Bioinformatics.Org

2002 ORGANIZATION PLAN

September 20, 2002

MISSION STATEMENT

Bioinformatics.Org is an international organization which promotes freedom and openness in the field of bioinformatics. This is done by providing free and open resources for research, development and education so that such resources can be further developed. The organization is non-profit and maintains an Internet site by the same name where these resources can be accessed. Bioinformatics.Org hopes to lower the barrier to entering and participating in the field of bioinformatics, as access to cutting-edge resources can be prohibitively expensive for those working individually, in small groups, at poorly-funded institutions or in developing nations.

CONTENTS

1 BIOINFORMATICS.ORG PAST

1.1 The founding of Bioinformatics.Org

The Organization was founded by J.W. Bizzaro in the year 1998. It was initially, and still is, co-located at the University of Massachusetts Lowell, in the Center for Intelligent Biomaterials, where Prof. Kenneth A. Marx is the Director. The initial goal of the Organization was to build off of existing collaborations between J.W. Bizzaro, Ken Marx, and people at various institutions. A website was developed for the Organization, at that time called "The Biological Informatics and Computation (BIC) Group," to make available the bioinformatics software developed.

Very shortly thereafter, interest came from people who were not collaborators but merely interested in a place to host their own Free[1] and Open Source[2] bioinformatics software.

Below appear the founding (early) members of Bioinformatics.Org and their current affiliations. Affiliations as of the founding, if different, appear in parentheses.

J.W. Bizzaro, M.Sc.
University of Massachusetts Lowell
(Boston College)

Kenneth A. Marx, Ph.D.
University of Massachusetts Lowell

Peter D. St. Onge, M.Sc.
University of Toronto, Canada
(McGill University, Canada)

Gary Van Domselaar
Wyeth Research, Massachusetts
(University of Alberta, Canada)

Mark Luo
Freelance consultant in California

[1] Defined by the Free Software Foundation (www.fsf.org)
[2] Defined by the Open Source Initiative (www.opensource.org)

1.2 Motivations: Ethical issues in the field

The founders of the Free and Open Source Software movements of recent decades mistakenly believed that scientists work in an environment where everything is shared freely and openly, and they sought the same for environment for themselves. These ideas served as a portrait of an ideal, ethical environment for the scientific process to function, although it seemed only non-scientists were actively trying to embody them. As recently as 1998, no Internet resources were dedicated to Free and Open Source Software for bioinformatics. Additionally, the ideas hadn't been applied to certain ethical issues in the field.

Bioinformatics.Org is a scientific organization (i) promoting freedom and openness[3] in all of bioinformatics, not just for software; (ii) providing free and open resources to all bioinformaticists; and, (iii) practicing freedom and openness in research, development and education. Bioinformatics.Org was born out of this perceived need for such in bioinformatics.

1.2.1 Commodities in bioinformatics

Bioinformatics.Org became concerned with ethical issues in the field of bioinformatics, particularly as they pertain to the environment in which the scientific process functions. Of great interest to the founders of the Organization are the commodities in the field and how they affect this process.

Bioinformatic methods and materials[4] are commodities as the result of the precedent set by the patenting of pharmacological drugs and industrial technologies. One well-known example of this "commoditization" is the plethora of patent applications submitted for genes during the Human Genome Project, a practice that continues today. Many genes for HIV, cancers and blindness, for example, are currently owned. As with pharmaceuticals, newly discovered sequences are considered unique chemical compounds and are thus patentable.

Patents are also pending on algorithms and software. Commercial bioinformatic applications are priced into the millions of dollars per institution, restricting the re-use of the methods. In the US, algorithms and software can be patented.

Also of concern is the process of scientific communication. Copyrights on published research results in most cases belong to publishing companies and not to the researcher or granting agency. The availability of methods and materials men-

[3] Both freedom *and* openness are ideas that can be applied to practicing bioinformatics.

[4] In bioinformatics, the methods are the algorithms and software used to analyze the materials, which are the biological data.

tioned in the publications is often not required. And, subscriptions to these publications can cost thousands of dollars.

And these issues do not just affect those in industry. Academic institutions have been increasingly concerned with capitalizing on intellectual property. The Bayh/Dole Amendment has made this a legal requirement in the US.

1.2.2 What are the implications?

Scientists are inhibited in the use of bioinformatic methods and materials, and in the use of educational materials. And the effects are far-reaching. It has been especially restrictive for people in developing nations, but even businesses are inhibited, being limited in the application of basic science toward product development.

1.2.3 Why do we care?

We believe that science is different from most other human endeavors, that science is ultimately for the advancement of humanity.

Nearly all scientific discoveries made today are based on earlier work (both methods and materials) that had been given away. As Isaac Newton said, “I stand on the shoulders of giants,” referring to the work that had been done before him.

And we believe that a free and open environment helps to meet scientific needs. First, there is the need to communicate. In the Hindu proverb of the “Blind Men and the Elephant,” several blind men come across an elephant. One man feels a leg of the elephant and thinks an elephant is a tree; another man is sprayed by the trunk and thinks it is a hose; another man feels the tail and thinks it is a rope; another man feels an ear and thinks it is a fan; and, another man feels the side and thinks it is a wall. In this proverb, an inability to see the larger picture lead to incorrect conclusions. The same can be true for the scientific process when there is an inability to communicate.

There is also the need to confirm published results. In a free and open environment, mistakes can be found and fixed, and new directions and approaches can be taken.

And, of course, there is the need to learn and to educate. Students learn from example.

These are the same arguments for the use of Free and Open Source Software in general, but they apply to bioinformatic data and educational materials as well.

Science, in addition to Open Source Software, is a "gift economy"[5].

In a survey of 1,240 geneticists from 100 US universities[6], 86% said that sharing had decreased or stayed the same; 47% had been denied access to data; 28% of those denied could not confirm published results; 73% say that withholding data is slowing progress in the field; 63% say that it is harming relationships among peers; And, more than 50% say that it has adversely affected their own research or the education of their students.

1.2.4 What keeps others from stealing?

If you give things away, what keeps others from taking credit[7]? This is the most common response to the suggestion that there be more sharing in science, and it brings up a very important issue. Scientists would not copy text without giving credit (reference), because there is a level of self-policing in the discipline that prevents it. As a student in academe, one would fail a course or face expulsion for plagiarizing literature. The same must logically apply to methods and materials. Freedom and openness are themselves inhibited by this form of plagiarism, and the ideas cannot be fully realized until the plagiarism is recognized and dealt with.

