compounds – Team members would add new compounds to the project list by clicking the "add" button and typing in the NVP # for the compound that they wanted to add. Similarly, compounds could be removed by checking the box next to the compound and clicking "remove"

Adding a compound to the project list would automatically tag the compound with associations to a specific project, target, disease indication, current project development phase, etc. This would make tagging with project associations part of the workflow and thus eliminate the need for individuals to specifically navigate to the Compound Homepage to tag compounds with project associations. Similarly removing a compound from the list, particularly if the compound was in a more advanced development stage, would automatically tag the compound with a "termination" tag.

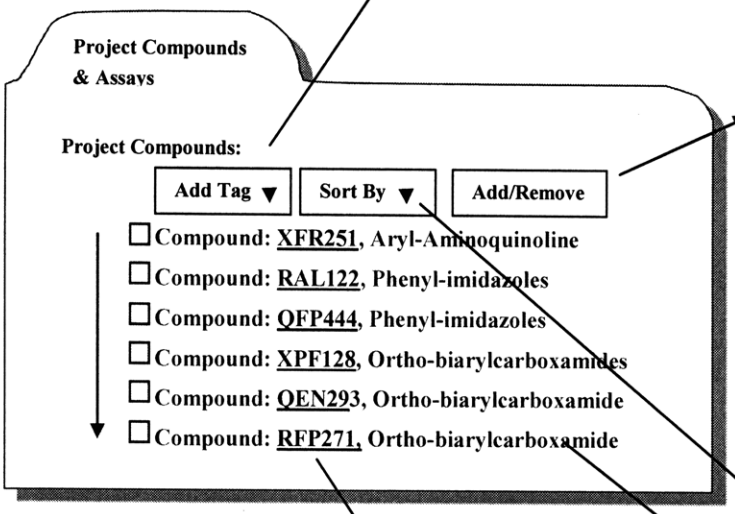


Figure 37: Project Compounds and Assays (Tab 2)
Note: All project names, compound codes, targets, etc. are arbitrary and do not represent actual ongoing work at Novartis.

Sorting – The list of compounds could be sorted by any number of different tags.

Additionally, the tags could be displayed next to the compound list. In this specific example tags associated with different chemical series (Aryl-Aminoquinoline, Ortho-biarylcarboxamide, etc) are displayed.

Permissions - Permissions for adding/removing compounds could be restricted to team members only if teams had security concerns.

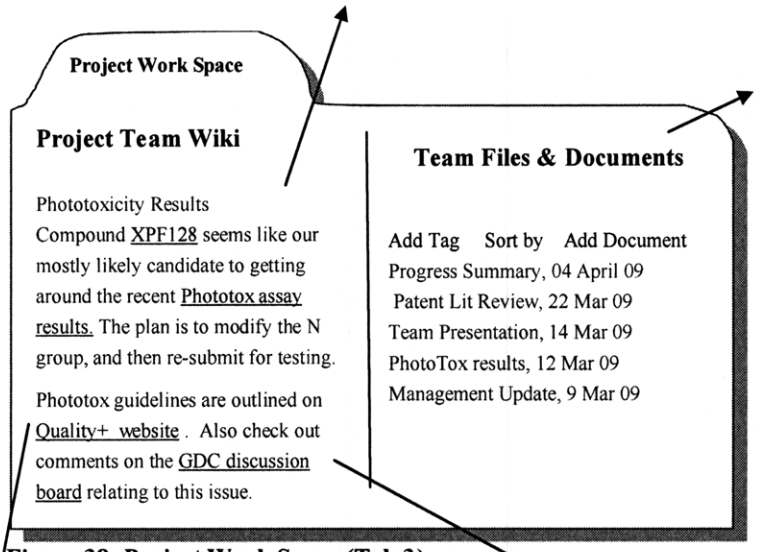
Compound Homepage Link – Each chemical compound on the project list would link directly to its Compound Homepage enabling the scientist to instantly find a wide variety of information associated with the compound.

[95] Concept envisioned by Deborah Igo, The Novartis Institute for BioMedical Research, NITAS

10.5 Project Work Space (Tab#3)

The project workspace would provide a platform based working environment for project team members.

Wiki Space – A project wiki would give teams a common platform for posting output related to tacit knowledge work. This environment should be unstructured allowing teams to post, collaborate, and capture team project knowledge as they see most relevant.



File & Document Storage – Teams currently store project team information on a shared team space in Lotus Notes. Migrating links or an interface to these documents to either the new SharePoint environment or a wiki would allow these documents to be viewable to the organizational at large.

Note: actual files should be stored independent of the wiki system in a corporate document management system. For example, Cisco uses SharePoint's environment for an interface to view documents but all documents are stored within Documentum and simply linked into SharePoint.

Figure 38: Project Work Space (Tab 3)
Note: All project names, compound codes, targets, etc. are arbitrary and do not represent actual ongoing work at Novartis.

Linking - A wiki would allow team members to help one another Navigate throughout the organization, establishing links to resources that that they find helpful for the team. Allowing the user community to build this linking network will be essential toward enabling "Google-like" intranet search capability.

Discussion Boards- The wiki space could either directly incorporate team discussion boards or link to specific discussion boards elsewhere in the organization.

Chapter 11: Compound Terminations

One of the benefits of organizing compounds in lists on the project workspace (Figure 37) would be the ability to systematically capture information regarding the removal/discontinuation of compounds from a specific project. Novartis currently captures information associated with project terminations in a project tracking database. Although this information provides insight into why the last remaining project compound was terminated it often does not provide visibility as to why compounds earlier in the project were discontinued.

11.1 User Research

According to the user interview research, an overwhelming majority of scientists felt that capturing information associated with a chemical compound termination was not only needed, but extremely valuable for the organization.

"Information on compound failures would be invaluable or our group. In our team meetings we talk about other compounds that we are interested in but often we have absolutely no context of if they've been worked on before, etc. Just knowing the historical context of where this compound was used in other projects, what were the issues associated with it would be powerful."

"It's criminal we don't have post-mortems of why compounds fail. We wind up making the same mistake 3, 4, or even 5 times. We're still reining in these in on hERG and phototox issues.

11.2 Process for Removing/Terminating Compounds

Part of the reason why compound terminations are not adequately captured by NIBR is that there is not a specific process in place to do so. Adding a compound removal function to the project compound list would establish such a process and tightly integrate it into the typical project workflow.

In order to remove a compound from the project list users could check the box to the left of the compound (Figure 37) and click the "add/remove" button. Depending on the specific development phase the project was in, requirements for documenting rationale for compound removal could be enforced before the compound could be removed from the project list. Table 12 below outlines a few possibilities. Additionally, when compounds were removed from the project team lists, associated termination rationale would be applied via tags to allow this metadata to be captured and later indentified throughout the entire organization.

Table 12: Requirements for Removing Compounds from Project Homepage List

Project Development Phase	Requirements for removing compound from project list	Information Captured
Early Stage Target Discovery & Assay Development (D0, D1)	None	N/A – Targets, not compounds are the focus of early development work. It is unlikely that specific compounds would be associated to particular projects at this point in time.
Hit Finding/High throughput Screening (D2)	None	N/A – High throughput screening assays often generate hundreds of thousands of primary hits. Mostly, likely it would not be worth capturing the compound history (in this particular format) at this early stage in the development process.
Lead Optimization (D3)	• Team members would fill in simple database form • Tag compound(s) with relevant discontinuation category	• Indicate if decision pertains to one specific chemical compound, or all compounds in the series. If the series is being discontinued all compounds within the series would be annotated with the same remarks/rationale. • Rationale associated with discontinuing compound/series • Links to relevant testing results, documentation, etc.
Candidate selection process (CSP) & Selected for proof of concept trial (sPOC)	• Discontinuation summary Report • Associated documentation & relevant files • Review by Program Office scientific committee • Tag compound(s) with relevant discontinuation category	• Primary reasons for compound discontinuation • Secondary reasons for compound discontinuation • Supporting documentation (Review board meeting minutes, PowerPoint presentations, etc.)

11.3 Discontinuation Summary Reports

As part of this thesis project, capturing information for compound discontinuations was briefly explored for compounds in the CSP and sPOC development stages. Thirty nine different compounds that had been terminated in the either the CSP or sPOC stage were analyzed. A small subset of these compounds were reviewed with a team of chemistry, toxicology, and safety profiling experts in order to identify and classify root cause reasons for termination. Findings were then compiled into discontinuation summary reports. A specific example of a compound discontinuation summary report is provided below. A similar report format is suggested as a requirement for removing compounds from the project homepage list that have been discontinued at the CSP or sPOC project stage.

Note: Compound identifications, disease indications, etc. have been changed to keep information confidential.

Discontinuation summary Report

Compound: PZR571

NCC code: PZR571A
Other Compounds in package: ZFY476, XJM861
DA: Oncology
Indication: Cancer

Color coding Key:
"P" Primary = serious problem, would discontinue development
"S" Secondary = Problem could potentially be solved, development path would be difficult. This reason alone would not necessarily dictate immediate termination but could be problematic in the future.

Reasons for Discontinuation

- **Toxicity "P"**
  - **Hepatotoxicity (liver cell necrosis) "P"**
- **Formulation "P"**
  - **Solubility "P"**
  - **Batch-to-batch variation "S"**
- **Active Metabolite "S"**

Summary Comments

Toxicity, hepatotoxicity

The compound was tested in 2 week rat toxicology studies and found to exhibit hepatotoxicity at all tested dose ranges. Studies were unable to establish a no observed adverse effect level (NOAEL).

"A 2-week oral gavagetox study in male rats was carried out. Minimal focal and/or single cell necrosis of hepatocytes, often accompanied by hæmorrhage and slightly increased inflammatory foci were seen in 2/5, 3/5 and 2/5 rats at 30, 75 and 150 mg/kg, respectively.Thus, a **NOAEL was not established**." – DADB Meeting Minutes 21 Sept 08

Formulation

Compound had solubility issues in formulation and exhibited batch-to-batch variation.

"PRZ571 Issues

- Low solubility and need for special formulations (microemulsion)
- Different properties of different compound batches (solubility, PK, etc.)
- Very low solubility in standard microemulsions (1% currently) limit drug load to 10 mg for a large soft gelatine capsule (Capsugel No 22 oblong)
- Effective human dose needs to be 20 mg (2 capsules) or lower"

" ~ DADB Meeting Minutes 21 Sept 08

"The solubility in MEPCs was rather low and insufficient for studies beyond phase I." – Leo Wild* email 08 Sept 08

Summary Report (Continued)

Active Metabolite

PZR571 has two active metabolites. One of those, ZFY476, was investigated as the lead compound for development but was later discontinued because it could not be formulated. As compound efficacy would have been determined on both the dosing of PZR571, as well as the presence of active metabolite, dose determination would be very difficult.

"Drug substance has two active metabolites" DADB Meeting Minutes 21 Sept 08

"In addition development would have been very complex due to the fact that ZFY476 is a active metabolite of PZR571 (in vitro even more potent than the parent!). The metabolism to ZFY476 (sulfone) proceeds via MQR223 (sulfoxide) - another active metabolite found in vivo!" ~Leo Wild* email 08 Sept 08

Additional Information Links

DADB Meeting Minutes 11 Sept 07

Excerpt from FIP meeting 21 Sept 08

Leo Wild* email 08 Sept 08

Project contacts*

Project Team Leader: Grey Nyberg

Chemistry Lead: Leo Wild

Other team members: Carin Gray, Michael Stoner, Gabriel Schmidt

Discontinuation Review Committee

Scientific Experts: Bill Egerton, Urban Laszlod

Figure 39: Compound Discontinuation Summary Report Example

*Note: Compound identifications, disease indications, team names, etc. have been changed to keep information confidential. Names are fictitious and are not those of actual Novartis employees.

Primary and Secondary Discontinuation categories

Analysis of compound terminations revealed that the majority of compounds do not necessarily have one specific reason or rationale associated with the compound termination. Often the story is more complex, with a variety of potentially problematic issues. As it is important capture all rationales for organizational learning, termination reasons were classified into either primary or secondary categories. Primary reasons, designated in the report by "P", involved serious problems that on their own would be justification for eliminating a compound from a project. In the example above (Figure 39), primary discontinuations included hepatotoxicity and formulation solubility issues. Secondary rationales, designated in the report by "S", are problems that could potentially be solved but where the development path would be difficult requiring additional resources, development time, etc. The secondary reason alone would not necessarily dictate termination. However, a compound with multiple secondary issues might be too difficult to move forward.

