G E N I

Global Environment for Network Innovations

GENI End-User Opt-In Overview

Document ID: GENI-SE-OI-01.4

July 15, 2009

Prepared by:

The GENI Project Office

BBN Technologies

10 Moulton Street

Cambridge, MA 02138 USA

TABLE OF CONTENTS

1   Document Scope

This section describes this document's purpose, its context within the overall GENI document tree, the set of related documents, and this document's revision history.

1.1   Purpose of this Document

This document provides summary of the GENI control framework requirements.

It is a DRAFT, to be used for discussion in the GENI Facility Control Framework working group.

Once approved, it can be used as a guide to judge the completeness of prototype control framework designs, and as a guide to their continued evolution.

1.2   Context for this Document

Figure 1-1. below shows the context for this document within GENI's overall document tree.
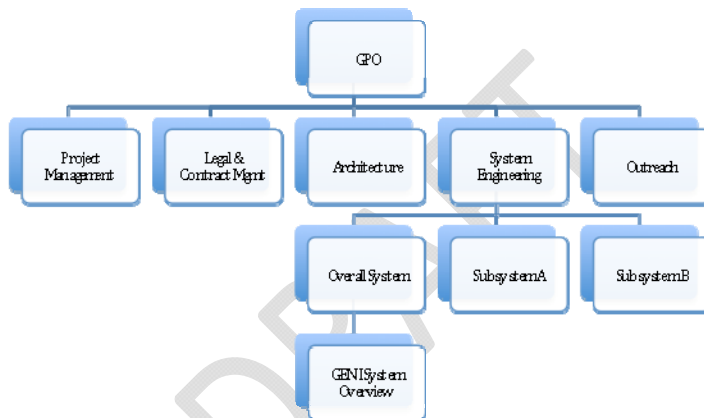
Figure 1-1. This Document within the GENI Document Tree.

1.3   Related Documents

The following documents of exact date listed are related to this document, and provide background information, requirements, etc., that are important for this document.

1.3.1   National Science Foundation (NSF) Documents

Document ID	Document Title and Issue Date
N / A	

1.3.2   GENI Documents

Document ID	Document Title and Issue Date
GENI-SE-SY-RQ-01.9	GENI System Requirements (DRAFT) January 16, 2009 http://groups.geni.net/geni/wiki/SysReqDoc
GENI-SE-SY-SO-02.0	GENI System Overview September 29, 2008 http://www.geni.net/docs/GENISysOvrvw092908.pdf
GENI-INF-PRO-S1-OV-1.12	GENI Spiral 1 Overview September 29, 2009 http://www.geni.net/docs/GENIS1Ovrvw092908.pdf
GENI-SE-CF-RQ-01.3	GENI Control Framework Requirements (DRAFT) January 13, 2009 http://groups.geni.net/geni/wiki/GeniControlFrameworkRequirements
GENI-SE-CF-PLGO-01.2	PlanetLab GENI Control Framework Overview (DRAFT) January 14, 2009 http://groups.geni.net/geni/wiki/PlanetLabGeniControlFrameworkOverview
GENI-SE-CF-PRGO-01.4	ProtoGENI Control Framework Overview (DRAFT) March 25, 2009 http://groups.geni.net/geni/wiki/ProtoGeniControlFrameworkOverview
GENI-SE-CF-ORGO-01.2	ORCA GENI Control Framework Overview (DRAFT) January 14, 2009 http://groups.geni.net/geni/wiki/OrcaGeniControlFrameworkOverview
GENI-SE-SY-TS-UC-LC-01.1	Lifecycle of a GENI Experiment (DRAFT) March 4, 2009 http://groups.geni.net/geni/wiki/ExperimentLifecycleDocument
GENI-SE-SY-STRO-01.1	GENI Structure Overview (DRAFT) June 15, 2009

1.3.3   Standards Documents

Document ID	Document Title and Issue Date
N / A	

1.3.4   Other Documents

Document ID	Document Title and Issue Date

1.4   Document Revision History

The following table provides the revision history for this document, summarizing the date at which it was revised, who revised it, and a brief summary of the changes. This list is maintained in reverse chronological order so the newest revision comes first in the list.

Revision	Date	Revised By	Summary of Changes
-01.1	4/10/09	H. Mussman	Early draft, for use in WG.
-01.2	4/21/09	H. Mussman	Rewrite, to include input from WG meeting on 4/14/09.
-01.3	5/9/09	H. Mussman	Revise and shorten.
-01.4	7/15/09	H. Mussman	Modify after separate GENI Structure Overview document was written.

2   GENI Entities, Actors and Agreements

See "GENI Structure Overview" document.

2.1   Aggregates that Provide Resources

A GENI suite has these entities that provide resources for experiments within the GENI suite:

- Multiple aggregates (or aggregate components), which are physical entities that provide resources to authorized users for experiments.

Associated with each aggregate are these actors:

- A management authority that is responsible for the management of the aggregate, and which can delegate selected functions to other actors. (Should this be renamed "aggregate authority"?)
- An aggregate administrator is an actor who has been delegated the responsibility for assigning the resources provided by that aggregate.
- An aggregate operator is an actor who has been delegated the responsibility for maintaining the operational readiness of that aggregate.
- The management authority can enter into agreements for the aggregate.
- An aggregate may often be provided by a research institution, e.g., a research group at a university; then, that research group becomes the "management authority", perhaps in combination with other university officials and/or groups.