1.2.5 Are these ideas anti-business?

Some people hold that the ideas of freedom and openness are inhibitive to business. Afterall, the marketplace operates on the principles ownership and resale at a profit. Perhaps businesses which commoditize bioinformatic methods and materials would take issue with us, but our advocacy is only for the implementation of these ideas in the environment where the scientific process functions. For the vast majority of businesses, even those in pharmaceuticals and biotechnology, the commidization of science is itself inhibitive. Most of these businesses invest a large percentage of their research and development budgets to reproduce and replicate what their competitors already have. If scientific methods and materials were public goods, there would be no concern for their theft, and these expenses would be saved.

[5] Eric S. Raymond, of the Open Source Initiative, has characterized the Open Source Software movement as a "gift economy."

[6] JAMA 287:473-480 (2002)

[7] Or "stealing." Is it credit or profit that motivates ownership?

2 BIOINFORMATICS.ORG PRESENT

2.1 Organizational structure

The structure of the Organization appears in Figure 1. “Bioinformatics.Org” is the name of the Organization and covers several services offered to members. First and of greatest importance is “The Open Lab,” which is the name for all project hosting services and divided into research, development and education. More information about The Open Lab can be found in §2.4.

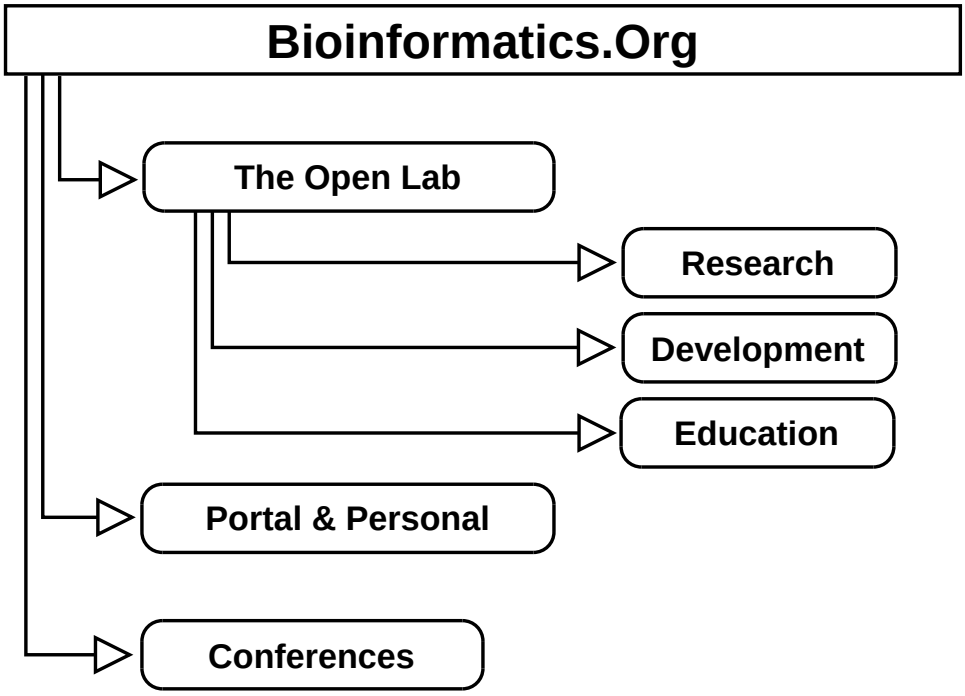

Figure 1: The present structure of the Organization

Bioinformatics.Org provides additional services and resources for its members, particularly through the website. Some are portal services, directing visitors to important information that may not be on the website. Some of the more significant portal services are the “BiO News” forum, the Events Calendar, the Software Registry, and the “BiO Jobs” forum.

More personalized services are rendered to members of Bioinformatics.Org, some of which are under development. Members are able to post “opportunity sought” messages on the BiO Jobs forum. And, recently, members have been

able to post biographical information and a *curriculum vitae*, making it easier to be located by potential employers and colleagues with similar interests.

Another part of the Organization deals with conferences, where members and potential members can meet in person and meet the officials of Bioinformatics.Org. The Organization holds an annual meeting, currently in conjuction with the O'Reilly Bioinformatics Technology Conference[8], and it is present at various other events through media partnerships with conference coordinators.

2.2 Officials

Officials of the Organization are the members responsible for its operation. All officials are currently volunteers (they are not payed).

2.2.1 Trustees

The Board of Trustees is being formed. The principal trustees are legal trustees, and those chosen so far appear below.

J.W. Bizzaro, M.Sc. Chairman
University of Massachusetts Lowell

Gary Van Domselaar Vice Chairman
Wyeth Research, Massachusetts

Mark Luo
Freelance consultant in California

Jon K. Stewart, J.D.
Larsen & Gruber LLC, Utah

2.2.2 Honorary trustees

Bioinformatics.Org will have a number of honorary trustees, who are not legal trustees but can vote on issues before the Board. The purpose of this position is to give members, both organizational and individual, who have contributed significantly to the Organization, a say in its direction.

[8] conferences.orielly.com/biocon/

2.2.3 Officers

In addition to trustees, Bioinformatics.Org has executive positions for people who are involved in the daily operation of the Organization. Below are the current officers and their positions.

J.W. Bizzaro, M.Sc. President
University of Massachusetts Lowell

Gary Van Domselaar Vice President & Chief Technology Officer
Wyeth Research, Massachusetts

Mark Luo Treasurer
Freelance consultant in California

Holly Miller, Ph.D. Secretary & Chief Public Relations Officer
State University of New York, Stony Brook

2.2.4 Advisors

Bioinformatics.Org has a Board of Advisors. Advisors are chosen by the Board of Trustees and are people whose experience is valued. Their names appear below.

Prof. Kenneth Marx, Ph.D. Chief Advisor
University of Massachusetts Lowell

David Lapointe, Ph.D.
University of Massachusetts Medical School

Prof. Robert Harrison, Ph.D.
Georgia State University

2.2.5 System administrators

In addition to those mentioned previously, below are some of the more active volunteer system administrators:

Marcos Oliveira de Carvalho

Joseph Landman, Ph.D.