11.4 Standardized Classification of Termination Categories

In order to effectively tag and analyze compound terminations standard nomenclature, as well as user defined categories, should be associated with each compound. It will be important that the defined terminology matches that in the project management databases so that the organization can have common naming conventions for both project and compound terminations. Termination categories have been defined for current and future project management databases, PICture and Porto. A scientific review committee should be brought together to analyze these termination categories and apply specific case studies to make sure that the classifications are complete and accurate. Additionally, project team members need to be given the flexibility to apply user generated tagging related to compound terminations. Even with a carefully defined upfront categorization scheme it will be impossible to envision and capture every situation. With relatively cheap digital storage the more tagging information that can be associated with a particular compound, the better.

11.5 Termination Analysis Committee

The author's efforts tracking down information associated with compound failures revealed that compound terminations tended to very political. While a large majority of Novartis employees were very open to talking about their experiences with certain compounds, others were highly guarded and hesitant to share information. Several internal sources additionally re-iterated that the rationale for a compound termination was often very dependent on the information source and that it was therefore necessary to speak with a large number of employees in order to capture the complete story. For the reasons highlighted above, it would be prudent to establish a scientific committee to review compound/project terminations for CSP, sPOC molecules and author the

discontinuation summary report. Elements of committee structure and leadership are highlighted below.

Committee Structure:

- **Program Office Leadership**- The NIBR Program Office functions as an overall umbrella organization and provides a number of support services to both the NIBR disease areas and platform groups including portfolio analysis. A committee structured to analyze compound terminations and compile overall lessons learned would structurally fit well under this umbrella.
- **Scientific experts** – It will be paramount to include scientific experts on the committee. Not only will such individuals establish the credibility and legitimacy of a committee within the larger scientific organization, scientific experts are also needed to evaluate the complex issues associated with termination decisions. Program Office members cannot effectively do this alone.
- **Dedicated position** -Committee membership could be a full time position that would rotate every six months to a year giving various scientists throughout NIBR a valuable learning opportunity for analyzing the portfolio at large.

Committee Benefits

- **Unbiased Analysis**– Decoupling compound termination analysis from the project team would help to establish termination root causes without the team/ organizational biases.
- **Organizational Learning** – The committee could serve as a hub for organizational learning allowing top management to capture lessons learned and ensure appropriate business processes were in place to identify potential issues earlier in the development process. For example, NIBR currently has a series of "Quality Plus" teams that are instrumental in analyzing specific issues like cardiac arrhythmia (hERG). The committee would simply be a higher level extension of the Quality Plus program.

Additional Actions by Management

The concept of "fail fast" is often prevalent within the Pharmaceutical Industry. The idea is that a company should quickly try to eliminate compounds that are not successful and fail them as early in the development cycle as possible. This avoids spending time and resources on something that will ultimately never become a viable drug. This approach often leads companies to focus on quickly moving onto the next project and overlooks valuable knowledge that can be obtained by learning lessons from project failures, compound histories, etc.

"Most people more interested in working on next project rather than documenting on why a project failed or documenting terminated project. Very rarely are project post-mortems put together"

Additionally, individuals fear failure and the implications for their personal career:

"Individuals have specifically told me that they don't want to be associated with a project failure."

NIBR leadership needs to institute a culture where it is not only acceptable to fail, but almost expected. Drug discovery is immensely complicated with an inherent high failure rate: For every 5,000 – 10,000 compounds tested, 5 of those will reach clinical trials and only one will become an FDA approved drug (Figure 7). Furthermore, the pace associated with drug discovery process often forces scientists to make decisions based on the information that they have available at the current time. Culturally, this should be recognized, understood and accepted.

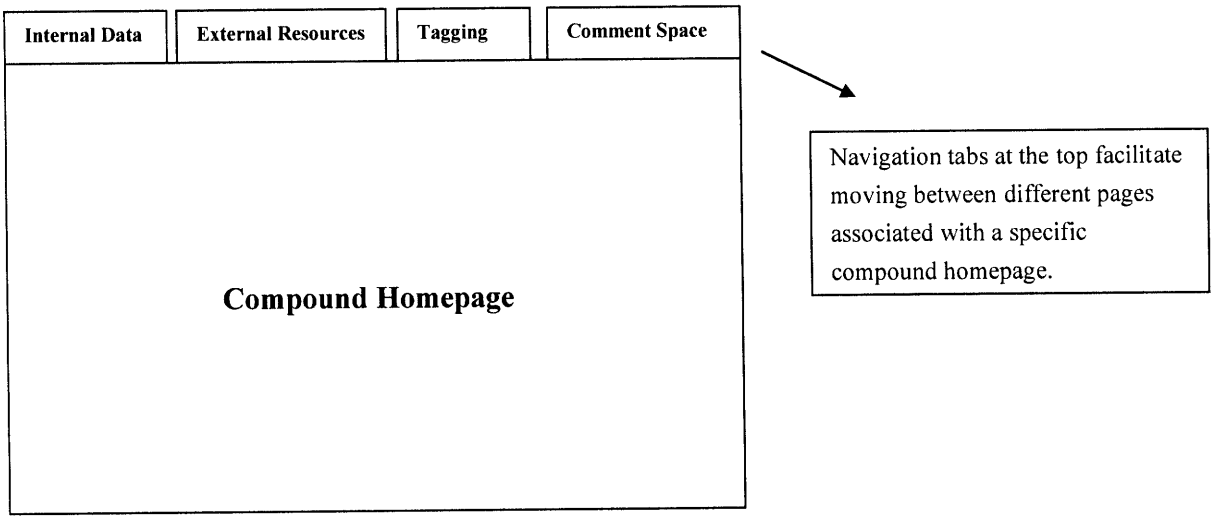
"Reasons for termination are incredibly sensitive information because of both science and politics. Many decisions have been made with incomplete data and nobody wants to be second guessed. You might have made a 1st guess on minimal data."

Instead of having employees run from failure because it is a "bad thing" the emphasis needs to be placed on organizational learning from that failure - "What have we learned? How could we have caught this earlier in the development process? What can we do next time to test for this earlier?" Additionally, management could emphasize capturing "compound experiences" rather than "failures." Compounds are often dropped by certain projects only to be picked up later for use in other indications, delivery mechanisms, etc. Very often discontinuations are not true "failures."

Chapter 12: Compound Homepage

Similar to the Project Homepage structure, each individual compound within the Novartis library would have a designated "Compound Homepage". The homepage would serve as a user editable knowledge management platform for individual compounds.

The Compound Homepage would be structured as a series of associated team web pages with navigation tabs at the top of each to facilitate moving between pages.



12.1 Internal Data Page (Tab#1)

This would be the 1st page /tab displayed when visitors navigate to the Compound Homepage site. The internal data tab would either directly display or link to organizational data associated with a particular compound. The overall goal would be to unify diverse sources of information in once easy to find place and allow user to find information with the "fewest number of clicks possible."[96] The internal data page is a non-user editable portal and simply displays or links to information from other data sources.

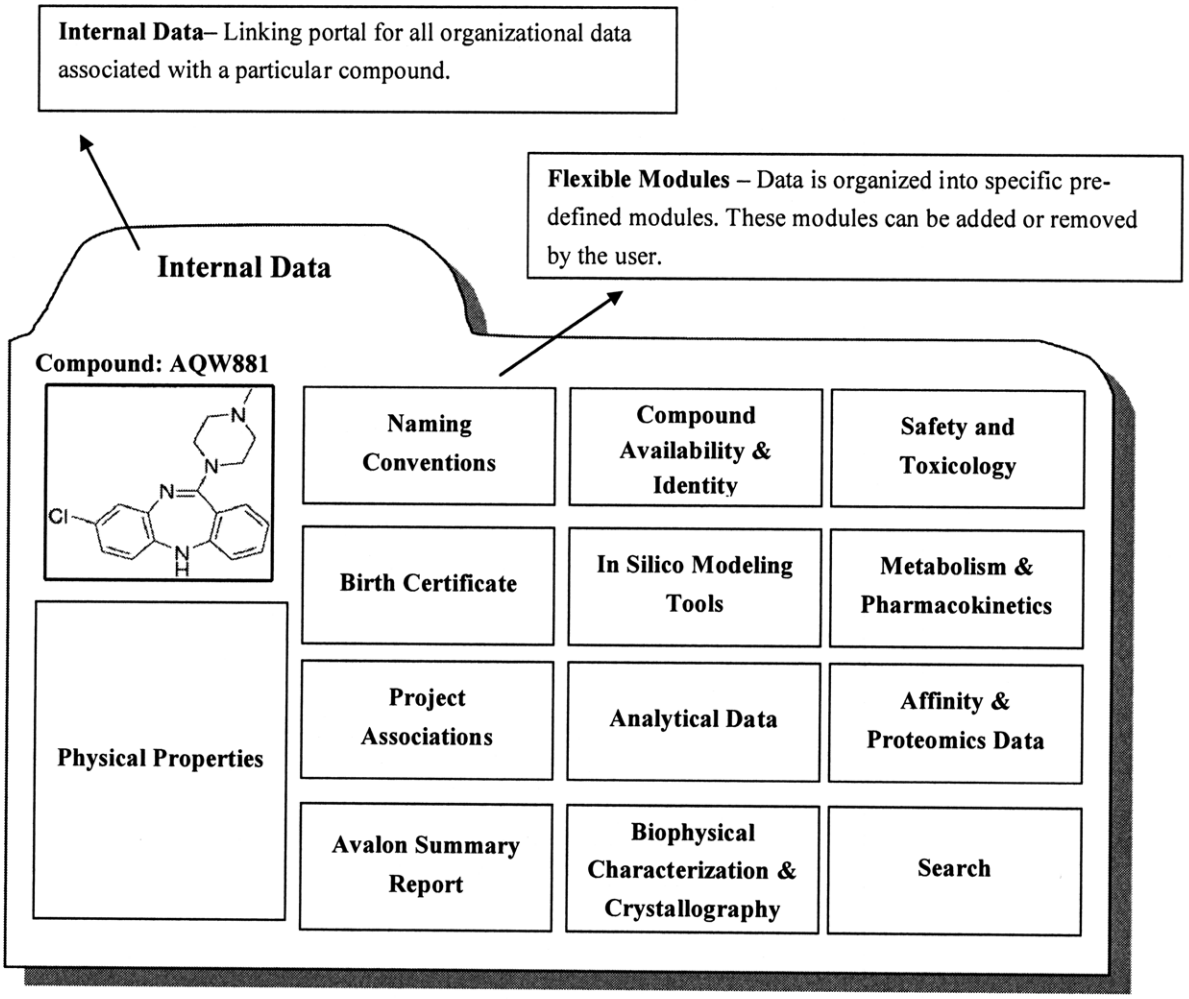


Figure 40: Novartis Internal Data (Tab #1)
Note: All project names, compound codes, targets, etc. are arbitrary and do not represent actual ongoing work at Novartis.

[96] Dejan Bojanic, The Novartis Institute for BioMedical Research, Center for Proteomic Chemistry

Modularity

Throughout the user research interviews scientists expressed a wide range of preferences for the types of data displayed on the Compound Homepage. Preferences not only varied by organization and workflow activity but also by personal preference. Based on this feedback, data was broken down into specific modules. These modules could be either added or removed by the end users similar to a Yahoo or iGoogle homepage (Figure 41).