2.2   Slices that Utilize Resources

A GENI suite has these entities that provide utilize resources for experiments within the GENI suite:

- Multiple slices (abstract entities) that "contain" resources to support experiments (abstract entities) within the GENI suite.
- Each slice contains from zero to many resources (named "slivers"), obtained from zero to many aggregates. See Figure 2-1.
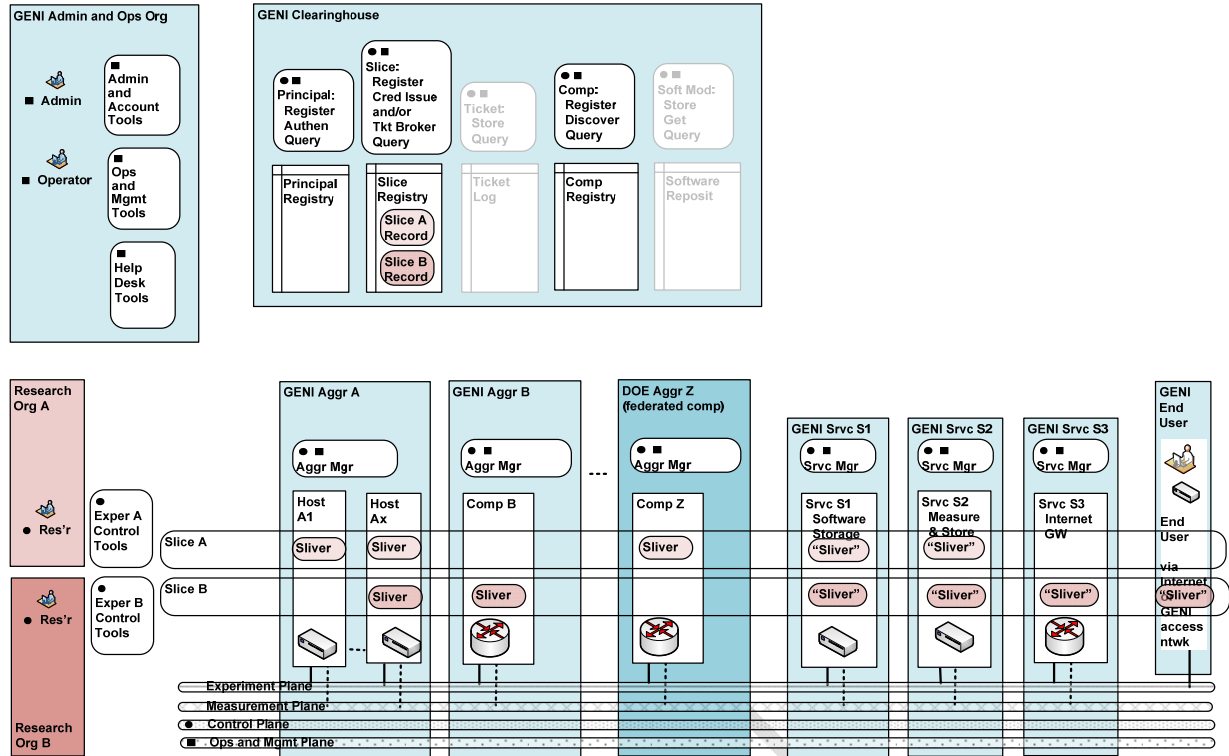
Figure 2-1. Two GENI Slices.

Associated with each slice are these actors:

- A slice authority that is responsible for the management of the slice, and the experiments conducted with the resources assigned to that slice, and which can delegate selected functions to other actors.

- A slice administrator is an actor who has been delegated the responsibility for getting resources assigned to a slice.

- A PI is an actor who has been delegated the responsibility for allowing certain researchers to request that resources be assigned to the slice, and then utilize the resources that have been assigned to the slice.

- A researcher is an actor who has been allowed to request that resources be assigned to the slice, and then utilize the resources that have been assigned to the slice to conduct an experiment.

- The slice authority can enter into agreements for the slice.

- A slice is almost always utilized by researchers at a research institution, e.g., a research group at a university; then that research group becomes the "slice authority" , perhaps in combination with other university officials and/or groups.

It is expected that most research institutions will both provide and utilize resources, and thus they will have both a management and a slice authority. This can be considered the "reference case", with exceptions for research institutions that only provide or utilize resources, but not both.

2.3   Necessary Agreements

An underlying set of agreements is necessary between these principal actors in a GENI suite, if the GENI suite is to provide useful service to all of the associated research institutions:

- The management authorities of all aggregates, that are responsible for the management of all aggregates and for maintaining the operational readiness of those aggregates.

- A slice authorities of all slices, that are responsible for the management of all slices, and the experiments conducted with the resources assigned to those slices.

These agreements must include:

- A resource usage agreement, that would typically allow resources from any (or all) of the aggregates to be used in any (or all) of the slices, and the rules (or policies) for assigning resources, particularly any exceptions.

- Cooperative operations agreements.

- Acceptable use agreements, and other agreements on best practices for managing the suite and conducting experiments.

3   Opt-In Users

GENI opt-in users are those users who participate or "opt-in" to a GENI experiment, and become part of the slice.

A GENI suite may optionally include these "opt-in" users:

- Opt-In users who choose to "opt-in" to a GENI experiment within the GENI suite.
- Opt-In users who are pulled into a GENI experiment within the GENI suite as a group, e.g., because a site decides to participate in an experiment or service.
- 3rd-party users, who are outside the GENI suite, but may be undesirably affected by a GENI experiment from within the GENI suite (this may be considered a case of "unintended" opt-in by such users)

The number of opt-in end users may easily exceed the number of research users. Unlike research users, the opt-in users may not be individually registered and authenticated in the GENI clearinghouse or aggregates (probably not in most cases), so experiments will need to provide recording, tracking, and security/privacy functions for their opt-in users when they operate on GENI.

These users may access GENI through the general-purpose Internet, or through a native GENI network (for example at a university LAN that is part of a GENI project).

Some users may choose to opt-in to a GENI experiment to utilize an application or service that is hosted in a GENI experiment.

Some users may choose to opt-in to a GENI experiment and make their own resources available to others.

Sometimes a site participates in an experiment or service, and its users are brought along as a group; individual users should, however, be able to opt-out. Such users may have no understanding of GENI, even though they are running an application or service that takes advantage of GENI resources.

Also, other "innocent" users may have their traffic routed to or through a GENI application or service. This would be done to include real-world users and traffic to make experimental results relevant to real-world networks.

3rd-party users are those users who are outside the GENI suite, but may be undesirably affected by a GENI experiment from within the GENI suite. When this happens, it may be considered a case of "unintended" opt-in by such users.

4   Opt-In Goals

A principal GENI goal is to attract and retain users who choose to participate or "opt-in" to a GENI experiment, and become part of the slice.