Martin Kucej

Jayson Falkner

Dmitry Mozzherin, Ph.D. **Josh Goodman**

Satykam Goswami

John S.J. Anderson **Brendan Glackin**

Most of the system administrators are university students interested in bioinformatics. And many have stated that they hope their contributions to the Organization will help establish them in the field.

2.3 Membership

The Organization has memberships at two levels: individual and organizational.

2.3.1 Individual

As of September 2002, there are about 4,000 “individual” members world-wide. This type of membership is the basic membership in the Organization. There is no membership fee, and membership doesn’t expire.

In addition to having access to all of the services at Bioinformatics.Org, individual members can vote on issues affecting the Organization. The vote is initiated by the Board of Trustees and is considered a non-binding recommendation to the Board.

As can be seen in Figure 2, membership growth increased in rate during the summer of 2001. The rate in growth and the number of members are much greater than might have been predicted and are taken as a measure of the success of the Organization.

Individual membership at Bioinformatics.Org currently requires that the person give their real name and e-mail address (of course, in addition to username and password) but not any other information. Unfortunately, that means we do not have an accurate assessment of who the members are and where they come from.

We do, however, have e-mail addresses and can acquire some information from them. Figure 3 shows the different e-mail primary domains for the members. Most are, of course, in the .com domain, as many people use e-mail services such as yahoo.com, hotmail.com and aol.com. The same is true for the .net domain.

Below are the top 10 (in quantity) secondary domains for each primary domain. Again, .com and .net addresses seem to be mostly from commercial e-mail services, but there are many addresses from computer, pharmaceutical and biotechnology businesses which do not make the top 10.

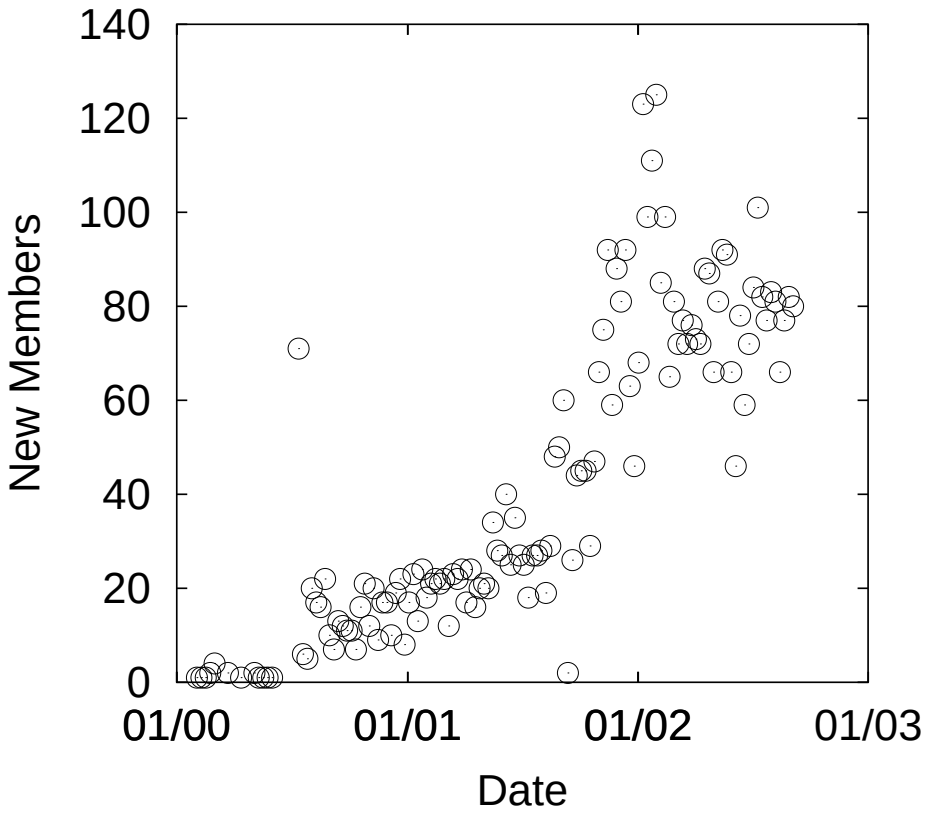

Figure 2: The number of new members per week, since early in 2000

Of note in the .com list are the domains rediffmail.com and indiatimes.com. These are e-mail services in India. Along with the number of .in domains, we can discern that more than 10% of our members are from that country.

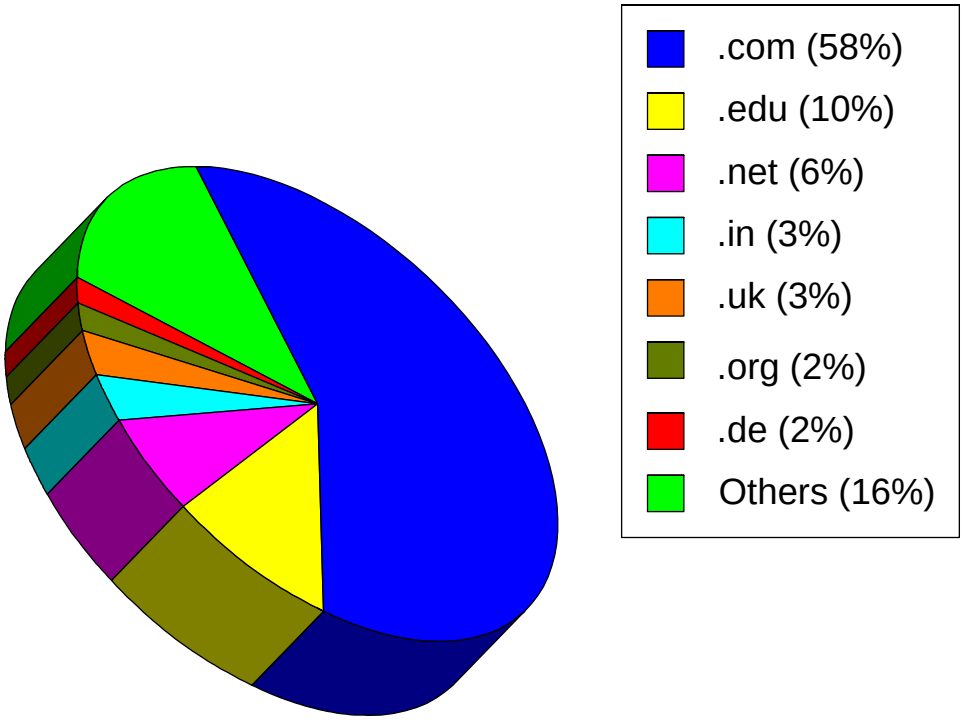

Figure 3: E-mail address domains of members

Top 10 .com	
yahoo	810
hotmail	674
rediffmail	296
indiatimes	64
aol	49
lycos	23
msn	21
rr	20
email	19
attbi	18