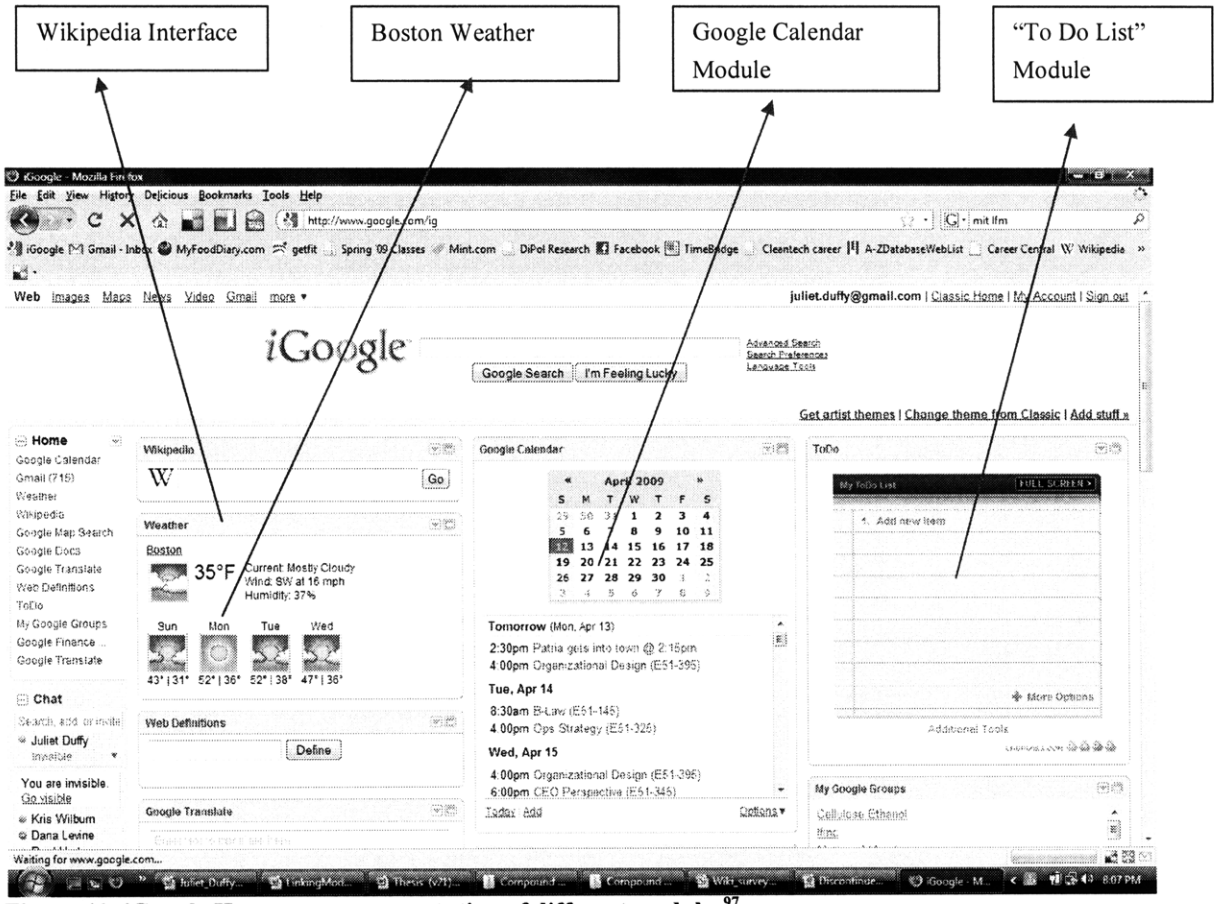


Figure 41: iGoogle Homepage, representation of different modules[97]

[97] iGoogle site. *Google,* n.d. Web Accessed 20 April 09 http://www.google.com/ig

Data Warehouse

Rather than connecting the various modules directly into existing databases it will be essential to create a backend data warehouse (Figure 42). Creating a data warehouse would translate information into a standardized format and have multiple advantages including: scalability, loading speed, maintenance and backend access for various computational science groups. Data warehousing would allow new modules to be added with various ease.

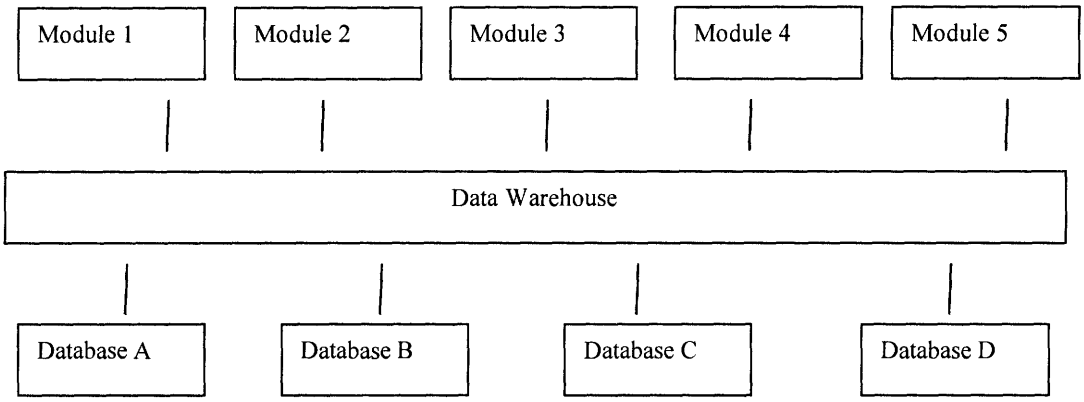
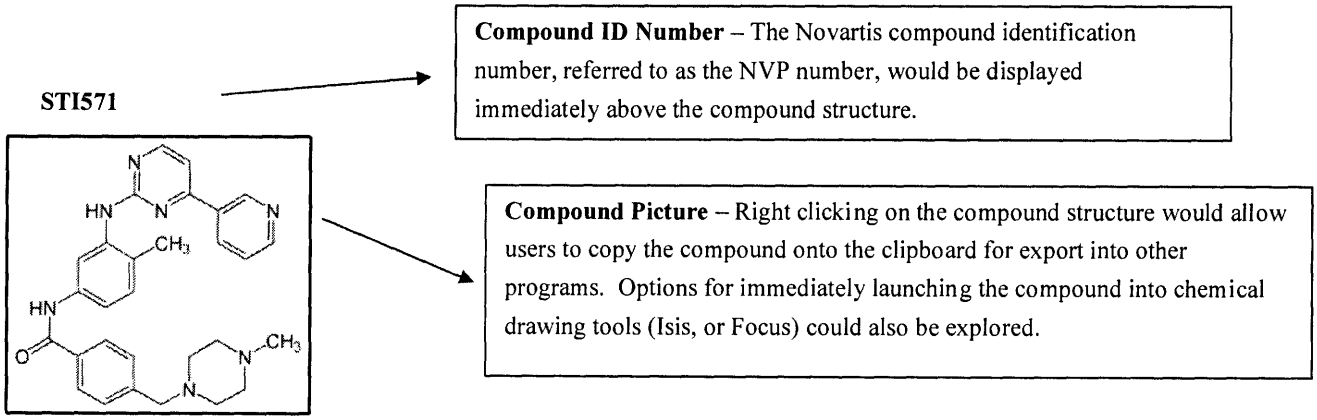


Figure 42: Data Warehouse

12.1a Compound Picture Module

Scientists communicate chemical structures in pictures. The first module on the compound homepage would not be user editable and would automatically display the chemical structure in the upper left hand corner of the Compound Homepage.



Note: All project names, compound codes, targets, etc. are arbitrary and do notrepresent actual ongoing work at Novartis.

12.1b Naming Conventions Module

A wide variety of naming conventions exist for chemical structures. Examples include scientific identifiers (IUPAC, InChI), digital representations (InChIKey, SMILES) chemical society/Pharmacopoeia names (PhEur, USP), research codes, trade names, marketing brand names, etc. Additionally, a number of commonly used scientific databases assign specific identifiers for each chemical compound. For example, the Chemical Abstracts Service (CAS) is a comprehensive database that searches scientific and patent literature. CAS assigns a unique chemical identifier number, a CAS Number, for each chemical. Providing scientists with access to chemical synonyms and chemical identifiers facilitates both internal and external search.

Synonyms (internal) – Many compounds have multiple internal codes/names. Older compounds often have a Ciba-Geigy or Sandoz # associated with them. Chemicals further along in development will have a development or clinical code #. User research indicated a high preference for being able to view both internal and common external synonyms on the front webpage (vs. buried deeper in links). As a result these are displayed up-front.

Naming Conventions Module

Synonyms (internal)

Ciba-Geigy #	CGP57148
Development Code	CGP57148B

Common Synonyms (external)

CAS Number	152459-95-5, 220127-57-1
ATC Code	L01XE01
PubChem	5291
DrugBank	APRD01028
ChemSpider	5101

Links to Other Names
View all synonyms
IUPAC, InChI, InChIKey, SMILES
UltraLink Preferred Names

View all synonyms – All synonyms are not reflected on the front page of the module due to space limitations. This link would open up another pop-up window that displayed all the synonyms in one place.

IUPAC, InChI, InChiKey, SMILES – Scientists expressed interest in being able to view the IUPAC, InChI, InchiKey and SMILES name but felt that listing the complete name directly within the Naming Conventions module will take up too much space. Therefore, users are given the option to link to these names. Opening the "IUPAC, InChI, InChIKey, SMILES" link would display the following table in a pop-up window:

IUPAC	4-[(4-methylpiperazin-1-yl)methyl]-N-[4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl]benzamide
InChI ?	1S/C29H31N7O/c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-16-14-35(2)15-17-36/h3-13,18-19H,14-17,20H2,1-2H3,(H,32,37)(H,31,33,34)
InChIKey	KTUFNOKKBVMGRW-UHFFFAOYSA-N

? **Question Mark (mouse over)** - A question mark mouse over symbol that displays "click for info" upon mouse over could be located next to various name items. Clicking would produce a pop-up window explanation for the object. For example:

"The IUPAC International Chemical Identifier (InChI, pronounced "INchee") is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web." ~ InChI Wikipedia Article

UltraLink Preferred Name –The Novartis IT group has assembled a library of compound nomenclatures as part of the UltraLink project. UltraLink not only captures chemical names but assigns a naming priority hierarchical structure. The synonym browser is currently available and will be easy to link to the naming conventions module. A possible link to the UltraLink preferred names, and its utility to scientists should be further explored:

UltraLink Priority StructureExample: Zoledronic Acid

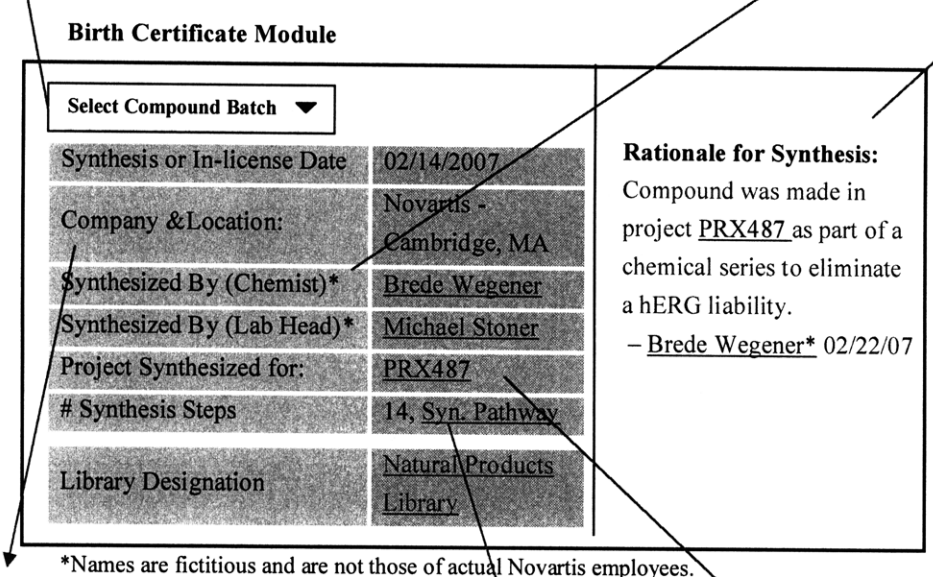
1. INN (International Non-Proprietary Name)	zoledronic acid
2. BAN (British Approved Name)	zoledronic acid
3. USAN (US Approved Name)	zoledronate disodium; zoledronic acid
4. INNM (Modified International Non-Proprietary Name)	zoledronate disodium
5. BANM (Modified British Approved Name)	zoledronate disodium
6. BP (British Pharmacopoeia)	Not Available
7. PhEur (European Pharmacopoeia)	Not Available
8. USNF (US National Formulary)	Zoledronate disodium; zoledronic acid
9. USP (US Pharmacopoeia)	Not Available
10. BPVet (British Pharmacopoeia - veterinary)	Not Available
11. Name	CGP 42446; ZOLEDRONIC ACID; Zoledronic acid; zoledronic acid
12. Synonym	Aclasta; Reclast; ZOL 446; ZOL-446; Zoledronate; Zomera; Zometa; anhydrous zoledronic acid; zoledronate; zoledronate TTS; zoledronic acid monohydrate
13. Research Code	CGP-42446; CGP-42446A; CGP042446; ZOL-446

12.1c Birth Certificate Module

A wide variety of information is related to the registration or "birth" of a compound. Scientists expressed a strong desire to have access to names of individuals that synthesized the compound, location of synthesis, synthesis rationale and other information relating to the compound registration.

Select Compound Batch – Often different batches of a particular compound are synthesized. A drop down menu would allow a user to select the compound batch that they wish to see. By default the first synthesis batch would be displayed.

Chemist, Lab Head – The Chemist and Lab Head names are currently captured as part of compound registration process. Providing these names hyperlinked to the Novartis directory will allow scientists to directly follow-up and ask questions with the appropriate colleague.