Or, a site decides to participate in an experiment or service and its uses are pulled into a GENI experiment within the GENI suite as a group.

Or, the traffic of "innocent" users is routed through a GENI experiment.

 See GENI System Requirements at http://groups.geni.net/geni/wiki/SysReqDoc .

See early GENI documents.

In all cases, this will add realism to the GENI experiment and make the experimental results relevant to real-world networks.

Sometimes, it will facilitate an experiment that cannot otherwise be accomplished. For example, analyzing the flows of real traffic is difficult to do in today's research environment.

Furthermore, it is an essential step in building a large-scale long-running GENI experiment with real users. In turn, this is an important step towards launching a new network structure or service.

However, there are responsibilities and risks involved with opt-in. The responsibilities must be understood and met. The risks must be understood, accepted and minimized.

Risks for opt-in users include service quality, service continuity, security and privacy. The opt-in users must understand these risks, and typically accept them through "informed consent". The researchers must minimize these risks to be able to attract and retain opt-in users.

There are legal requirements for the privacy of opt-in users that must be met by the researchers. In addition, an experiment may utilize opt-in users as "human subjects", and then the researchers must have the experiment approved by their Institutional Review Board (IRB).

In all cases, the researchers must minimize the risk of disruption to 3rd-party users, and deal with any disruption in an efficient manner.

One way to meet responsibilities and minimize risks is to for all actors to agree on a clear set of best practices, and then follow them.

5   Opt-In Use Cases

5.1   Overview

To date, these opt-in use cases have been identified:

- Use Case 1: A user chooses to participate in a GENI experiment to receive a service
- Use Case 2: A user chooses to participate in a GENI experiment to contribute resources for others
- Use Case 3: A group of users are pulled into a GENI experiment
- Use Case 4: A 3rd-party user is undesirably affected by a disruptive GENI experiment

5.2   Use Case 1: A User Chooses to Participate in a GENI Experiment to Receive a Service

5.2.1   Description

A researcher sets up an experiment on GENI to provide a service to users and/or others.

Individual users learn of the experiment, and make a choice to opt-in to the experiment for its service.

If they fully understand the nature of the experiment and the service, they have provided informed consent when they make the choice to join the experiment.

An experiment may include a separate content provider, perhaps from another experiment.

5.2.2   Implementations

Google Beta services are examples of large scale, commercial applications of this use case. All users make a choice to opt-in after being attracted by useful and "cool" experiments.

Another example is Many Eyes, a long-running visualization service provided by IBM researchers. See http://services.alphaworks.ibm.com/ManyEyes/ . Information on the Many Eyes visualization service is being sought from Wendy Kellogg and/or Martin Wattenberg – IBM.   It is one of many applications hosted on the IBM alpha Works platform http://www.alphaworks.ibm.com/

Larry Peterson reports in http://www.cs.princeton.edu/~llp/policy.pdf that PlanetLab currently supports many services provided by researchers that attract many opt-in users: "Researchers also deploy continuously running network services that attract a user community. This is generally a symbiotic relationship: users gain some benefit by using the service (e.g., faster downloads) and researchers learn how the service behaves under real workloads. To give a sense of the magnitude of such services, PlanetLab typically carries 6TB of Internet-bound traffic on behalf of 1 million users each day.

Content Distribution Networks (CDNs) are the most successful example of continuously running services on PlanetLab. These CDNs effectively interpose themselves between clients that access content and servers that provide content."

GENI projects: currently none

5.2.3   Issues

A key issue is how to attract and then retain opt-in users.

A one-day workshop was held on July 22, 2008 in Cambridge, MA, that was structured to solicit ideas from experts interested in technology, and its impact on society. White papers were submitted by participants in the workshop:
http://groups.geni.net/geni/attachment/wiki/GeniOptIn/White%20paper%20from%20cycle%201%20participants.pdf

Reports from the workshop are at
http://groups.geni.net/geni/attachment/wiki/presentations/Partridge-Opt-In-Workshop-v1.0.ppt and at http://groups.geni.net/geni/attachment/wiki/GeniOptIn/Opt-In%20Workshop%20Report-v21.5.pdf .

One conclusion was that the way to attract end-users is (mostly) through getting great applications into GENI slices. Thus, we need to think about how to ensure GENI is application-friendly.

Also, it may be helpful to have a "marketplace" to match network researchers with application researchers, because these are likely to be distinct.

Various approaches were identified to attract and retain valuable "emergent users".

Issue: When users make a choice to opt-in, what is the mechanism used to make the choice? Typically a web interface? If there is no web interface as part of the service, what can be done?

Issue: When users make a choice to opt-in, how can we insure that we have informed consent? What information must be provided? Is there a way to standardize on the approach to assure that it is done properly every time? Is there a standardized software package that can be utilized?

Issue: Privacy of individual opt-in users is essential, to meet both ethical and legal requirements.

5.3   Use Case 2: A User Chooses to Participate in a GENI Experiment to Contribute Resources for Others

5.3.1   Description

A researcher sets up an experiment on GENI to provide a service to allow users to share resources with the researcher and/or others.

Individual users learn of the experiment, and make a choice to opt-in to the experiment and make their own resources available to others, such as capacity on a PC or a mobile node. They may have to download code to their node, and run it, etc.

The user should fully understand the nature of the experiment, thus have provided informed consent.

In addition, the user may gain access to a resource or service in the GENI experiment.

5.3.2   Implementations

A long-running implementation of this use case is the SETI@home project with over 3 million users. See http://setiathome.ssl.berkeley.edu/

GENI project: 1645 Million Node GENI project (Justin Cappos – U Washington). See
http://groups.geni.net/geni/wiki/MillionNodeGENI

This project is currently focusing on a platform for educational cloud computing called Seattle.