Top 10 .net	
usa	30
earthlink	22
netscape	16
hanmail	16
263	16
att	11
verizon	10
hinet	9
gmx	9
pacbell	6

Top 10 .edu	
mit	20
harvard	12
umn	11
uga	10
ufl	10
washington	9
stanford	9
wisc	8
umich	8
msu	8

Top 10 .org		Top 10 .gov	
bioinformatics	19	nih	15
cgiar	11	lbl	8
e-cell	6	usda	4
ieee	5	ncifcrf	3
computer	4	lanl	2
g-language	3	fda	2
acm	3	uspto	1
wadsworth	2	pnl	1
systemsbiology	2	ornl	1
nodalpoint	2	nrel	1

In the summer of this year, the International Society for Computational Biologists[9], the major professional society in the field, reported that they have about 1,300 members. That would make Bioinformatics.Org about 3 times the size. However, the ISCB charges a fee for membership, which expires every year.

2.3.2 Organizational

Organizational memberships, akin to corporate partnerships, are open to all formal organizations, for-profit or non-profit. There is no fee for this type of membership, and there are no voting rights[10]. Organizations may obtain this type of membership through an agreement to collaborate on a project or by becoming a sponsor of Bioinformatics.Org.

2.4 The Open Lab

"The Open Lab" is the name for the project hosting services at Bioinformatics.Org and currently comprises all assets of the Organization[11].

[9] www.iscb.org

[10] None beyond those granted to individual members who may also be employees or members of the organizational member. An agreement, however, may be made whereby the organizational member has a temporary seat on the Board of Trustees.

[11] "The Open Lab" was also the second name of the Organization, just prior to the name "Bioinformatics.Org."

2.4.1 Facilities

Bioinformatics.Org is located at the University of Massachusetts Lowell, in the Chemistry Department, at the Center for Intelligent Biomaterials. Prof. Kenneth A. Marx is Director of that Center.

2.4.2 Computer hardware

Bioinformatics.Org maintains the hardware listed in Table 1.

Table 1: Current Computer Hardware

Make	Processor	Memory	Storage
Dell PowerEdge 1300	500 MHz Pentium III	128 MB	9 GB
Dell PowerEdge 300	800 MHz Pentium III	128 MB	65 GB
Macintosh G4 Server[a]	533 MHz G4	512 MB	10 GB

[a]Contributed by Apple Computer as part of the Co-Lab Program (§2.5)

2.4.3 Research resources

Bioinformatics.Org hosts several online bioinformatic databases. Among them are A Database of Immigrant Genes, A Database of Leukemia Genes, and A Database of Pancreatic Cancer Genes.

The Organization also hosts online bioinformatic research tools, some with varying degrees of database usage. These include COMBOSA3D, PeCoP, SeWeR, SMS, and Sequence Extractor.

Members can subscribe to the BioMail service, which is a digital agent that routinely searches for newly published research articles with matching search criteria and e-mails the results back to the subscriber.

The registry and repository for bioinformatic software, currently under development, allow researchers to locate Free and Open Source Software, whether at Bioinformatics.Org or elsewhere.

The Organization also offers free compute time for members working on projects. This service will be much improved by the Co-Lab Program (§2.5).

2.4.4 Development resources

Bioinformatics.Org uses a project management system (PMS) based on the SourceForge system developed by VA Software[12]. It is Web-based and uses MySQL for the management of information. The PMS is used to offer the resources listed below and more.

- Bug tracking
- Task assignments
- Shell account
- Website
- FTP
- Concurrent Versions System (CVS)
- Mailing lists
- News and forums
- Domain name hosting

2.4.5 Educational resources

Bioinformatics.Org already hosts some educational resources. Some projects in The Open Lab are for education: for example, *TPE des 1res S du LFB 2001-2002* and *La perception des couleurs par l'oeil*, which are French-language, biochemistry education projects.

The site also features a bioinformatics FAQ (Frequently Asked Questions), perhaps the most extensive, general bioinformatics FAQ in the world. It has been contributed to by more than 30 people.

2.4.6 Hosted projects (all types)

There are more than 70 projects hosted at The Open Lab. Several of them are permanent features. Figure 4 shows the growth in projects hosted.

Below are the names of most of the projects currently hosted at Bioinformatics.Org.

A Database of Leukemia Genes
A Database of Pancreatic Cancer Genes
A Smith Waterman algorithm on GreenTea
AcE
AnnHyb
ANTHEDNA
Biochemical Network Visual Designer
Bioinformatics Benchmarking System
Bioinformatics Package Repository
biolib fortran library
BioQuery
Cardiovascular Gene Database