*Names are fictitious and are not those of actual Novartis employees.

Rationale for synthesis– Synthesis rationale was one of the most requested items during the user research. A mandatory data entry field could be added to capture this information during compound registration. That data field could then be displayed in the birth certificate module. Alternatively, the section could be a user editable wiki where scientists could annotate compounds that they have made in the past. A default message such as "Rationale is missing. Please help annotate me!" might encourage user

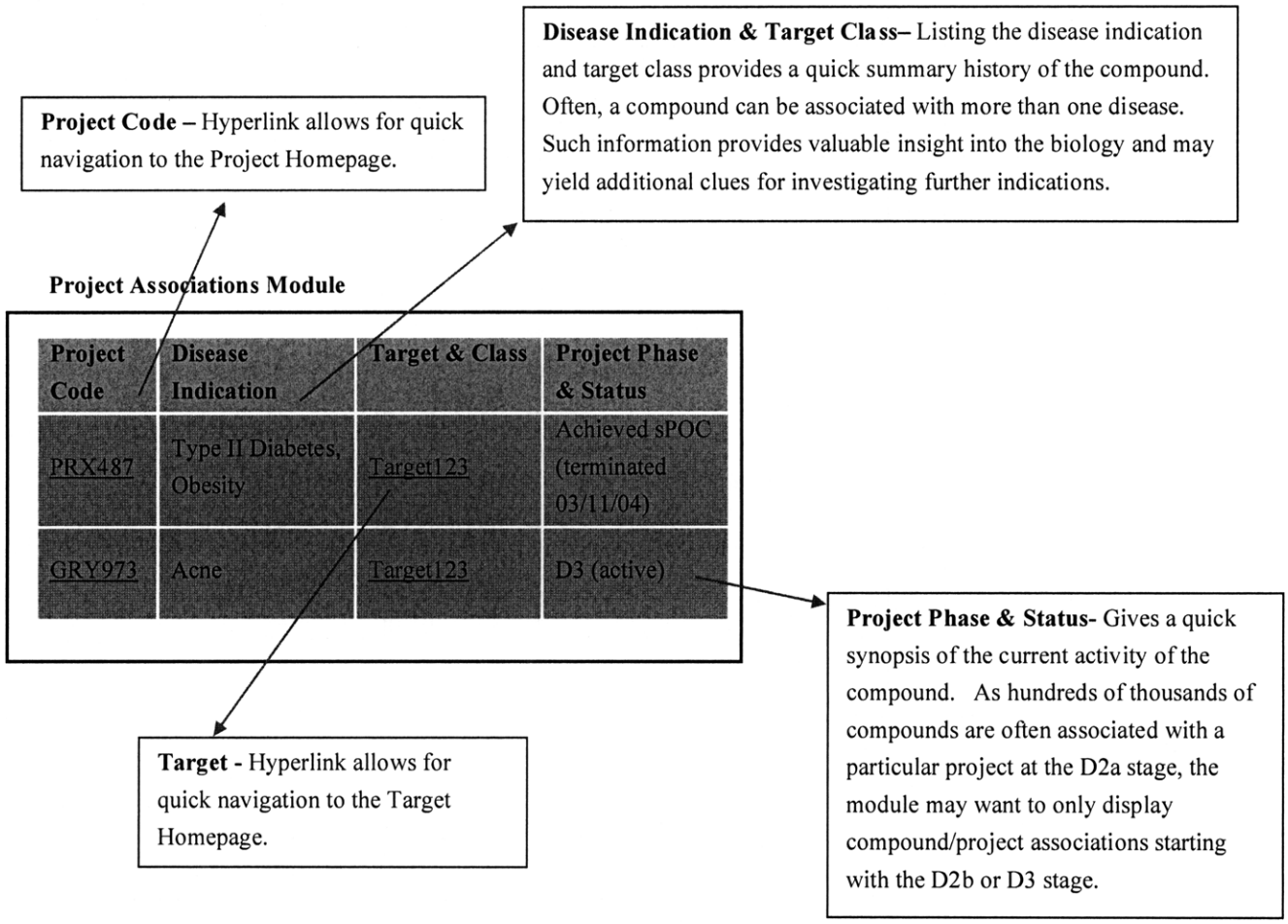
Company & Location– Chemical compounds are synthesized at a variety of research locations within Novartis. Additionally, many compounds are either in-licensed from other companies or purchased as part of a collection from an external chemical supplier. Providing clues about a compound's origins is often very helpful for tracking the compound history and understanding previous scientific work.

Project Synthesized for: – Individuals that were interviewed for user research routinely stated their desire to know what project the compound was synthesized for. This information needs to be captured as part of the compound registration process. Displaying it in the birth certificate module will provide a linkage to the Project Homepage where additional information about the project is available.

Synthesis Pathway: – Chemists that routinely synthesize chemical compounds, as well as Analytical scientists who perform NMR, specified that it would be helpful for them to be able to view the synthesis pathway. Such information also has the potential to save a chemist a significant amount of time & effort. Before beginning a synthesis they could simply perform a structural similarity search and view how a similar synthesis was performed. It's possible that if the compound was synthesized recently a fellow chemist might have some chemical material left over from the prior synthesis providing a jump start. Additionally, synthesis information is valuable when a compound near s development and will have to be produced on a large scale.

12.1d Project Associations Module

A chemical compound can be associated with many different projects. It's useful for scientists to be able to quickly investigate the compound's role in a particular project and potentially follow-up with project team members. A simple table that associates compounds with projects would help facilitate this task.

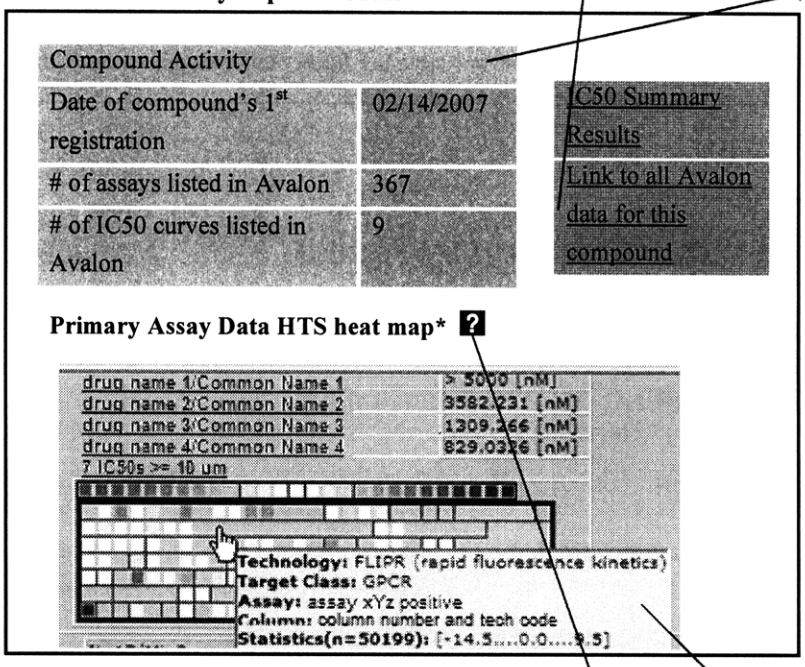


12.1e Avalon Summary Report Module

Avalon is the Novartis data warehouse that provides assay results and additional information on chemical compounds. It will be essential to provide both Avalon summary data as well as a link into Avalon so that users can easily extract additional information from Avalon, access testing results, etc.

Link to Avalon – Hyperlink would allow for quick navigation into Avalon. The hyperlink would pass the compound identification directly to Avalon so that a user would not be required to re-sign in and re-enter the NVP# to execute the search.

Avalon Summary Report Module



Compound Activity– Users expressed their desire to be able to quickly see high level Avalon statistics. The # of assays listed in Avalon would indicate the amount of information listed in Avalon for a particular compound. As one interviewee animatedly expressed "I want to know how much info is in Avalon and if it's worth it to open Avalon to look at this info."

The # of IC50 curves listed gives an indication of the compound's level of promiscuity. For example, hundreds or thousands of IC50's curves would indicate that the compound was highly promiscuous or a frequent hitter.

Both # of total assays as well as # of IC50 results would need to be referenced in the context of the compound registration date. A compound that was added to the library in the past few months would have very few assays & IC50 curves. This would not be an indication of an inactive compound but rather a function of how long it had been in the library.

* Heat map & IC50 Summary Results were developed by at Novartis by Bernd Rohde.

? Question Mark (mouse over) - A question mark mouse over symbol that displays "click for info" would be displayed next to the heat map. Clicking the symbol would produce a pop-up window explanation for the object. In this case the following pop-up window would appear:

What is this heat map?
The primary heat map represents a collection of CPC primary screening data put together on 01/2008. It is a one-time snap shot of primary screening data and is **NOT** necessarily inclusive of all primary assays that have been run by CPC.

How do I interpret this?
Each box on the grid represents a particular assay. The assays have been grouped first by technology then by target in order to give a visual representation of primary screening results.
The coloring represents the amount of activity (% inhibition) measured in the assay.
Grey = Assay was not run
White = No activity (0% inhibition)
Red = The darker the red coloring the greater the % inhibition (+)
Blue = The darker the red coloring the greater the % inhibition (-)

Assay grid (mouse over) - The heat map allows users to mouse over individual squares. Upon mouse-over assay statistics are displayed.

12.2 External Resources Page (Tab#2)

The 2nd tab displayed on the Compound Homepage would provide a linking portal to external data associated with a particular compound. As discussion in the User Research chapter, propriety concerns with passing compounds through the internal firewall would need to be addressed and data resources would most likely need to be located within the Novartis firewall. Table 6, within the User Research Chapter highlights the external databases that interviewees indicated they used most often.

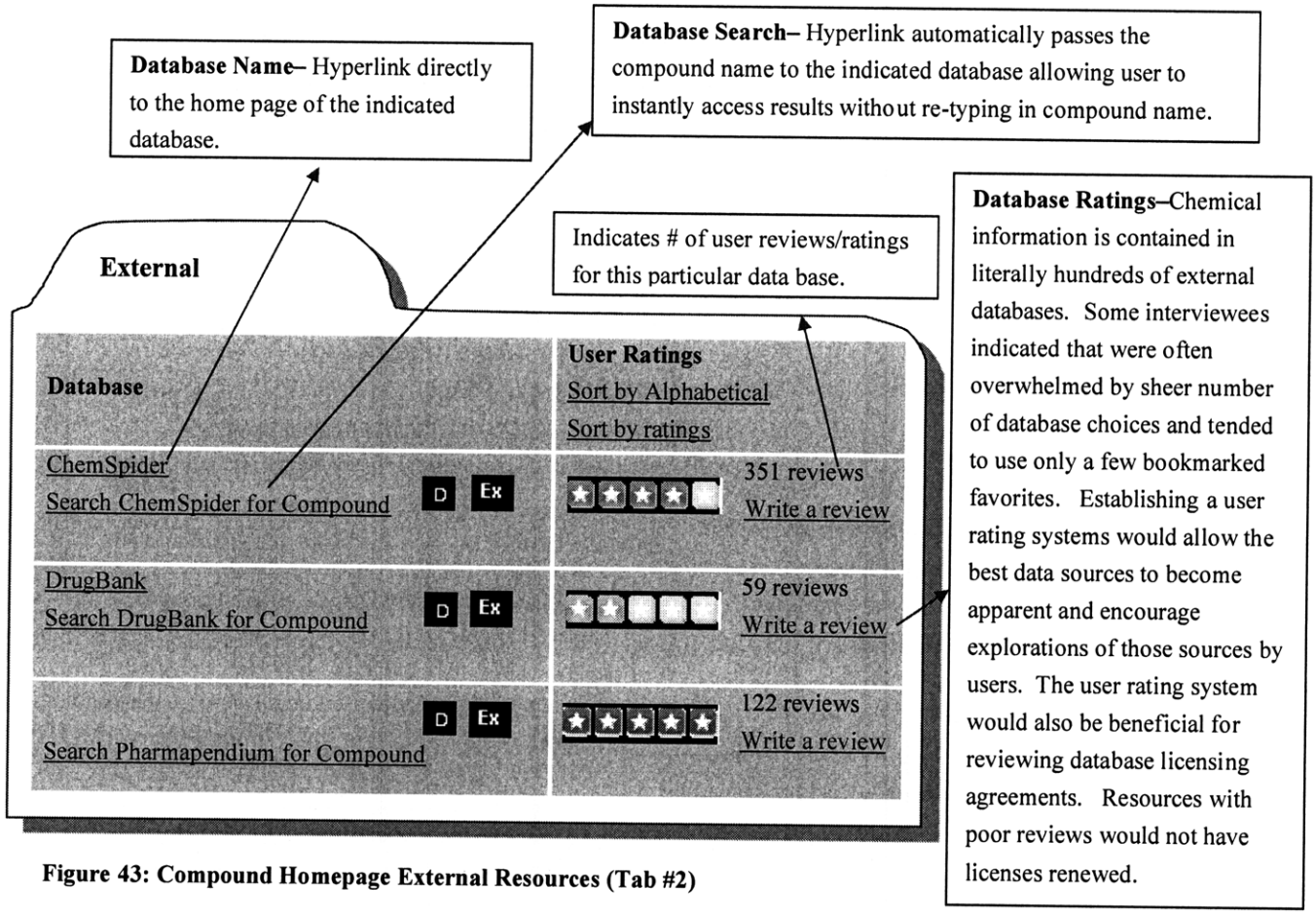


Figure 43: Compound Homepage External Resources (Tab #2)

D **Letter "D"(mouse over)** - Mousing over the symbol displays "click for database description". Clicking the symbol would produce a pop-up window that gives a brief overview of the database. For example, clicking the "D" symbol next to ChemSpider would display the following:

Database Description*–ChemSpider is a chemistry search engine open free to charge to the public. Chemical structures are indexed into a single searchable repository (*taken from ChemSpider)

Ex **Letter "Ex"(mouse over)** - Mousing over the symbol displays "click for examples of information found in this database". Clicking the symbol would produce a pop-up window that provided specific example of information that could be retrieved for that particular database.