See SIGCSE 2009 paper on Seattle: A Platform for Educational Cloud Computing at
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/Cappos-SIGCSE2009.pdf

See slides on Seattle from GEC3 are at
http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203a%20%20seattle.ppt
x

A live demo of a user contributing a resource to Seattle was presented at GEC4. See
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/Seattle%20--%20GEC%204.pdf

This wiki page has a very pure view of Seattle (no Seattle GENI).
https://seattle.cs.washington.edu/wiki/SeattleComponents

The node manager API also has a pure view (no Seattle GENI).
https://seattle.cs.washington.edu/wiki/NodeManagerDesign

GENI project: 1646 HomeNet Project (David Anderson – CMU) . This includes "Prototype
Support for Heterogenous Testbed Resources: Integrating cluster, broadband, and wireless emulation
nodes into the "Proto- GENI" Framework". See http://groups.geni.net/geni/wiki/CmuLab

5.3.3 Issues

Issue: Contributing resources entails significant risks for the user since they are typically running
code on their own PC or node that has been provided by the researcher and/or other users.   These
include stability, security and privacy risks. Every effort must be made to mitigate these risks.

Issue: This user must fully understand the nature of the experiment, thus have provided informed
consent.

Issue: If others are going to using the shared resources, can the user who is contributing the
resources limit what can be done with them? Perhaps just agree to a policy before they are contributed?

5.4   Use Case 3: A Group of Users are Pulled into a GENI Experiment

5.4.1   Description

A researcher sets up an experiment on GENI to provide a service to a group of users.

The experiment may include a separate content provider, perhaps from another experiment.

In some cases, a site decides to participate in the experiment or service, and its users are pulled into
the experiment as a group. Note that such users may have little or no understanding of GENI, even
though they are running an application or service that takes advantage of GENI resources. Typically,
individual users are informed of the nature of the experiment and its service, and are then able to "opt-
out".

Another option is to have a researcher sets up an experiment on GENI to analyze the network traffic
of "innocent users" that has been rerouted through the experiment from a network entity, such as an
ISP.   In such cases, it is typically not possible for them to make a decision to "opt-out". If this can be
done on a large scale, it can provide information to researchers that they are unable to gather today.

5.4.2   Implementations

Larry Peterson reported that PlanetLab has hosted services where a site participated in the service, and its users were brought along en masse.

Larry Peterson reports in http://www.cs.princeton.edu/~llp/policy.pdf that PlanetLab currently supports many services provided by researchers that attract many opt-in users: "Researchers also deploy continuously running network services that attract a user community. This is generally a symbiotic relationship: users gain some benefit by using the service (e.g., faster downloads) and researchers learn how the service behaves under real workloads. To give a sense of the magnitude of such services, PlanetLab typically carries 6TB of Internet-bound traffic on behalf of 1 million users each day.

Content Distribution Networks (CDNs) are the most successful example of continuously running services on PlanetLab. These CDNs effectively interpose themselves between clients that access content and servers that provide content."

Henning Schultzrinne reported that including network traffic from a group of users has been done on a small scale at IETF, SIGCOMM, and Internet2 meetings. The attendees are given the choice to login to a different wireless network, knowing that their traffic will be analyzed.

David Kotz reported that the Dartmouth Internet Security Testbed http://www.cs.dartmouth.edu/~dist/ is being setup to capture and analyze traffic from "innocent users" without informed consent. See http://groups.geni.net/geni/wiki/CRAWDAD_Experience for an explanation of the approval process that was necessary.

GENI Projects: 1650 Regional Opt-In project (Matt Mathis – PSC)

The Regional Opt-In at http://groups.geni.net/geni/wiki/RegionalOptIn is currently developing techniques that can be used for wholesale opt-in of large populations of innocent users. This project was reviewed at GEC3 http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203c%20%20Mathis-regOPT.pdf and GEC4 http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/OptInGEC4.pdf

 The Opt-In Requirements document at   http://groups.geni.net/geni/wiki/OptInReqs and http://groups.geni.net/geni/attachment/wiki/OptInReqs/OptInReqs.pdf outlines requirements for a strong wholesale opt-in mechanism for GENI. When fully deployed it would permit GENI experimenters to request that ISPs redirect traffic from a huge population of innocent users through GENI infrastructure. These users are innocent in the sense that they do not have to do anything at all to participate, and might not even be aware that they are doing so. Key to wholesale opt-in is that it fully engages the Institutional Review Board (IRB) process and that all participants are motivated by their own self interests to do the right thing.

5.4.3   Issues

Issue: If group opt-in is arranged by a service provider, how do we provide for informed consent to "opt-out"?

Issue: Can a group "innocent" users be included in an experiment without their agreement?   If so, what process must be followed? What if they are customers of an ISP?

Issue: How can we provide for "fast restoration" when the experiment fails, or must be brought down?

Issue: How do we deal with a user whoe has been pulled into an experiment, notices the difference, feels there has been a security breach, and complains?

5.5   Use Case 4: A 3rd Party User is Undesirably Affected by a Disruptive GENI Experiment

5.5.1   Description

An experiment on GENI that can affect and disrupt 3rd-party users and/or others. This is an unintentional "opt-in" of the disrupted user.

This can be caused by any experiment in GENI, including use cases 1, 2 or 3.

Often the researcher did not anticipate this result.

Often, network experiments involving the Internet probe a user's machine, and they consider this to be malicious.

Occasionally, a malicious researcher could knowingly cause a disruption.

The disrupted 3rd-party users can be within GENI.

The disrupted 3rd-party users may outside of GENI, reached via the Internet or another network.

Such disruption must be detected, and then contained or stopped.

There may be a direct complaint from a disrupted 3rd-party user, and it must be possible to trace back the disruptive traffic, through the aggregate that handled it, to the experiment that caused it.

The disruption may be detected within GENI before it disturbs a 3rd party user.

5.5.2   Implementations

Larry Peterson reported a wealth of PlanetLab experience with this use case.

See Understanding and Resolving Conflicts in PlanetLab, by Larry Peterson - Princeton Univ, at

http://www.cs.princeton.edu/~llp/policy.pdf and slides from GEC3 at http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%202a%20%20llp_policy.ppt

5.5.3   Issues

Issue: Need capability to detect disruptive traffic where possible.

Issue: Need capability to trace back the disruptive traffic, through the aggregate that handled it, to the experiment that caused it.