[12] www.vasoftware.com

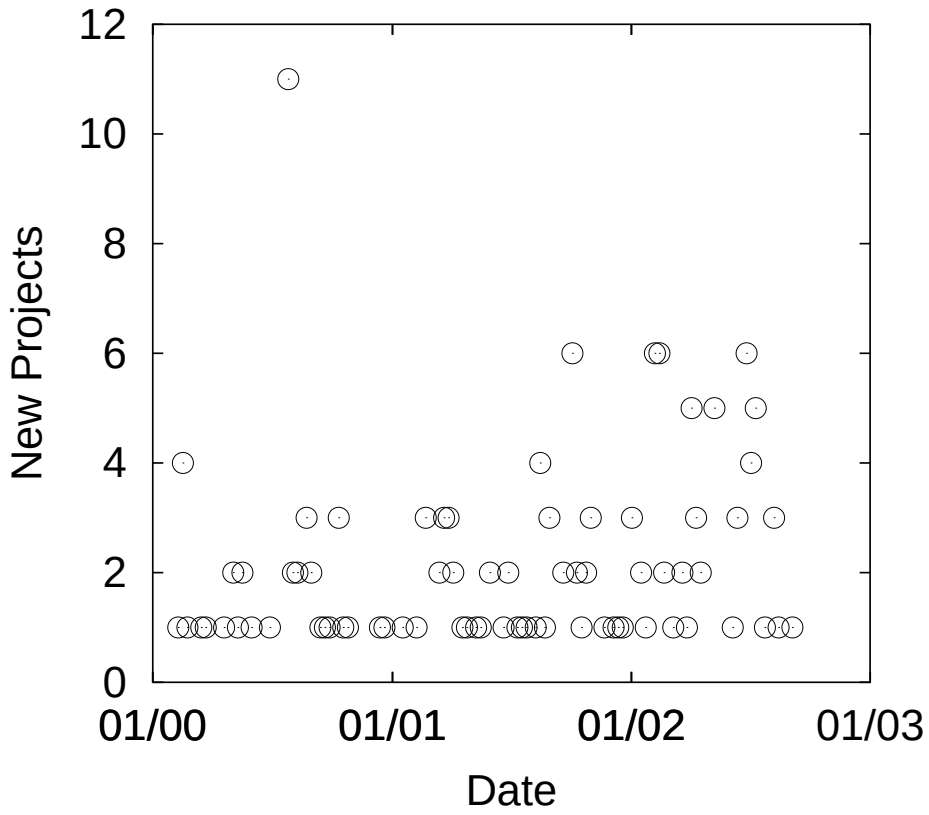

Figure 4: The number of new projects per week, since early in 2000. Data include all projects, whether viable and active, inactive, created for test purposes, or ultimately deleted.

CGInformatics
COMBOSA3D
DeMasker
DEODAS
Distributed Computing Power Project
dnacgr
E-CELL Simulation Environment Version 1
E-CELL Simulation Environment Version 2
E-CELL Simulation Environment Version 3
EnzymeLab
FishermansFriend:The Old Man And The Sea
FLIMS
FragSize
FreeMOL Molecular Modeling Environment
G-language Genome Analysis Environment
gBioSeq (Linux DNA sequence editor)
Gene Duplication
GeneGnome
Genpak
Genquire
GFIT
Ghemical
International Crop Information System
JaMBW
La perception des couleurs par l'oeil
Librarian

Mac5 phylogenetic reconstruction
MassQuest
Mavric
MBRC
MeltSim
molbiol_lectures
Multi-Genome Navigator
Network Tools & Applications in Biology
OpenGL based 3D Superimposer
PCalc - Primer Concentration Calculator
Perl Knowledge-Based Objects
Persistently Conserved Positions
Piper
Populations
Preclinical datamanagement
Project For Endangered Species
Protein Structure Toolkit
seKs sequence editing KDE suite
Sequence Extractor
Sequence Pattern Modeling and Searching
SLRI Bioinformatics Toolkit
SolanGENE
SpecBio02
SPRIG - Informatics for Plant Biology
StructureKit
Systems Biology Workbench
Tandem Repeat Occurrence Locator
Text2Knowledge
The Biodatabases Open Lab
The database of Immigrant genes
The Macroshack
The Open LIMS Project
The Open Notebook Project
The Operon Group
The Sequence Manipulation Suite
ThreeAAAnalysis
TPE des 1res S du LFB 2001-2002
W-curve
Wide Area XML System

There are two reasons for hosting a project with an online project management system. First, a project may not be affiliated with an institution with its own website. Second, a project may be developed by people in disparate places, thus requiring project management to be online.

The Open Lab competes mostly with SourceForge.net, the generic[13] project management system from which The Open Lab's own PMS was derived. SourceForge.net has roughly the same number of bioinformatic software development projects hosted there but more than 47,000 non-bioinformatic.

The advantages that we promulgate over SourceForge.net are as follows.

1. Our focus is on bioinformatics, with bioinformatic open labs (categories) coming.
2. We host web-based, server-side executables.
3. We host web-based databases.
4. Project members have shell-level access to CVS and other resources.
5. Shell account use is not limited to compiling.

[13] SourceForge.net is specifically for software development projects but not specifically for bioinformatics.

6. Storage is not limited.

7. We can accommodate special needs, because we're smaller (about 70 projects vs. more than 47,000) and can keep track of members more easily.

8. Bioinformatics.Org is non-profit, and even our resources are Open Source.

9. Project members can give a presentation about their project at the Annual Meeting (at the O'Reilly Bioinformatics Technology Conference), and the presenter attends for free.

A comparison of the ratios of members:projects between Bioinformatics.Org and SourceForge.net reveals an interesting difference. Bioinformatics.Org's ratio is about 50:1, while SourceForge.net's is about 10:1. We take this as an indication that our members are joining for more than just the opportunity to create software development projects. And so our emphasis on research, educational, personal and portal services should increase.

2.5 The Co-Lab Program

Bioinformatics.Org is seeking to collaborate with the manufacturers and vendors of cumputer hardware and software, on the development of "The Open Lab" project hosting resources, for the mutual benefit of all involved. The "Co-Lab" program, as it is called, allows collaborators to co-locate products either on-site or off-site, with the option of partaking in the global Bioinformatics.Org Grid.

Bioinformatics.Org addresses the problem of the high barrier to entering the field of bioinformatics. Access to cutting-edge bioinformatic resources are prohibitively expensive for most practicing and neophyte bioinformaticists. To lower that barrier, the Organization hosts resources, under the name "The Open Lab," that can be accessed free of charge.

However, the computing resources required for Bioinformatics.Org to fulfill its mission are not growing with the membership numbers. The Organization therefore proposes a mutual-benefit collaboration for the development of The Open Lab with manufacturers and vendors of computer hardware and software.

2.5.1 Options

Two options exist for the placement of hardware and software products:

Product Co-Location This is much like the option available to members of the Open Source Development Lab[14]. Hardware and software products would be located (placed) at Bioinformatics.Org facilities, for use by our members. However, the products would be used specifically for bioinformatics, and use would include the execution of applications (for research) as well as porting.

The Bioinformatics.Org Grid Co-Lab members would keep products at their location (or somewhere other than Bioinformatics.Org) and take part in the Bioinformatics.Org Grid. This is a Grid Technology[15] project, creating a globally distributed computer network for bioinformatics research. Members may optionally co-locate products at Bioinformatics.Org facilities for participation in the Grid.

And Bioinformatics.Org is seeking Co-Lab members on three levels:

Assistant Level The member chooses either product co-location or participation in the Bioinformatics.Org Grid. Bioinformatics.Org will perform systems administration while the member maintains a service contract.