Examples of Info find in ChemSpider:
Additional external data sources, commercial suppliers, naming conventions, prediction of physicochemical properties based on ACD/Lab algorithms.

12.3 Grouping/Tagging Page (Tab #3)

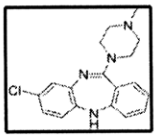
The 3rd tab displayed on the Compound Homepage would allow individuals to tag individual or groups of compounds, explore tagging relationships, and annotate tagging groupings via a wiki.

Compound Picture– As the "Tabbing/Grouping" tab encourages browsing to other compound homepages it will be important to orient the user and remind him/her which Compound Homepage they are viewing. As individuals don't typically remember NVP# structural associations, a picture of the structure will be imperative. Having the picture displayed on this page will also help avoid tagging mistakes and insure that users are tagging the correct compound.

Tagging / Grouping

Tagging– Users could tag a single compound or apply a tag to a group of compounds.

AQR851



Add a user tag to NVP-AQR851

Apply a user tag to multiple compounds

The number of group members is listed to right of each tag. Clicking on the members brings up a NPV & structure picture list of all compounds in that group.

A scroll bar allows users to scroll though the list of tags.

Pre-defined Tags

Suggest a pre-defined Tag

- Project AQW365 *(545 members)*
- Project QWR883 *(48 members)*
- Type II Diabetes *(17 members)*
- Stearoyl-CoA -validated target hit (*23 members)*
- Target Family GCPR, validated hit*(1285 members)*
- Natural Product Library *(10,233 members)*
- Synthesized by Brede Wegener *(459 members)*
- Structural Similarity *(100 members)*
- CSP (terminated) *(42 members)*
- Terminated, Phototoxicity "p"*(17 members)*

User Applied Tags

Add a user tag to NVP-AQR851

- Greg Stone Climax List *(1,345 members)*
- Ciba-Geigy Terminated List (2000-1977) *(781 members)*
- Failure to co-crystallize with protein *(9 members)*
- Addition increases transcription of Gene ABC *(13 members)*
- Unstable compound powder *(11 members)*
- Education office case studies of phototoxicity *(7 members)*

Tags that are underlined hyperlink to an associated "tag" wiki. This wiki allows for the creation of additional tag related metadata and provides a user forum for discussing the group of tagged compounds. Potentially, users could also rate the usefulness of the grouping/particular tag. An example Tag Wiki is below:

Figure 44: Compound Homepage Tagging/Grouping

Pre-Defined Tags – These tags would be tags that have already been defined by the organization, or metadata that is captured systematically (see discussion below). Pre-defined tags are **color coded green** in order to allow the user to quickly distinguish them from user created tags.

User Defined Tags – User defined tags can be created at any time by any member of the NIBR community (see discussion below). User-defined tags are **color codedmaroon** in order to allow the user to quickly distinguish them from pre-defined tags.

Tagging Group Wiki – Wiki space will be open to user community to encourage discussion and further annotation of the tagged group. Throughout the user research interviews scientists expressed a strong desire to comment on groups of compounds. This wiki provides this.

Tagging Group Created by: Carol Mahol

Date Created: 11/24/2008

Initial Group Members: WRT421, YIR554, ZRI271, AQR851, PMQ395, JKQ539, RJS223

Rationale for Group Creation: I created this group as a result of digging up some examples of compounds that had phototox issues for an upcoming education office class on 12/08. Additional info on these case studies is posted here

---

Wiki Section Guidelines for posting to this wiki Example Posts

Successful/Not-successful case studies

Several of the compounds in this group were successfully modified to eliminate the phototox issues (YIR554,RJS223) while others resulted in termination (AQR851) ~ Carol Mahol 11/24/08

Structural Similarity- All of these compounds share a common structural elements that allows for the breakdown into a reactive metabolite XYR..... ~ Bill Egerton 11/28/08

12.3a User Applied Tags

User defined tags could be created at any time by any member of the NIBR community. As discussed in previous chapters, allowing the user community to add metadata and categorize content via folksonomy has several advantages: the work of categorization is spread out among the entire community, the categorization scheme is flexible, time delay is minimal (e.g. you don't have to wait for information to be added to a database by IT data organizers), and tags reflect the information structure and relationships that people actually use[98]. Additionally, user created tagging will allow individuals to share relationships and knowledge about specific chemical compounds throughout the entire research organization in real time, irrespective of organizational boundaries. Several hypothetical examples of user created tags are highlighted below:

Education Office Case Studies of Phototoxicity Case Study

The Education Office at NIBR develops a number of scientific classes in order to train employees on a wide variety of topics. In a hypothetical example a member of this organization named Carol is creating an education course specifically to address phototoxicity. Carol has spent hours compiling many detailed case studies of compounds that were phototoxic. Sometimes the phototoxicity issue was resolved by the project teams, other times the compound was terminated. Carol decides to tag her associated case study group of compounds. She navigates to the NIBR Compound Homepage of AQR851, clicks "apply a user tag to multiple compounds", enters in a list of compounds in addition at AQR851, and creates a tag category called "Education Office Case Studies of Phototoxicity".

Not only does Carol choose to tag the group of compounds, but she also fills out a tagging wiki (illustrated above). The tagging wiki captures additional metadata associated with the tag including: the name of tagger, date tag group was created, rationale behind adding the tag, etc. It also provides a forum for additional discussion of the tagged group. Carol uses the wiki to further annotate her group of compounds: "Several of the compounds in this group were successfully modified to eliminate the phototox issues (YIR554, RJS223) while others resulted in termination (AQR851)".

Bill, a member of the Global Discovery Chemistry Group (GDC), stumbles upon Carol's grouping while looking at the Compound Homepage of ZRI271 for an unrelated problem. Bill is one of NIBR's experts on phototoxicity so Carol's tagged grouping immediately interests him. He clicks on the *"7 members"* hyperlink to the right of Carol's tag on the ZRI271 Compound Homepage.

---

[98] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

From there is he able to view all of the compounds that Carol has tagged. Bill compares structural similarities of the compounds in Carol's group and comments on the tag wiki regarding his observations.

Furthermore, Bill doesn't know Carol but now has instant access to her contact information via the "People Homepage" by clicking on the hyperlink "Carol". Bill picks up his phone and calls Carol to discuss a few of the phototoxicity case studies in more detail. During the course of their conversation Bill also suggests two other compounds that would be very valuable for Carol to add in her class next month. Carol then manages to persuade Bill to stop by her scheduled class in December as a guest speaker. Two individuals who previously didn't know one another are now connected and actively dialoging to help address a drug discovery challenge. Both Bill and Carol are tremendously benefiting from the interaction. The entire time required for Carol to create the initial group of tags? Just 7 minutes.

Unstable Compound Powder Case Study

Jason Brucker, a very meticulousness organic chemist in the NIBR GDC organization keeps a very detailed record of his chemical inventory. Jason has worked in drug discovery synthesizing chemicals for a number of years. His experience has taught him the value of keeping certain chemicals in his personal inventory as they are often useful starting points for a synthesis, eliminating the need to perform a number of prior reaction steps. Jason stores these extra chemicals in his laboratory typically in powder form. As some chemicals are prone to degradation over time, Jason will often run a quick identity check on the mass spec if he is using a stock powder that he knows is older.

Over the years Jason has compiled a list of compounds that have proven from his personal experience to be unstable over time. One day Christina, one of Jason's colleagues asks Jason for this list. Christina is synthesizing something very similar to what Jason has produced in the past and wants to make sure that it's not on Jason's "unstable" list. Christina suggests that Jason should post this list to the Compound Homepage, as it might potentially benefit other chemists within the organization. Following Christina's advice, Jason navigates to the Compound Homepage and tags eleven different compounds with the user tag "unstable compound powder". He annotates the tag wiki as to why he created the grouping.

As Jason is very detailed oriented, he has saved old mass spec results, as well as a PowerPoint slide detailing a proposed mechanism of action of breakdown from some of his compounds that have degraded. Annotating the tag wiki inspires Jason to dig up these documents from his personal hard drive. As NIBR's policy is to not store files physically in the SharePoint wiki site, Jason places

these files in one of his public directories on Documentum, the company's new document storage system. From Documentum Jason obtains a hyperlink to the location of his documents. He embeds this hyperlink in the tagging wiki simply stating *"spec results and PowerPoint detailing proposed mechanism of degradation located* here*"*

Six months later, Steve, a project team leader is reviewing high throughput screening results for his project. Compound AQR851 appears as a very strong candidate with a large percentage of inhibition of the project's target. Before running further testing Steve decides it might be prudent to poke around on the Compound Homepage to see if he can learn anything about the history of AQR851. As Steve scrolls through the list of user generated tags he notices that compound AQR851 is tagged with *"unstable compound powder."*As this particular tag is a maroon colored user generated tag Steve initially treats this with a bit of skepticism and realizes that the information may be a bit open to interpretation. Still, this information is concerning. Steve clicks on the tag to open the tagging wiki. He reads Jason Brucker's comments and follows the link to Jason's PowerPoint slide which proposes a mechanism of degradation for AQR851. Jason has clearly documented that AQR851 degrades into compounds X & Y. Steve now wonders if his assay results were caused by AQR851, compound X or compound Y. He contacts the compound library management group to share his concerns and request mass spec identity testing of the AQR851 powder held in the compound archive. One week later testing results come back. Sure enough, compound AQR851 has degraded. In fact, the powder is mostly now compound X, not compound AQR851!

The organic chemists on Steve's project team quickly synthesize a fresh batch of compound X. When the team repeats the target specific assay testing with a purified solution of compound X they are able to verify that it was compound X, and not compound AQR851 that causes the large percentage inhibition of the target. Compound X is not even currently in the Novartis library! Compound X is submitted for library registration. Steve navigates to Compound X's homepage to add the "rationale for synthesis" and to annotate his team's experience with compound X. He links this to the comments on Jason's tagging wiki and the compound AQR851 homepage.

User Applied Tags and Tagging Wikis Summary

From just the two examples above the reader can see how user generated compound tags, coupled with the power of user annotated wikis could be immensely powerful for NIBR. User generated tags and wikis are particularly important as they provide a very flexible way for individuals to capture relationships and knowledge work as it is happening. As seen from above, often the tags are unpredictable and could not have been foreseen in an upfront based, top-down organizational scheme.

12.3b Pre-determined Tags

Individual vs. Group Questions

How do "pre-determined" tags differ from user generated tags, and how are these also useful for the organization? Pre-determined tags are nothing more than the metadata that already exists for a particular compound. In other words, reviewing "pre-determined" tags is just another way of simply viewing data that exists for a particular compound.

Within the Internal data section tab of Compound Homepage, different linking modules were explored. The modules provide links to or display information relating to one particular compound. For example, suppose that a scientist wanted to locate the chemist responsible for synthesizing compound AQR851. He/she would navigate to the AQR851 Compound Homepage, click on the Internal Data tab, scroll down to the "birth certificate" module and then look at data field titled "synthesized by chemist." The name *"Brede Wegener"* appears so the scientist knows that Brede Wegener synthesized compound AQR851.