Issue: Need capability to stop disruptive experiment.

Issue: Experiments that probe the Internet (or another network) should have capability to implement a "blocked call list" so that probing of a user's machine can be blocked.

Issue: Should minimize overhead on operators.

Issue: Are there ways the strategy Planet Lab adopts is not applicable to GENI?

6   Opt-In Mechanisms

6.1   Choice

On the "attract users" side, we need to provide some mechanisms to choose to join an experiment:

- Explicitly "point your browser at" an experiment.
- Run "opt-in" software (e.g., OpenVPN, or in general, client-side software you have downloaded).
- Transparently high-jack certain classes of traffic (e.g., using OpenFlow to divert certain traffic).

6.2   Connectivity

Opt-in users may access a GENI experiment via:

- The general-purpose Internet
- Through a native GENI network (for example at a university LAN that is part of a GENI project)

Opt-in connectivity mechanisms include:

- Physical layer connectivity – e.g., wireless
- Network layer connectivity – e.g., destination, host
- Application layer connectivity – e.g., web, cache, VoIP

Many opt-in users will access a GENI experiment via the Internet:

- For network layer connectivity, an Internet Gateway to/from GENI will be required; see section xx.
- For application layer connectivity, an appropriate DNS entry will be required, that points to a valid public IP address, that reaches a GENI experiment via an internet gateway.
- In some cases, it may be possible to use the DNS redirect mechanism.

6.3   Internet Gateway to/from GENI

6.3.1   Description

Gateway between GENI and the Internet

(Need to define basic gateway functions.)

6.3.2   Implementations

GENI Project: 1601 Virtual Tunnels (Nick Feamster – Georgia Tech)   See GENI Project on Bringing Experimenters and External Connectivity to GENI at http://groups.geni.net/geni/wiki/DTunnels   Presentation from GEC3 at http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203b%20bgp-mux-gec3.ppt

GENI Project: 1650 Regional Opt-In (Matt Mathis – PSC) See GENI Project on Regional Opt-In at http://groups.geni.net/geni/wiki/RegionalOptIn Presentation from GEC3 at

http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203c%20%20Mathis-regOPT.pdf   Presentation from GEC4 at
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/OptInGEC4.pdf

   Opt-In Requirements doc at http://groups.geni.net/geni/wiki/OptInReqs and
http://groups.geni.net/geni/attachment/wiki/OptInReqs/OptInReqs.pdf This document outlines requirements for a strong wholesale opt-in mechanism for GENI. It would permit GENI experimenters to request that ISPs redirect traffic from a huge population of innocent users through GENI infrastructure.   It includes the requirements for the gateway that handles the traffic.

6.3.3   Issues

Issue: Tagging of traffic, so can trace trace back the disruptive traffic, through the aggregate that handled it, to the experiment that caused it.

Issue: Monitoring of traffic, to detect disruptive behavior.

Issue: Monitoring of traffic, to detect anomalous behavior.

Issue: Mechanisms for reacting to problems.

7   Opt-In Risks to Entities and Actors

4) User risks, third party impacts   Henning

7.1   Risks to Aggregates

Gather too many resources.

Disrupt other slivers.

7.2   Risks to Opt-in Users

Loss of privacy.

Loss of service.

Virus on contributed node.

User risks
-resources (bw, CPU, ...)
-disclosure
-disruption of exp. service
-legal exposure
-security
-interference with other services

7.3   Risks to 3rd-party Users

Disrupt them.

Third part impact
 -what (probing)
 -opt-out

8   Reference Single-Suite GENI Structure and Agreements

8.1   Structure

A straightforward single-suite structure for GENI is based on a two-level hierarchy (see Section 5.2), and includes a GENI suite authority. The GENI suite authority has agreements with all of the management and site authorities such that:

- All management authorities agree to follow a resource use policy that makes resources available to all slice authorities (all slices) in the GENI suite (perhaps with some exceptions).
- All actors agree on cooperative operation policies.
- All actors agree to follow a common set of best practices.

See Figure 7-1.
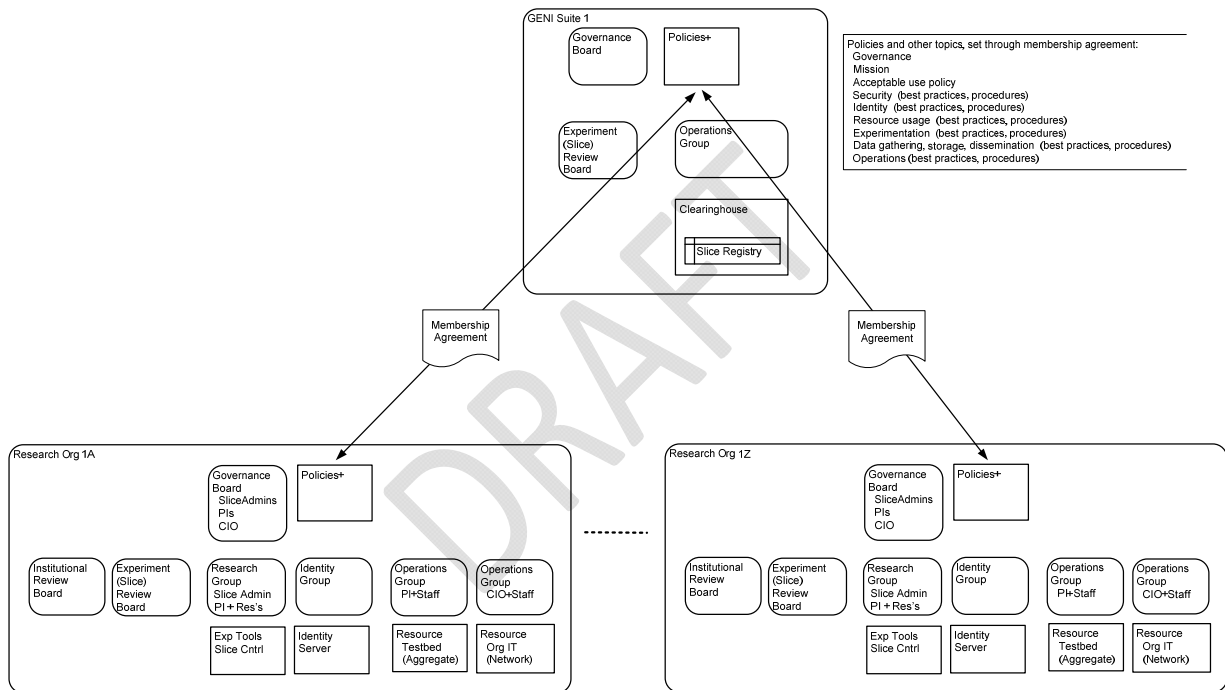
Figure 7-1. Reference Single-Suite GENI Strucutre.