Associate Level In addition to the aforementioned, the member chooses one person to be an honorary member of the Bioinformatics.Org Board of Trustees, and the member contributes to the cost of the facilities.

Principal Level In addition to the aforementioned, the member donates the primary computing hardware and software to run the Bioinformatics.Org website and gets a ”Powered By” logo on each page. And, the member chooses one person to be a principal member (instead of an honorary member) of the Board of Trustees. There is only one Principal Level member.

2.5.2 Benefits of collaboration

There are several benefits to joining the Co-Lab program:

Product Exposure Scientists using the resources would essentially be test-driving research systems, which would influence future purchasing decisions. *More than 4,000 bioinformaticists world-wide* would thus have exposure to your products.

[14] www.osdl.org

[15] For information on Grid Technology, visit escience-grid.org.uk

Platform Development Software development projects in The Open Lab would have access to multiple hardware platforms for porting. Thus, potentially, dozens of Free and Open Source bioinformatic libraries and applications could become available for these. This is the primary goal of OSDL, of which IBM, Hewlett Packard, Intel, Red Hat and others are members.

For example, there are currently more than 30 software projects at Bioinformatics.Org that run in the UNIX environment. They range from a gene finding accuracy evaluation tool (AcE), to a tandem repeat occurrence locator (TROLL). Even more projects, not yet hosted with the Organization, may choose to relocate because of the availability of certain platforms.

Low Cost You would retain ownership of your hardware, and Bioinformatics.Org would perform systems administration. The Organization is not asking for a donation of hardware but for assistance in developing a resource that would be of great benefit to both the bioinformatics community and to your business.

Publicity With a membership of more than 4,000 people, Bioinformatics.Org is one of the largest bioinformatic organizations in the world and currently more than twice the size of the International Society of Computational Biologists (ISCB)[16]. Bioinformatics.Org would publicly acknowledge the relationship on the website (including graphics where appropriate), in a press release, and in newsletters.

Public Relations The Organization also represents the Open Source bioinformatics community. Bioinformatics.Org will grant use of its name and graphics for press releases and publications.

2.5.3 Participants

Current Co-Lab members include

Apple Computer, Inc.
Cupertino, California

Xinit Systems[17]
London, UK

[16] www.iscb.org
[17] A formal agreement has not yet been finalized.

GreenTea Technologies, Inc.[17]
Mountain View, California

2.6 Practicing freedom and openness

All information, processes and educational materials at Bioinformatics.Org are free and open. The project management system is itself in CVS and able to be downloaded by the public.

Without the use freely available software, the Organization may never have begun. The commercial equivalences of what we use would have cost tens of thousands of dollars. The generosity of those who have made these resources freely available have inspired us to make our own resources such, contributing back to the community. This is in fact how the community grows.

2.7 Website activity

Webserver activity shows that more than *one million* files are downloaded from the server *every month*. And there are approximately 500 unique visitors to the front page every day.

2.8 Accolades

Bioinformatics.Org is either mentioned or featured in the following publications:

Linux Magazine June 2002

Salon.com February 25, 2002

Science September 21, 2001

Genetic Engineering News September 15, 2001[18]

The Organization has received messages from members and visitors expressing gratitude for our services. Below appears one such testimonial.

[18]The Organization subsequently appears on Genetic Engineering News's list of the 100 "most influential and important biotechnology websites" for 2001.

July 24, 2002

Wow! This is the best thing I've seen in years! I was collecting free online-ware to feature on a new section of GeneBrowser.com when I stumbled upon Bioinformatics.Org and was astounded at the SMS software, which more than duplicated a day's worth of searching the entire web (for online-ware). I can't tell you how useful it is to have all of this in one place. My lab has stumbled with various licensing restrictions for over a year, at a cost of more than $4,000, just to have one user, one copy, one computer, non-accessible by other lab members. Usually, it is not possible to use it at all because of the license restrictions. We recently had two months of downtime because of incompatibilities, bugs, and licensing restrictions.

Clague Hodgson, Ph.D.
Scientific Director
Nature Technology Corporation

3 BIOINFORMATICS.ORG FUTURE

3.1 This year

Things to accomplish this year include incorporating, formation of the Board of Trustees, formalized executive positions, and obtaining non-profit status. Donations to Bioinformatics.Org will be tax-free and tax-deductible. The Organization is also seeking external funding via grants and partnerships.

3.2 1 to 5 years

3.2.1 Open labs and structure changes

Figure 5 shows the new structure of the Organization. The major change is the inclusion of “open labs.” Open labs are divisions of The Open Lab into subspecialties.

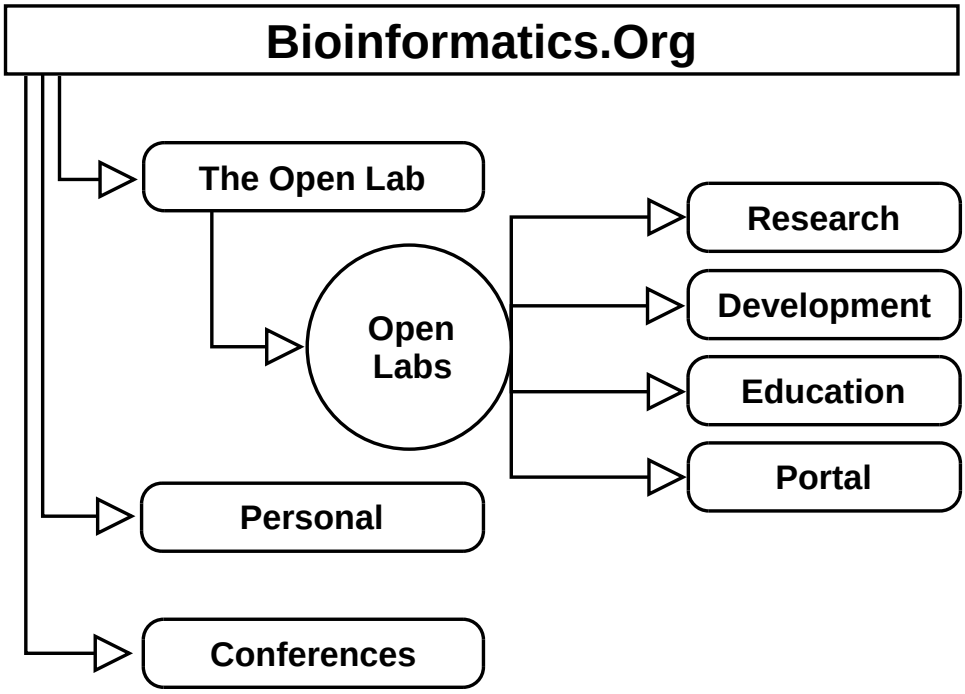

Figure 5: New structure of the Organization

Below are some of the established or nascent open labs.