This is great if the user just wants to know about compound AQR851. However, what if the user wanted to explore that same question in a different way? What if the scientist kept randomly encountering compounds synthesized by Brede Wegener that just happened to be of particular interest. The scientist might want to view all of the chemical compounds that had ever been synthesized by Brede Wegener. Navigating to randomly selected chemical compound homepages and searching for "Brede Wegener" as the synthesizer would not only be tedious, in a collection of millions of compounds it would be impossible. So how would an individual accomplish this task? By simply organizing the information in a different way, via "pre-determined" tagging, the task is remarkably easy. A simple algorithm search of an already existing Novartis database would identify all of the chemical compounds that had been synthesized by Brede Wegener. These could be tagged with *"synthesized by Brede Wegener"*. Once tagged it would be easy to view these as a group of compounds. The scientist would simply navigate to the "tagging /grouping tab," scroll through the list of pre-determined tags to locate the "synthesized by Brede Wegener tag." Immediately they would see that this group has 459 members meaning that Brede has synthesized 459 other chemicals (Figure 44). Clicking on the "459 members" hyperlink would display a list of NVP#'s and associated pictures of the structures of all 459 group members. Navigation and further exploration of these compounds would be easy as each NVP# would be hyperlinked back to that particular compound's homepage.

Questions related to an individual compound are compared with situations / questions that might be asked of a group of compounds (Table 13). The Internal Data tab or linking functionality of the Compound Homepage would usually be best suited to answer individual compound questions while the Tagging/Grouping tab would allow users to explore questions relating to an entire group of compounds.

Table 13: Examples of Individual vs. Group Compound Questions

Subject	Individual Questions	Group Questions
Compound Origin	I'm interested compound AQR851. I'd love to talk to the scientist that made this compound. Who synthesized AQR851?	There are a lot of really interesting compounds that I've discovered came from contract lab RTY. What are all the compounds we have in the library that were made by this same contract lab?
Target Class	I suspect that compound RQB751 is a promiscuous hitter. What are all of the specific validated targets that RQB751 has hit?	I know that RQB571 hits enzyme target XYZ but what are all of the other chemical compounds that are validated hits against this enzyme XYZ?
hERG	I'm thinking of including compound RYU123 into my project. Oh wait, comments and links to assay results on the Compound Homepage shows that it has unsolvable hERG issues. Let's not use this compound.	I'm studying hERG activity. What are all the chemical compounds that have been implicated in hERG? What can I learn about the commonalities of this group?

Note: All project names, compound codes, targets, etc. are arbitrary and do not represent actual ongoing work at Novartis.

Search vs. Pre-determined Tags

In the previous example of searching for all compounds synthesized by Brede Wegener, why not just execute a search for "compound synthesis Brede Wegener" instead of accessing the information through pre-determined tagging structure? Although searching is very powerful and will also be included in the Novartis web structure, specifically listing pre-determined compound tags within the Compound Homepage structure has several advantages over a simple search.

First of all, search architecture would need to be in place to allow searching to return the appropriate information. Perhaps the data relating to the compounds that Brede Wegener has synthesized is located in the compound registration database. Does the current Novartis wide search engine query this database? If not, maybe the database does have a search interface but it would require a user to first find this interface. This means that a scientist who is new to the organization would have know that a) a database exists for compound registration b) it contains the

name of individuals who synthesized the various chemicals c) the database is called ABC d) the web interface to search this database is located at hyperlink xyz123. Secondly, even if the Novartis wide search engine queried this database in addition to other organization information it might return links that are unrelated to this specific question. For example, the organic chemistry group's internal webpage might be a hit "the GDC compound synthesis group supports all of NIBR's synthesis needs……. Group members include: Jack Johnson, Brede Wegener, Jennifer Duffy, etc."

Lastly and most importantly, the pre-determined tagging structure encourages random browsing and the exploration of relationships in a way that searching does not. Synonymous with browsing through the Amazon.com bookstore, a user might not know exactly what he/she is searching for but none-the-less stumble across powerful, interesting information. The tagging structure of the Compound Homepage website establishes the framework for this random browsing and actively encourages the exploration of relationships between compounds, assays, side effects, disease indications, people, etc.

12.4 Compound Wiki Page (Tab #4)

The 4th tab displayed on the Compound Homepage would allow the Novartis user community to collectively compile knowledge related to a specific compound and display this information to the entire organization. Unlike the "tagging wiki" information that is posted on the compound wiki would pertain only to one specific compound (vs. a group of compounds).

Wikipedia Editing Organization– The Compound wiki should adopt the structure of Wikipedia and allow users to view an article, contribute to an editing discussion specifically about the contents of the article, edit the page, add new sections to the article and view the page editing history.

History– The names of article authors, dates of editing contributions, as well as previous edit versions of the wiki can be viewed via the "history" hyperlink."

Wiki Guidelines– Guidelines are posted for users who are relatively unfamiliar with wikis, how to edit, etc. The guidelines section links to several examples of compound wiki pages that are "ideal wiki pages". These example pages can serve as a template for the creation of new content.

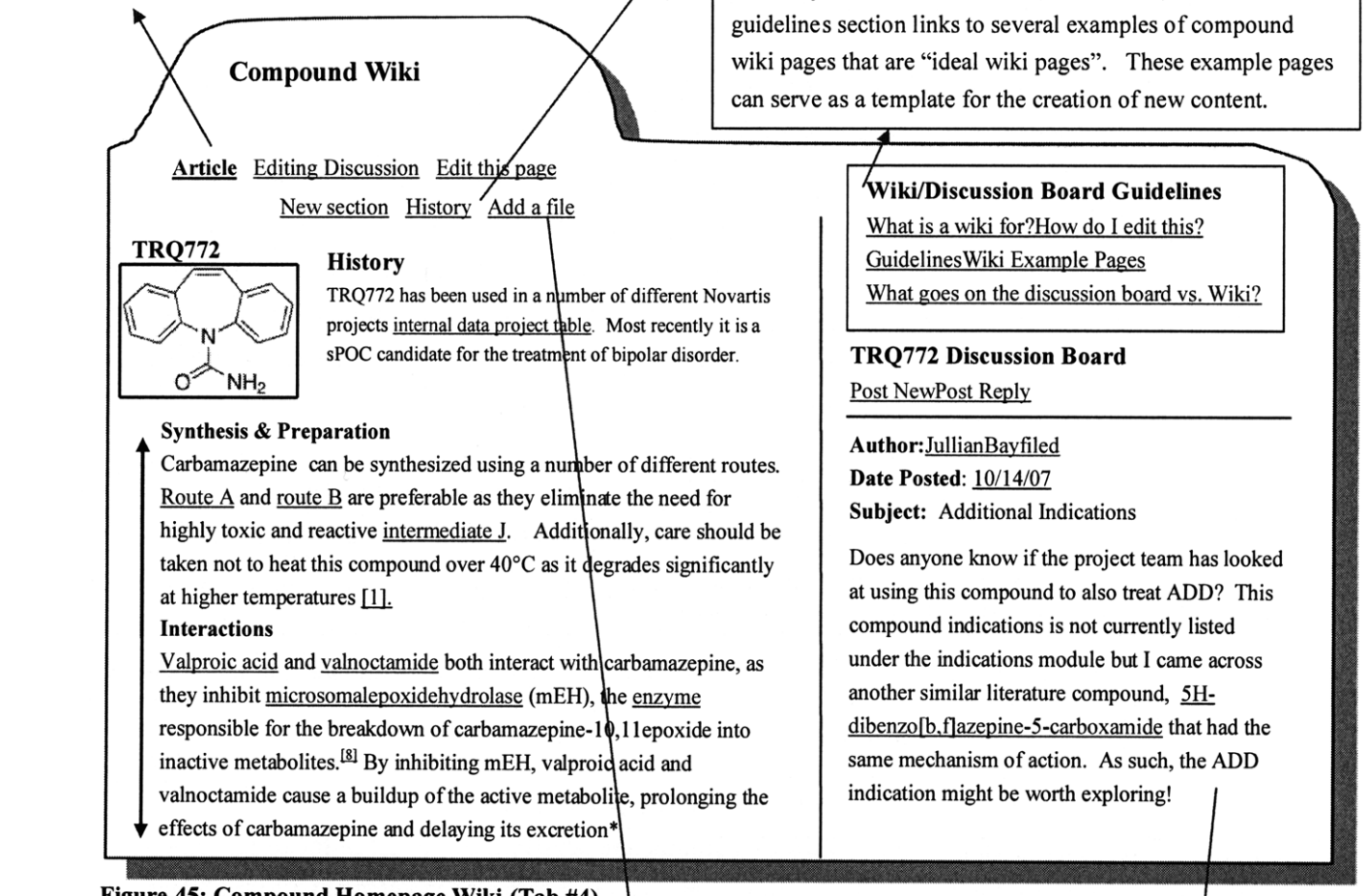


Figure 45: Compound Homepage Wiki (Tab #4)

Add a File – Interviewees throughout the user research stressed the importance of being able to attach files to the wiki. Observations further confirmed that compound information is kept in a variety of PowerPoint presentations, picture bitmatfiles, excel documents, MS Word files, etc. Users need to be given a place to actively store this valuable information so that it is accessible to the entire community. The "add a file" hyperlink would allow a user to add a file into a publically viewable folder of the Novartis document organization system that is associated with the particular compound, in this case compound TRQ772. The user could then obtain a hyperlink for this document to place in the wiki or discussion board.

Discussion Board– The compound discussion group encourages scientific discussion about a particular compound in the form of forum posts and replies to the topic. This discussion board is included to ensure that the wiki article is compilation of knowledge (think compound encyclopedia page) and does not turn into a discussion forum for questions, ideas, etc.

Wiki vs. Discussion Board Content

It is important to distinguish between a compound wiki and discussion board related to a specific compound. The compound wiki is a collaborative type document that captures information about the compound. The document is created and updated by the user community as the compound moves through the drug discovery process. The further along in the drug discovery process, the more will be known about a particular compound, and the more can be added to the wiki article. The compound wiki can be thought of as a flexible, ever changing encyclopedia article about a particular compound. Although discussion about the wiki may take place on the "editing discussion tab" this discussion is restricted to conversation and dialogue about what to add to the encyclopedia article, how the article should be formatted, edited, presented, etc.

Conversely, the compound discussion board provides a forum for <u>any</u> type of discussion about the compound. It does not necessarily have to be related to any of the content present in the wiki. The example on the TRQ772 discussion board above illustrates this point. The user Jullian Bayfield has posted a comment about a potential additional disease indication for the compound. As the Attention Deficit Disorder (ADD) disease indication has not necessarily been investigated it is not currently part of the collective knowledge compound encyclopedia. That said, the comment is very powerful and needs to be captured. The compound discussion board is the appropriate forum to do so.

The compound discussion board is currently presented on tab #4 to the right of the Compound Wiki article. This amplifies the distinction between what is on the wiki vs. what goes in the discussion board. That said, this may not be the appropriate location. Further investigation should take place to see if the discussion board should be incorporated on the compound wiki page (tab#4), or if another tab should be added to the Compound Homepage specifically for a compound centric discussion group.

Compound Wiki (tab #4) vs. Internal Data linking portal (tab #1)

If the internal data linking portal (Compound Homepage tab #1) captures all of the organizational information associated with a particular compound why is it necessary to also have a wiki? Yes, the wiki provides a text-based format for explaining information and results. In a sense it's a summary overview of the body of knowledge associated with a particular compound. The internal data tab directs individuals to very specific data while the wiki provides the high level compound experience or summary overview of that accumulated knowledge.

Upfront Organizational Structure

Rather than impose an upfront organizational outline, it is recommended that users be allowed to define the organizational structure of the wiki. Certain sections may be very relevant for certain compounds while other sections may not pertain at all. To assist users in determining what type of information can and should be written about in the wiki, specific examples of model compound wiki pages are provided in the user guide at the top of the wiki. Scientists can review these pages to get an idea of what structure is appropriate. In addition, editors are encouraged to debate structure and organizational content of the wiki in the "editing discussion section."

Maintaining Quality / Dealing with Controversy

In a wiki type environment, anyone can create or delete the content of another user. This inherent structure leads to an internal quality control mechanism. The founder of Wikipedia, Jim Wales explains "The wiki model is different because it gives you an incentive when you're writing. If you write something that annoys other people, it's just going to be deleted. So if you want your writing to survive, you really have to strive to be cooperative and helpful."[99]

[99] Rand, Matt "Best of the Web: Extreme Blogging," *Forbes*, 13 Dec. 04. Web. Accessed 21 Feb 09 http:www.forbes.com/best/2004/1213/bow001.html.