This is the choice that has often been assumed in defining GENI. See the GENI System Overview document at http://www.geni.net/docs/GENISysOvrvw092908.pdf for more details.

When multiple suites are federated and/or affiliated, a more complicated set of agreements is required.

8.2   Entities and Actors

The GENI Suite Authority runs an important physical entity, the GENI Clearinghouse, which in turn contains a Slice Registry (and several other important functions).

In each of the research organizations, agreements should be made with a governance board that includes Slice Administrators, PIs and the CIO.

Several actors have been identified in each research organization, representing functions that will be required for efficient operation, including extensions for end-user opt-in.

These actors include:

- Operations group of CIO and staff.
- Operations group of PI and staff.
- Identity provider group. (new)
- Research group, including Slice Administrators, PIs and Researchers.
- Experiment review board. (new, suggested by PlanetLab experience, and important to screen opt-in experiments)
- Institutional review board. (new, driven by certain opt-in experiments)

8.3   Agreements

Agreements between the GENI Suite Authority governance board and the research organizations governance boards are expected to include:

- Governance
- Mission
- Acceptable use policy
- Security (best practices, procedures)
- Identity (best practices, procedures)
- Resource usage (best practices, procedures)
- Experimentation (best practices, procedures)
- Data gathering, storage, dissemination (best practices, procedures)
- Operations (best practices, procedures)

8.4   Example PlanetLab Agreement

Policies and other topics are set when a research organization joins the PlanetLab Consortium.

First and foremost, an organization must contribute resources (at least two nodes) before it can utilize resources.

And it must abide by an acceptable use policy.

See Understanding and Resolving Conflicts in PlanetLab, by Larry Peterson - Princeton Univ, at

http://www.cs.princeton.edu/~llp/policy.pdf and slides from GEC3 at http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%202a%20%20llp_policy.ppt

Furthermore, it must agree to common best practices for management. Here is a summary of what is expected:

Appendix: Framework for an Agreement

The following summarizes the expectations placed on researchers, hosting sites, and network testbed operators. We propose it as a framework for a community-wide agreement about managing wide-area network testbeds.

Researchers

• Follow established best practices. This includes first running experiments on a researcher's home site, and then ramping up slowly on the greater testbed.

• Be responsive. This includes monitoring experiments (and support mailing list), and not leaving experiments running on auto-pilot.

• Honor opt-out. This means maintaining blacklist addresses for third-party sites that object to experimental traffic.

• Keep any logs containing user-identifying information private. This also involves being responsive to legal requests to access these logs.

Hosting Sites

• Place nodes in a DMZ. Do not filter ports or ICMP ping packets.

• Forward complaints and concerns to the testbed operators.

• Do not disconnect nodes as this affects all researchers.

• If a compromise of the substrate is suspected, bring nodes into *safe* mode pending resolution.

Network Testbed Operators

• Actively manage the testbed. This includes keeping nodes up-to-date with the latest security patches and monitoring the tesbed for suspicious behavior.

• If a compromise of the substrate is supected, bring suspicious nodes into *safe* mode for further investigation.

• Be responsive to hosting site and third-party complaints.

• Suspend slices that do not adhere to best practices. Suspend users that consistently ignore best practices.

• Engage the research community in defining best practices for each type of experiment. Educate researchers as to these practices.

• Honor opt-out. This means maintaining blacklist addresses for third-party sites that object to experimental traffic.

• Allow hosting sites to opt-out of unwanted slices.

8.5   Agreements to Mitigate Opt-In Risks

Agreements between the GENI Suite Authority governance board and the research organizations governance boards are expected to include:

- Governance
- Mission
- Acceptable use policy

- Security (best practices, procedures)
- Identity (best practices, procedures)
- Resource usage (best practices, procedures)
- Experimentation (best practices, procedures)
- Data gathering, storage, dissemination (best practices, procedures)
- Operations (best practices, procedures)

9   Best Practices for Planning an Experiment

6) Experiment review process   Helen, Joe

9.1   Goals

List of documents

- Set of guidelines that can be handled to a graduate student running their experiments for the first time. Condensed form, much less than 50 pages.

- Best-practices document targeted towards a graduate student (2-3 pages)

Light weight to a full blown review process.

From self review to an institutional review board.

9.2   Experiment Review Process

The researcher should use this process when planning an experiment:

1. Researcher writes out experiment plan using a standard template.
2. Researcher checks experiment plan against established best practices for "setting up and running an experiment.
3. Researcher submits experiment plan for review by Suite Advisory Board, which considers both suite (network) issues and potential people issues.
4. If Suite Advisory Board judges that there may be one or more people issues, they forward plan for consideration by appropriate Institutional Review Board.

1st self examine

2nd advisory board, reviews network issues, looks for people issues.

3rd if people issues, passes to IRB. GENI specific IRB?

9.3   Provide Experiment Review Process Capability

10 Best Practices for Setting Up and Running an Experiment

On the "deal with consequences" side, we also need some mechanisms (some useful capabilities)

o audit mechanism + ability to suspend an experiment

o blacklists so we don't bother people that don't want to be bothered

o maybe there are useful and general purpose data archiving mechanisms

o a whole bunch of policy (best practices) statements (i.e., there are no mechanisms that solve the problem for us)

10.1 Utilize Slice Registry

Include experiment (slice) registry so that everything can be tracked.

10.2 Identify Slice Traffic Flows

Identify traffic leaving GENI with the slice that created it.