BioClusters
: Computing cluster (e.g., Beowulf) topics

BioDatabases
: Database topics

BioDevelopers
: Software development topics

Other, possible open labs are listed below.

BioResearchers
Research topics

BioEducators
Education topics

BioProtocols
Protocols, standards, ontologies, etc.

BioDownloads
Software registry and repository portal

BioLinks
Website links

BioEthics
Topics in ethics

BioEvents
Events and calendar portal

BioArt
Free graphics for biology

Note that each of the open labs will have portal aspects, and each will cover research, development and education where applicable and to varying degrees. The change in structure will therefore place portals under The Open Lab.

3.2.2 Improving research resources

More computing resources will be sought and of various platforms. This includes clusters and supercomputers. The project management system will be modified for research and education projects. And, research resources will be accessible via the Bioinformatics.Org Grid.

3.2.3 Toward in-house research

It is a goal that Bioinformatics.Org will someday be a research institute with in-house research projects. We would like the research to be “open access.” That is, research projects will be conducted akin to Free and Open Source Software development projects, where people will be able to view, comment on, and contribute to active, pre-publication projects.

3.2.4 Improving educational resources

Bioinformatics.Org would also like to improve the available educational resources. We would like to offer online courses, tutorials and primers in topics covered by the open labs. We have also planned on developing an integrated glossary, where bioinformatic terms on the website will be dynamically hyperlinked to their glossary definitions.

3.2.5 Estimated budgets

The projected, estimated budgets for the next three years are in Table 2. The budgets are blue-sky and depend upon the acquisition of external grants and the generosity of sponsors.

Table 2: Estimated Budgets 2003-2005

Year	2003	2004	2005
	US $	US $	US $
New Hardware	5,000	10,000	20,000
Hardware Repair	1,000	2,000	4,000
Personnel	0	20,000	40,000
Facilities	0	0	0
Travel	1,000	3,000	5,000
Promotion	1,000	3,000	5,000
Totals	8,000	38,000	74,000

None of the items, other than facilities, would be crucial to fund. Since we are beginning with $0, we *can* continue that way almost indefinitely. Computer hardware could remain at a few servers, using housing, Internet access and electricity

at a university, as is our current situation. The cost of facilities, however, would have to be maintained and at a comparitively high cost. In order to grow, it would therefore be necessary to have no facility expenses for several more years.

3.2.6 Required facilities

The Organization will need facilities beyond what it currently has at the University of Massachusetts Lowell. The immediate need is mostly for data center space. Figure 6 shows the floorplan for a small data center with workstation space. This will likely suit the needs of the Organization for a couple years (until 2005).

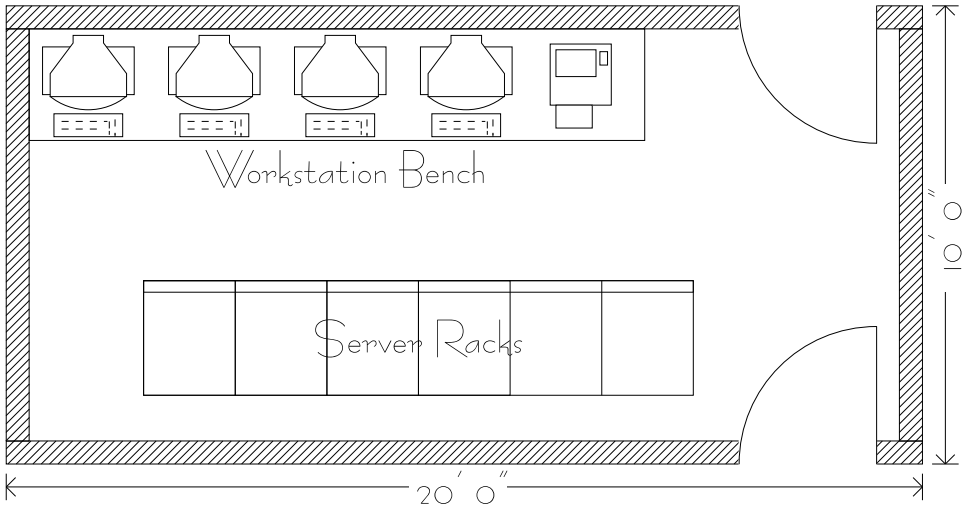

Figure 6: Small data center

By the third year, the Organization should have several personnel (employees) and additional computer hardware. That would require a separation of the workstation space and the data center, mostly due to the noise generated by a large amount of hardware.

Figures 7 and 8 show small and large “war rooms,” a popular design for groups working on a single project.

3.2.7 Sources of funding

In order to grow, the Organization must be externally funded. General sources of funding have been identified and listed below.