Chapter 13: Conclusion and Future Steps

Novartis needs to take several future steps to facilitate the creation of Compound Homepage and other elements of the new global scientific web environment. Currently, Novartis is actively undertaking several Web 2.0 and data improvement initiatives. Upgrades to many existing data systems are taking place, allowing metadata to be standardized and have common nomenclature. The company is also working on restructuring the internal web, implementing user driven SharePoint team sites, discussion boards, idea boards. Furthermore, several departments have made significant progress with the implementation of both target and safety profiling wiki environments.

Top Leadership and Organizational Commitment– In order to establish a global Web 2.0 environment, NIBR needs to officially commit to creating and implementing a new web-based navigation environment for scientific data. Currently a couple of departments are spearheading the effort but it is unrealistic to expect such an extensive initiative to be carried only at a departmental level. There needs to be a substantial commitment from top leadership to provide official resources and leadership support. The project needs to be staffed with multiple full time job positions. Individuals could be pulled from the NITAS organization, Center for Proteomic Chemistry, Global Discovery Chemistry or hired from an outside web consulting firm such as Avenue A. Razorfish (AARF).

Scientific Web Team–A "scientific web" leadership guidance team should be established. The team will guide and specify the overall organizational structure for the scientific web environment (i.e. Compound Homepage, Assay Homepage, Target Homepage, People, Project and Search). As many of the different spheres overlap and link together it is imperative that the system is first scoped out at a high system wide level with all of the working pieces in mind.

Build the System - Envision and build each of the different scientific spheres (target, compound, project, assay, people, and search).Make sure that user experience is similar between all systems so that users don't have to learn a different format or deal with a different interface when they jump from one world to the next. Additionally, each scientific sphere will need to be tightly integrated into the existing workflow to encourage adoption and provide "immediate benefit" for users.

Web 2.0 Training – Several literature review articles stressed the importance of training employees on how to use new Web 2.0 tools. John Castledine, the Director of Learning and Development for Pfizer's Global Research and Development (PGRD) Division, reflects on Pfizer's implementation of Pfizerpedia:

"There can be little doubt that for an organization to encourage the adoption of Enterprise 2.0 there must be a perceived overall benefit in doing so. To achieve sustainable change, it is vital that these benefits are presented from the frame-of-reference of the key stakeholders. For example: information overload is a major concern for most colleagues. Hence without understanding RSS feeds, blogs become yet more websites to add to your favorites list. Similarly, wikis and social bookmarking tags offer welcomed options to reduce email traffic within teams."[100]

Employees should participate in training to learning about wikis, blogs, RSS feeds, tagging, social networking, etc. Case studies of Flickr, Delicious, Yelp, Facebook and Wikipedia could be presented. Homework would include creating a blog, authoring content on Wikipedia, etc. This exposure will help employees understand and feel comfortable contributing to the Web 2.0 environment. This is especially important for older generations who have not necessarily grown-up actively participating in a user generated content environment.

Homepage Training/Organizational Communication- It will be very important for management at all levels to clearly and consistently communicate expectations associated with the new scientific web via training, talking about web environment in staff meetings, etc. Guidance needs to actively be established and communicated surrounding items such as: how to add document attachments to the wiki, the procedure for associating tags, etc. User community rules should also be clarified. For example, it's very important that users have the chance to disagree over posted content; however they should hash out editorial disagreements in the editing discussion section of the compound wiki, or perhaps an allocated discussion thread post rather than on the compound wiki article itself.

Management Involvement - Management's direct involvement in encouraging user contribution will be absolutely crucial for the adoption, usage and overall usefulness of the Web 2.0 environment. Many Web 2.0 tools require participation on a mass scale in order to achieve significant network effects. For example, if only a few users on Flickr tag photos, the majority of content never gets classified and searching via tags becomes worthless.

Lack of user participation becomes magnified in smaller communities. In large internet based user communities just a small number of users typically contribute content. McKinsey surveyed many users of online video sites to learn what motivates users and analyze user posting patterns. They discovered that just a few users posted the majority of the video content. Depending on the site 3-

[100] "Pfizerpedia: knowledge repository at Pfizer" *Chartered Institute of Personnel and Development, n.d.* Web. Accessed 21 Feb 09 http://www.cipd.co.uk/helpingpeoplelearn/_pfzrpd.htm

6% of the membership added 75% of the videos available for download and just 2% of the member base accounted for more than half of all videos viewed.[101] If the user community itself is very large, with millions of users such as Wikipedia, 3% still constitutes a substantial number of contributors. However, in a smaller based community with only 10,000 employees, contribution rates of only 3% could be problematic.

As the value of Web 2.0 systems are driven by the number of active contributors, it is imperative that management must actively promote the initiative rather than taking a hands-off approach and just relying on it to "take off" on its own. Harvard Business School Professor Andrew McAfee elaborates: "One huge fallacy of using Web 2.0 technologies within a corporation could be described as "if we build it, they will come."[102]

Different Approaches to Drive Contribution

Andrew suggests that corporations may actually have some distinct advantages over internet web communities in that they can actively influence user contribution[103]. One approach would be to specify that users must contribute, but not necessarily dictate or micromanage what exactly how and where they choose to contribute. For example, an employee's goal plan might specify that they need to contribute 200 times to the user community over the course of the next year. This contribution could be anything from tagging an object, editing a wiki, writing a section of a wiki, posting on a discussion board, rating content, etc.

Darren Lennard managing director a European investment bank Dresdner Kleinwort Wasserstein (DrKW) describes his personal experience in driving wiki adoption:

> *"To encourage use, Lennard put up an initial wiki page with a vague mission statement on it , e-mailed everyone to tell them about the new tool and what it could do, and encouraged them to start using it. Nothing happened. People weren't clear on what it was, what it should be used for or what its advantages were, so they stayed away. 'I realized that I had to be a lot more directive if I wanted behaviors to change' says Lennard, 'and I also had to put up wiki content that required users to get involved.' Lennard posted that agenda and action items of an upcoming meeting and suggested that people use the wiki for their*

---

[101] Bughin, Jacques "How companies can make the most of user-generated content" *The McKinsey Quarterly* (2007). Print.

[102] Brynjolfsson, Erik and Andrew McAfee "Beyond Enterprise 2.0: The Future of the Web", *MIT Sloan Management Review* 48. 3 (2007). Print.

[103] McAfee, Andrew. Personal Interview by Juliet Duffy. Aug 08.

responses to them. "I told my desk that I would no longer read e-mails on some topics" he says[104].

Other companies have experimented with additional enticements. In the McKinsey user generated content study, companies that were able to identify thought leaders and nurture users who posted quality content "boosted the overall number of contributors and improved quality of the postings." Other companies strove to make contributing and participation fun for employees. At Google employees placed online bets on the likelihood that particular ideas would be adopted. Still others, such as Intuit, used job rotation programs. Certain employees were periodically invited to contribute to the company's internal dialogue on a full-time basis. [105]

Regardless of the decided upon approach, Novartis leaders need to be actively involved in order to avoid the "if we build it, they will come" fallacy. User contributions will be essential to achieving the value and network effects of Web 2.0

Culture Change

Lastly, but most importantly, Novartis will need to establish an organizational culture that promotes information sharing, earning from and acceptability of failure, public dialogue, and trust. A number of the cultural barriers that Novartis will need to overcome are discussed in detail in the User Research and Novartis Culture and Three Lenses Analysis chapters. Actively working to change the culture must be a constant conscious initiative by leadership at all levels.

[104] McAfee, Andrew "Enterprise 2.0: The Dawn of Emergent Collaboration" *MIT Sloan Management Review*, 47. 3 (2006). Print

[105] Bughin, Jacques "How companies can make the most of user-generated content" *The McKinsey Quarterly* (2007). Print.

Conclusions

Web 2.0, the ultimate platform for tacit based knowledge work has finally arrived. User driven, collaborative platform based tools including wikis, web mash-ups, discussion boards, linkage based search engines, and tagging have the potential to vastly change how information is managed and how knowledge work is captured.

Web 2.0 has particular relevance for the pharmaceutical industry as drug discovery is an extremely knowledge intensive activity. Applying Web 2.0 tools to NIBR's targets, assays, people, projects, and compounds in a well thought out framework has the potential to yield tremendous productivity improvements in the drug discovery process.

In this new "scientific web" environment information will be set free. Knowledge will no longer be contained within organizational silos or behind data gatekeepers. It will flow freely to who-ever needs it regardless of where they sit in the organization. Pre-existing forced classification schemes will be eliminated so that they cannot bury important relationships into a forced one-dimensional format. Compounds, targets, and assays will use tagging to simultaneous classify information under multiple categories. This will enable users to instantaneously and effectively slice and combine the data any way that they can possibly imagine.

Lastly, effectively harnessing the collective intelligence of thousands of scientists within Novartis's worldwide research network will enable a paradigm shift. Where previously organizational size was an impediment, size will now be beneficial. A large, extremely knowledgeable user community can more effectively annotate metadata, hyperlink to important content, establish tags, and collectively author content. Such activities will not only improve the search ability of information but also allow important scientific connections to emerge linking biology to chemistry and furthering Novartis's understanding of disease.

Appendix A: User Research Interview Questions (full interview version)

Q1: Name
Q2: What organization are you in?
Q3: How long have you been at Novartis?
Q4: What is your educational background? (biologist, chemist, biochemist, etc.)
Q5: How long have you been in the Pharma Industry? Was this in a drug discovery role?
Q6: If you've had a job elsewhere in Pharma/biotech how did the data environment compare to what we have at NIBR?
Q7: What is your job function? What does your group do?
Q8: Do you deal with compounds on a 1:1 basis or more as a set?
Q9: What information about particular compound(s) do you **generate** on a regular basis? Where is this information that you generate currently stored? (ex: reports, databases, etc)
Q10: What is the Novartis culture like surrounding documentation of work? Do people check that things get documented appropriately? Do they care? Does this come up in yearend reviews?
Q11: What internal information about particular compound(s) do you **use/consume** on a regular basis? Where is that information stored? How often do you access each of these databases/info sources? Which ones are most important? (please rank priority)
Q12: What are the key pieces of info you need to make a decision(s) in your role/job?
Q13: How easy it is to find access to the internal information you are looking for? (scale 1-10 with 1=extremely hard)
Q14: What information is easy to find?
Q15: What information is hard to find? What makes it hard to find/access?
Q16: How much of your time do you spend looking for information? (avg. time per week) Any specific examples?
Q17: If you don't have access to a certain internal database how long does it take for you to get access? Any examples?
Q18: If you cannot get access to a database, or simply don't know where to find something what do you do?
Q19: What external data sources do you use? (ex: PubMed) and what information do you typically extract from these sources? How often do you access each of these databases/info sources (daily, weekly, etc)? Which ones are most important? (rank priority)
Q20: How often do you go outside of Novartis to access information in an external database? (daily, weekly, once a month)
Q21: Do you think you would reference external databases more frequently if these databases were available in one easy to find/click portal?
Q22: What wiki's (if any) are you aware of at Novartis?
Q23: Which wiki's do you currently use?
Q24: If all Novartis wiki environments were integrated into one, what would be the entry point that you would use most often? (Target, Profiling, Compound, HTS Pipeline)
Q25: Do you have any experience with using Wiki's (such as Wikipedia)? How much?
Q26: Have you ever edited or helped to create a Wiki?
Q27: Should Wiki be open to adding users content and links? To what extent? Who should input info? Open to anyone in Novartis? Project team members only?
Q28: How should the user added dialogue be structured? Pre-populated outlines to discourage random contributions?

Q29: How could we encourage people to add content to the wiki and avoid the blank page syndrome?
Q30: If the interface allowed for user input would you be interested in adding notes/text to keep the Wiki up to date?
Q31: How often do you see yourself contributing to the Wiki? (weekly, daily, monthly distribution
Q32: Have you ever used Facebook? (yes/no)
Q33: Have you ever used Linked-in? (yes/no)
Q34: Do you think the "friends" applications / ID of networks within Facebook could have relevance for compound Wiki? If so, how?
Q35: How could the friend/association groups be useful (other than just interesting info to have)?
Q36: Would a compound Wiki in general be useful to you? Would you use this tool (yes/no/maybe)?
Would you use the compound wiki in team meetings? How frequently?
Q37: Is there any specific information or functionality that you'd like to see included in Wiki?
Q38: Is there any kind of data analysis that could be done with Wiki that you couldn't do by just accessing databases individually?
Q39: What would these tools be? Could they change how you make decisions?
Q40: Would you be open toward participating in a test run of a pilot compound Wiki?
Q41: Compound Failures. Why do compounds fail? What would be useful for you to track?
Q42: What should the goal of the compound wiki be?