10.3 Audit Slice Resources

Provide audit capabilities for slice.

10.4 Provide Slice Shutdown Capability

Provide shutdown capabilities for a slice.

11 Best Practices for Enrolling Opt-in Users

5) How are users enrolled?   Joe, Helen

11.1 Provide Information for Opt-in Choice with "Informed Consent"

For example, here is a possible way for an experimenter and an end user to work together:

An experimenter wants to have opt-in participation on an individual basis. They publish a document that answers a set of standard questions so that participants (end users) can judge whether they want to play along and provide "informed consent". Questions might include:

What kind of experiment is this?

Who is responsible for the experiment?

What kind of data is being gathered by the experiment?

How long will the data be retained?

Will participation in this experiment affect other applications I'm running?

Will the experiment prevent my PC from going to sleep?

How much extra traffic will the experiment generate on my access link?

Should I run this application on a cell phone or other low-bandwidth device?

How are users enrolled?
1) User consent (elaborate)
2) Whole-sale opt-in
3) pinged

  -organizational agreement
-content providers
  -informed consent vs. consent

Consent
implicit
implied?
informed
check off
easy to find disclosure
recourse is opt out

"individualized and specific" are legal requirements

No PII

No IRB

2c) To donate resources (e.g., TORR)

Is this much different?

More risks

3) wholesale option w/o consent

Legality depends upon ISP contract with user

One plan: engage IRB, for case without informed consent Do a public notice, including a public notice on how to opt-out. Quick restoral.

PlanetLab has done this, capturing traffic to a web server from browsers

3b) who can do this?

Researcher – no, if federally funded research.

carrier System admin can do this at a carrier

How does Internet2 do this?

11.2 Provide Information for Opt-out Choice with "Informed Consent"

11.3 Always Allow Opt-out Choice

11.4 Provide Opt-in User Choice Capability

12 Best Practices for Gathering and Storing Experiment Data

3) Data retention   Aaron

12.1 Associate Experiment Data with a Slice and Manage Distribution

Data categories
-phone calls, email, URLs, packets etc
-essential vs. interested
 pairwise IP address

During experiment, it is useful to keep more data than is needed which might be helpful during later debugging.

12.2 Implementation and Operational Issues

Gathering logs and experiment data on GENI, and managing their distribution:

Issue: who can see what data?

Issue: isolation of experiments.

Issue: privacy of user data.

Issue: controlling access to data

Issue: assuring privacy of data.

Issue: assuring anonymity of data.

Data retention
 -PII
What is PII?
Listed by IRB?
gathered
 -stored
 -distributed
Destroyed

12.3 Identify and Manage User Identifiable Information

What is PII?
anonymization goodness

Is IP address, MAC address, length of transfer, time PII?

Deal with information in networking arena.

We are not talking about social security numbers or medical information which is more sensitive.

Helen: Get rid of PII term. Use Identifiable information only and personal is only a property.

12.4 Do Not Capture Content

Avoid capture of content!

12.5 Identify and Manage Identifiable Information in Network Logs

12.6 Current Implementations

GENI Projects

1628 Measurement System (Paul Barford – U Wisconsin)

13 Best Practices in Dealing with Disruption to 3rd-party Users

Third part impact

 -what (probing)

 -opt-out

13.1 Detect Traffic Flows from Disruptive Experiments

13.2 Provide for Shutdown of Disruptive Experiments

13.3 Manage Complaints from Disrupted 3rd-party Users

Detecting GENI experiments with disruptive behavior and managing them, including shutting them down.

Within GENI.

Outside of GENI, including disruption via the Internet.

See GENI System Requirements

See GENI Control Framework Requirements

13.4 Provide "Do-not-disturb" Lists of 3rd-party Users to Experiments

13.5 Current Implementations

Consider experience from PlanetLab by Larry Peterson – Princeton

GENI Projects

See PlanetLab GENI project

See ProtoGENI project.

14 References

1.1) End-User Opt-In WG Meeting at the GEC-1: Meeting minutes (October 2007) at http://groups.geni.net/geni/wiki/GeniOptinMinneapolis

1.2) End-User Opt-In presentation at the GEC-2: http://groups.geni.net/geni/attachment/wiki/GeniOptIn/geni-opt-in-0803.ppt

1.3) End-User Opt-In WG Meeting at the GEC-3: Meeting minutes, including all presentations (October 2008) at http://groups.geni.net/geni/wiki/OptInWGGEC3

1.4) End-User Opt-In WG Meeting at the GEC-4: Meeting minutes, including all presentations (April 2, 2009) at http://groups.geni.net/geni/wiki/GEC4OptInWGAgenda

1.5) GENI End-User Opt-In Scenarios and Capabilities Summary, by Harry Mussman - GPO at http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/102208b%20OptInWG_Scenarios_Capabilities_Issues_Documents.pdf

A one-day workshop was held on July 22, 2008 in Cambridge, MA, that was structured to solicit ideas from experts interested in technology, and its impact on society. Conclusions include: the way to attract end-users is (mostly) through getting great applications into GENI slices. Thus, we need to think about how to ensure GENI is application-friendly.

2.1) White papers were submitted by participants in the workshop : http://groups.geni.net/geni/attachment/wiki/GeniOptIn/White%20paper%20from%20cycle%201%20participants.pdf

2.2) Summary report from the GENI User Opt-IN Workshop (July 22, 2008) at http://groups.geni.net/geni/attachment/wiki/presentations/Partridge-Opt-In-Workshop-v1.0.ppt

2.3) Final report at http://groups.geni.net/geni/attachment/wiki/GeniOptIn/Opt-In%20Workshop%20Report-v21.5.pdf

2.4) Contributions on Many Eyes visualization service from Wendy Kellogg and/or Martin Wattenberg – IBM.

2.5) Many Eyes is a long-running visualization service http://services.alphaworks.ibm.com/ManyEyes/ that is hosted on the IBM alpha Works platform.

2.6) The IBM alpha Works platform http://www.alphaworks.ibm.com/   hosts Many Eyes and many other IBM services.