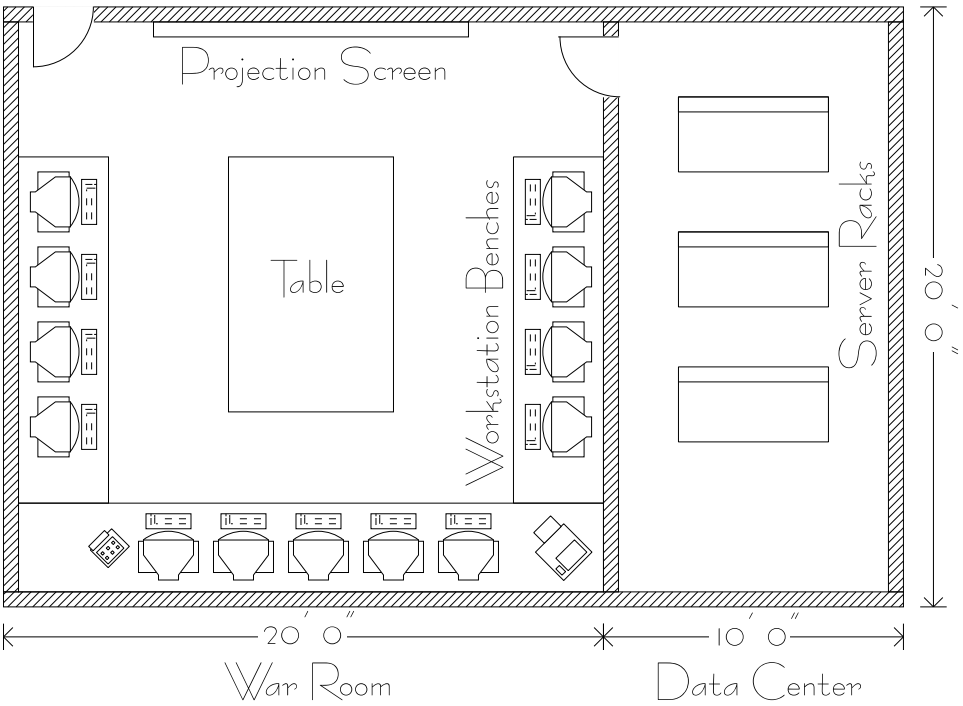

Figure 7: Small war room and data center

- Member donations
- Private foundation grants
- Business partnerships (organizational memberships)
- Business donations (organizational memberships)
- US Government grants

US Government grants will likely be in the long term, as a significant administrative infrastructure will be expected.

3.3 5 to 10 years

A long term goal for the Organization is to have a total annual operating budget of about US $1 million, which would be enough to pay for all budget items listed, including facilities and several employees. We would like to reach that goal within ten years (by 2012).

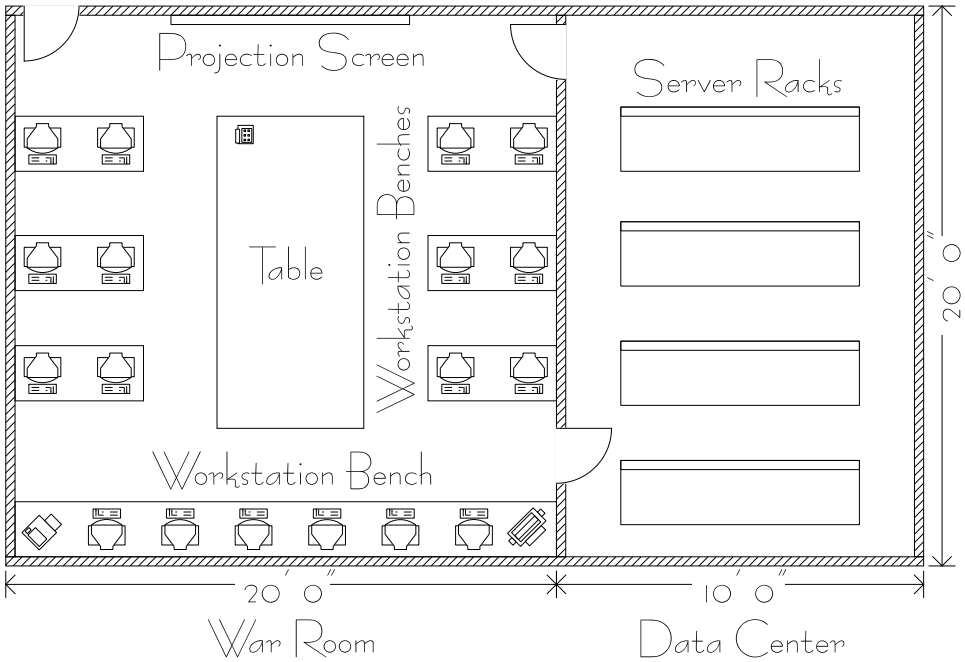

Figure 8: Large war room and data center

3.3.1 Required facilities

In 5 to 10 years (2007 to 2012), the Organization may have enough personel, possibly including some visiting researchers or “faculty,” to need some office space. Figure 9 shows a floorplan for adding offices to the war room design mentioned previously. This design is not meant to be specific.

3.4 10 years and more

The ultimate goal of the Organization is to become an established and highly regarded institute with substantial facilities and faculty. Bioinformatics.Org may be called at that time “The Bioinformatics Organization,” have possibly 30,000 square feet of space that it owns outright and several dozen personnel.

The Organization should, however, always maintain its Internet presence and keep to its mission.

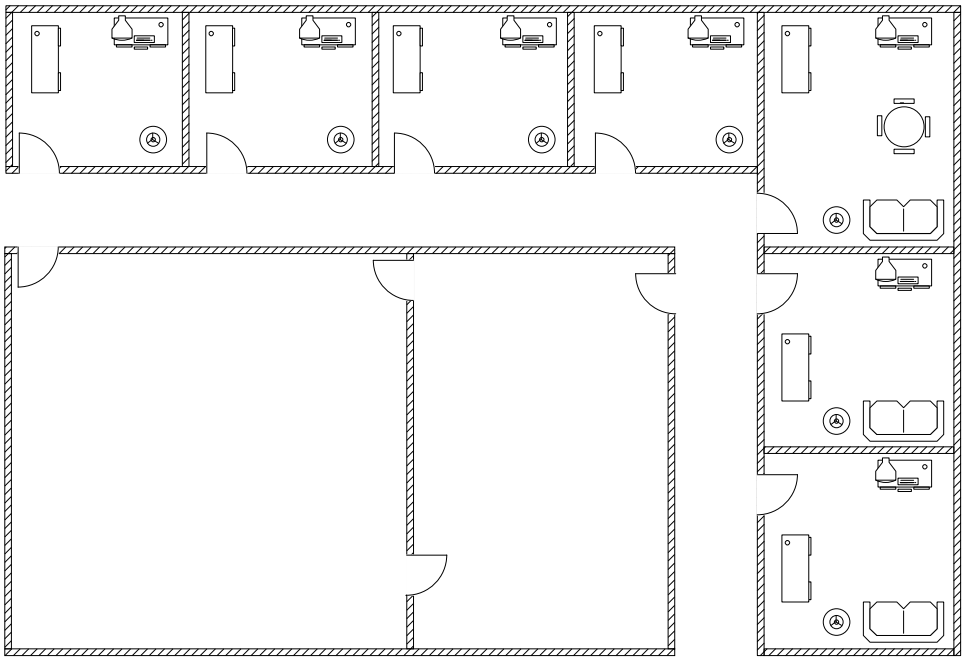

Figure 9: War room, data center and offices