Appendix B: User Research Interview Questions (abbreviated interview version)

Q1: Name
Q2: What organization are you in?
Q3: How long have you been at Novartis?
Q4: What is your educational background? (biologist, chemist, biochemist, etc.)
Q5: How long have you been in the Pharma Industry? Was this in a drug discovery role?
Q6: If you've had a job elsewhere in Pharma/biotech how did the data environment compare to what we have at NIBR?
Q7: What is your job function? What does your group do?
Q8: What is the Novartis culture like surrounding documentation of work? Do people check that things get documented appropriately? Do they care? Does this come up in yearend reviews?
Q9: Do you deal with compounds on a 1:1 basis or more as a set?
Q10: What information about particular compound(s) do you generate on a regular basis? Where is this info that you generate stored? (ex: reports, databases, etc)
Q11: What internal information about particular compound(s) do you use on a regular basis? Where is that information stored? How often do you access each of these databases/info sources? Which ones are most important? (rank priority)
How easy it is to find access to the internal information you are looking for? (scale 1-10 with 1=extremely hard)
Q12: What information is easy to find?
Q13: Do you think you would reference external databases more frequently if these databases were avail. in one easy to find/click portal?
Q14: Links to external databases? If so, what?
Q15: What Wiki's (if any) are you aware of at Novartis?
Q16: Which Wiki's do you currently use?
Q17: Have you ever edited or helped to create a Wiki?
Q18: Panel #1 functionality (linking)
Q19: Panel 2 functionality (general user input). What should be included? Who should be able to add contents? Should these contents be structured? If so how?
Q20: How could we encourage people to add content? Avoid the blank page syndrome?
Q21: Would you add content? How often (weekly, daily, monthly) Why or why not?
Q22:Have you ever used Facebook? (yes/no)
Q23: Have you ever used Linked-in? (yes/no)
Q24: Panel 3 application: Do you think the "friends" applications / ID of networks within Facebook could have relevance for compound Wiki? If so, suggest some useful "friends/network" groups
Q25: Would the friend/association groups be useful or just interesting info to have?
Q26: Of the three panels (panel 1,2,3) what would be the most useful to you? Should we emphasize one of these functionalities over the other?
Q27: Is there any specific information or functionality that you'd like to see included in Wiki?
Q28: Who would be the users at each stage of the development process (or stages that you are familiar with) how could this wiki improve quality or fail faster?
Q29: Would you be open toward participating in a test run of a pilot compound Wiki?
Q30: Compound Failures. Why do compounds fail - what would be useful for you to track?
Q31: Individuals that would be good to contact
Q32: Other comments

Appendix C: Example of Novartis Safety Profiling Report

(Note: Compound information is for a non-proprietary, publically released compound)

Preclinical Profiling Report: All Data **NVP-LGC420** Group Salts **2009-04-16**

P1: Physicochemical characteristics

1. Solubility (HT-Eq Sol)

pH 6.8 (g/l):	0.3781
pH 1.0 (mM):	0.7214

2. Ionization Constant (HT-pKa)

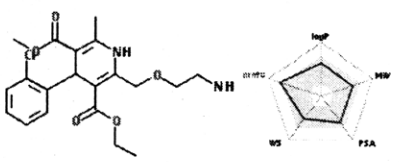
missing pKa: Y Calc.formula: B

3. Ionization Constant (HT-LogP)

Octanol/Water:

4. PAMPA (Passive Permeability) Class:

logPe pH 4.0	-5.8
logPe pH6.8	-4.7
logPe pH8.0	-4.1
log PAMPA	-4
calc FA [%]	99.8



MW: 408.89

P2/S5: Metabolic Clearance

	CYP Only		CYP+UGT	
	ER*	$t_{1/2}$	ER*	$t_{1/2}$
Rat		n/a		30.8
Mouse		n/a		n/a
Human		n/a		43.5
Dog		n/a		n/a
Monkey		n/a		n/a

* Hepatic Extraction Ratio

P2: CYP450 Inhibition

Assay	IC_{50}
3A4 Midaz	7
3A4 Testos	n/a
2D6 Bufur	n/a
2C9 Diclof	n/a

P3: Caco-2 (Active Transport)

Cell Permeability Class:	
PappA-B (cm/sec x 10^{-6}):	4.33
PappB-A (cm/sec x 10^{-6}):	4.41
Ratio (B-A/A-B):	1
Mechnism of transport:	Passive Transcellular
Comments:	Recovery < 80%

P4/S8: hERG Activity

Assay	Inhibition			IC_{50}	Rank
Binding (P4):	15.0%/1µM	36.0%/10µM	55.0%/30µM	18.89 µM	
Patch Clamp: QPatch (S8)			%/30µM	µM	
Patch Clamp: performed by PCS				µM	

S10a: Cardiac sodium peak current

Patch Clamp: Nav1.5 Quattro:	µM	

S10b: L-type calcium peak current

Patch Clamp: Cav1.2 Quattro:	µM	

P7: HC Micronucleus:

Result	TC_{50}	MaxConc
NEG/13.48uM	19.82	200

PCS:TK6 Micronucleus:	No Data
PCS:Mini Ames:	No Data

P5: Safety Pharmacology

Assay	IC_{50}	Assay	IC_{50}
5HT1A	> 10	h Motilin	> 10
5HT2A	> 10	M1	> 10
5HT2B	4.3	M3	> 10
5HT2C	14.5	NK1	> 10
5HT6	2	hr NT1	> 10
5HT7	> 10	op-delta	5
Ad1	> 30	op-kappa	1.2
Ad2A	> 30	op-mu	> 10
Ad3	> 30	TP	> 30
Alpha1A	5.7	Y1	> 30
alpha1	> 10	Y2	> 30
alpha2A	> 10	hr Y1a	> 10
alpha2B	> 10	hr Y2	> 30
alpha2C	10.4	r B2D	> 10
beta1	> 10	r GABA A	> 10
beta2	1.8	Nic(ns)	
beta3	28.5	r PCP	> 10
AT1	> 30	5HT3	2.3
B1	> 30	r Ca2+(L)	0.1
B2	> 30	r Ca2+(N)	> 10
CB1		r K+(ATP)	> 10
CCKa	> 30	AR	> 30
CCKb	> 30	ERa	> 30
CRF1	> 30	ERb	> 30
CRF2a	> 30	PR-B	> 30
D1	> 10	hr PXR	
D2	2.9	GR	> 30
D3	2.8	AdT	> 30
D4.4	> 10	DAT	3.3
ETa	> 30	NET	> 10
ETb	> 30	SHTT	> 10
h GABA B	> 10	COX-1	> 30
GIP	> 30	COX-2	> 30
GAR	> 30	MAO-A	> 10
GHS	> 30	h PDE3	> 10
H1	> 10	h PDE4D	> 10
H2	9.8		
H3			
MC3	16.5		
MC4	15.4		

Proteases	IC_{50}
s. hCaspase3	30.2
s. hCathepsinD	> 30
s. hMMP08	17.5
s. hThrombin	11

Kinases	IC_{50}
EPK ALK	>10
EPK CDK2A	>10
EPK FGFR3K	>10
EPK HCK	>10
EPK HER1	>10
EPK HER2	>10
EPK IGF1R	>10
EPK INS1R	>10
EPK JAK1	>10
EPK JAK2	>10
EPK JAK3	>10
EPK KDR	>10
EPK MK2	>10
EPK PDK1	>10
EPK PKA	>10
EPK PKBa	>10
EPK RET	>10
EPK SYK	>10
EPK TYK2	>10
EPK cABL	>10
EPK cABLT315	>10
EPK cKIT	>10
EPK cMET	>10
EPK cSRC	>10

Functional Assays	ago	antago
5HT1A	12.8	19.7
5HT2A		
5HT2B	> 10	11.7
Alpha 2A	> 10	> 10
D1	> 10	
M2	> 10	> 10
H2	> 10	
PR-B		

Appendix D: User Research Compiled list of Grouping Suggestions

# Users Suggesting addition	Tagging/Group Suggestion
11	Project/Program association
9	target hit (in general) – general consensus that these should be validated hits
8	Similar Chemo type (backbones and scaffold similarity) - desire to search against both internal and external compounds
8	Synthesized by this Chemist (First Name, Last Name)
5	Target families (I belong to family of GCRP, Kinase inhibitor, nuclear receptor, etc) Families I would include: Kinase, GPCRs, Channels and Nuclear Hormone receptors these are the four main groups. Others ex: "potent renin inhibitor" , protease
5	Disease association, indication
3	Target hit - safety profiling target hit / off-target hit
3	Act on a specific pathway, I exhibit activity in this pathway
3	Chemotype or chemical series designation (what compounds a scientist /project team made within a given series)
2	By project phase (I'm in phase D2a)
2	Individuals that have worked on this compound
2	Similar/same pharmacophore groups
2	Assays compound has been run in
2	Other compounds that have IC50 curves in a particular assay.
1	Target (validated hit is ambiguous, I'd like to see top 10% of compounds that are active against this target."
1	Potent hERG inhibitors
1	Promiscuous hitters in pharmacology safety profiling assay (different than just hitting one safety profiling...combo of a & b & c & d)
1	Promiscuity against regular target (like kinases) I'm validated against 10 kinase (hitting 3 or more bad)
1	Compounds that hit a certain safety assay (ex: compounds that have hit a hERG assay)
1	Going beyond Disease areas - Specific indications (ex: tumor suppressor)
1	I've been used in indication xxx for POC
1	I've been used in indication xxx for a full-phase development trial
1	Structure - *pair wise comparison - the number of compounds that are identical but one change away (one atom deleted, or added)
1	Structure – by changing functional groups. this compound with a phenyl but there is also the pryidol. Instead of phyenol but hetrocycle.
1	Structure – by substructure
1	Scaffold - purines, etc. esteroids, benzodyasipines - how many of the compounds in the archive that belong to the purine family ?
1	We all produce/or have been linked to xx side effect
1	Frequent hitter in xx assay technology
1	list of individuals within Novartis that have commented on a compound in the Wiki
1	Formulations (I'm used as a topical, pill, etc)
1	Compounds that have made it to CSP
1	Individuals that part of the project

1	Toxicity	
1	Categories of activities based on assay results ex: this group of compounds has Kinase activity	
1	Predicted activities based on structure (predict this will have kinase activity)	
1	Discontinued compounds (and it what stage)	
1	Compounds that have had a false (+) in HTS - if certain compounds kept coming up you could link this to maybe interference in a certain assay, etc.	
1	Binding to targets with assay information. Example: "we are active against the same target but in a different assay - equivalent of we went to the same school but in different years"	
1	Explore liabilities associated with compounds (safety, etc)	
1	Other NCBI databases	
1	In-licensed from another company or synthesized internally	
1	"Ugly compounds" – these may interact with a target but chemist won't touch it.	
1	Hit against families (kinase, GPCR) and even sub-families of those	
1	General Social Network relationships	
1	Compound is in an external patent	
1	Homologues were else is this binding - to the rat target? Human target? Etc	
1	Receptor family	
1	"Bad apples" We have a couple of thousand compounds that we know we hate~	
1	Where did compound come from? (i.e. contract lab in Singapore?)	
1	Hit in a particular assay	
1	Individuals that have ever requested this compound (in TRT) and for what purpose	
1	Compound structure analogs (viewing clusters of compounds) – with a slide able scale	
1	Reactive structure	
1	Frequent hitter (promiscuous compounds) – maybe broken down by assay?	
1	# of steps in synthesis	
1	Common known targets "HDAC inhibitors"	
1	Generic mechanism of action Ex: I belong to the family of cytostatics,I don't allow cells to replicate/anti poliforative.	
1	Series for how solubility got better optimizing compounds	
1	Products of parallel synthesis - prepared in a similar way (see diagram in Juliet notebook). This would be like all 24 compounds you prepared together in a parallel. Intermediate 1, intermediate 2, group these as parallel library, this would have to be defined by the individual chemist.	
1	compound group with good bioavilaiblity	
1	lousy stability in microsome - high in vitro clearance.	
1	very low inhibition IC50 (P2 on safety profiling) - I'm an inhibitor of 2C9 I'm an inhibior for 2D6	
1	Collaborators (other companies) that have worked on this compound – requested by strategic alliances	
1	Origin – where did this compound come from? Another company? A partner Collaboration?	