3.1) GENI Project on Regional Opt-In at http://groups.geni.net/geni/wiki/RegionalOptIn

3.2) Regional Opt-In presentation at GEC3 at
http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203c%20%20Mathis-regOPT.pdf

3.3) Regional Opt-In presentation at GEC4 at
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/OptInGEC4.pdf

3.4) Opt-In Requirements doc http://groups.geni.net/geni/wiki/OptInReqs and
http://groups.geni.net/geni/attachment/wiki/OptInReqs/OptInReqs.pdf

This document outlines requirements for a strong wholesale opt-in mechanism for GENI. When fully deployed it would permit GENI experimenters to request that ISPs redirect traffic from a huge population of innocent users through GENI infrastructure. These users are innocent in the sense that they do not have to do anything at all to participate, and might not even be aware that they are doing so. Key to wholesale opt-in is that it fully engages the Institutional Review Board (IRB) process and that all participants are motivated by their own self interests to do the right thing.

4.1) GENI Project on Million Node GENI at http://groups.geni.net/geni/wiki/MillionNodeGENI

4.2) SIGCSE 2009 paper on Seattle: A Platform for Educational Cloud Computing at
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/Cappos-SIGCSE2009.pdf

4.3) Slides on Seattle from GEC3 at
http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203a%20%20seattle.pptx

4.4) Justin reviewed million-node GENI project, also known as Seattle, and presented a live demo of a user contributing a resource to Seattle at GEC4. See
http://groups.geni.net/geni/attachment/wiki/GEC4OptInWGAgenda/Seattle%20--%20GEC%204.pdf

4.5) From Justin, here is some of the information we have that explains how Seattle works.   There are a few holes in this (notably how the SeattleGENI website provides resources to users), but I think this might be a good starting point.
This wiki page has a very pure view of Seattle (no Seattle GENI).
https://seattle.cs.washington.edu/wiki/SeattleComponents

4.6) The node manager API also has a pure view (no Seattle GENI).
https://seattle.cs.washington.edu/wiki/NodeManagerDesign

4.7) GENI Project by David Anderson, Patrick Gunn, on "Prototype Support for Heterogenous Testbed Resources: Integrating cluster, broadband, and wireless emulation nodes into the "Proto-GENI" Framework", at

http://groups.geni.net/geni/wiki/CmuLab

5.1) Contributions on Dartmouth Internet Security Testbed http://www.cs.dartmouth.edu/~dist/ and on CRAWDAD project http://crawdad.cs.dartmouth.edu/ , provided by David Kotz – Dartmouth College, and summarized at http://groups.geni.net/geni/wiki/CRAWDAD_Experience

5.2) Paper on "Legal Issues surrounding Monitoring During Network Research" by Paul Ohm, Douglas Sicker and Dirk Grunwald at http://portal.acm.org/citation.cfm?id=1298307

5.3) References on "Sharing Internet Measurements" by Mark Allman and/or Vern Paxson at ICSI, summarized at http://www.icir.org/mallman/research/proj-simr.html

5.4) Particularly, a talk at ACM SIGCOMM/USENIX Internet Measurement Conference, October 2007, on Issues and Etiquette Concerning Use of Shared Measurement Data. Copies of the paper are at http://www.icir.org/mallman/papers/etiquette-imc07.pdf .

5.5) See also the slides at http://www.icir.org/mallman/papers/etiquette-imc07-talk.pdf

 The slides are especially instructive to GENI opt-in policy.

5.6) Also, a related paper on The Devil and Packet Trace Anonymization at http://www.icir.org/mallman/papers/devil-ccr-jan06.pdf

5.7) Also, an earlier (2002) paper by Allman (when he was at BBN) that describes "A Scalable System for Sharing Internet Measurements", which can be found at http://www.icir.org/mallman/papers/ .   It seems like a straightforward approach to solving the general problem.

5.8) Aaron Burstein, "Issues Relating to Data Acquisition, Retention, Use and Disclosure" at http://groups.geni.net/geni/attachment/wiki/041409NYCOptInWGAgenda/051509%20%20burstein%20%20dataretention.pdf

6.1) Understanding and Resolving Conflicts in PlanetLab, by Larry Peterson - Princeton Univ, at

http://www.cs.princeton.edu/~llp/policy.pdf

6.2) Slides from GEC3 at http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%202a%20%20llp_policy.ppt

6.3) A description of the PlanetLab Consortium at http://www.planet-lab.org/consortium

7.1) GENI Project on Bringing Experimenters and External Connectivity to GENI at
http://groups.geni.net/geni/wiki/DTunnels

7.2) Presentation at GEC3 by Nick Feamstra, on "Bringing External Connectivity and Experiments to GENI", at

http://groups.geni.net/geni/attachment/wiki/presentations/OptIn%20WG%20%203b%20bgp-mux-gec3.ppt

8.1) GENI Project by Paul Barford, on "Instrumentation and Measurement for GENI", at

http://groups.geni.net/geni/wiki/MeasurementSystem

9.1) GENI Workshop on security, with slides from GEC4 at
http://groups.geni.net/geni/attachment/wiki/Gec4presentations/GEC%204%20-
%202.%20Matt%20Bishop%20-%20security%20wkshp.ppt

10.1) Aaron Burstein, "Toward a Culture of Cybersecurity Research" at
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1113014

10.2) Aaron Burstein. (2008). Conducting Cybersecurity Research Legally and Ethically. USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '08). Retrieved July 24, 2008, from http://www.usenix.org/event/leet08/tech/full_papers/burstein/burstein.pdf

10.3) Aaron Burstein. (2008). Amending the ECPA to Enable a Culture of Cybersecurity Research. Harvard Journal of Law & Technology, 22(1), 167--222. Retrieved October 14, 2008, from http://jolt.law.harvard.edu/articles/pdf/v22/22HarvJLTech167.pdf

10.4) Creative Commons provides free tools that let authors, scientists, artists, and educators easily mark their creative work with the freedoms they want it to carry: http://creativecommons.org/

15 Glossary

The current GENI glossary can be found at: **http://groups.geni.net/geni/wiki/GeniGlossary**