# Nomenclature Events User Requirements

Author: Lori Corbani

Created: January 26, 2000

Last Modified: January 22, 2004 15:40

# 1 Purpose of Document

This document describes the user requirements for processing Nomenclature Events in the MGI Editorial Interface (EI).

Most of this document originally appeared in the Requirements and Design documents for TR 1291.

# 2 Introduction

The nomenclature database (Nomen) serves as a repository for new, reserved and proposed mouse symbols (symbols which are private and not yet ready for public disclosure). MGD has no mechanism for handling public vs. private data; all data within MGD are considered public.

Nomen serves its purpose well. It enables the Nomenclature group to enter reserved and proposed mouse nomenclature into a data repository. And it enables the Nomenclature group to transfer this data electronically to MGD when the nomenclature has been approved and is ready to be publicized.

 The primary responsibility of Nomen is as the data repository for new, private (reserved), and proposed (pending) mouse nomenclature. MGD is the primary repository for public mouse nomenclature.

# 3 Definitions

- Nomen - the nomenclature database
- MGD - the mouse genome database
- Broadcast - term used to mean the transfer of data from Nomen to MGD and also the process by which the data is transferred (i.e. the Broadcast process).

# 4 User Permissions

4 tiers of users are defined:

1. **Tier 1**

    Read-only users

    - all MGI users

2. **Tier 2**

   1. Can edit entries created by them.

   2. Cannot edit entries created by other users.

   3. Tier 2 Users:

      - Lucette Glass

3. **Tier 3**

   1. All Tier 2 permissions

   2. Can edit *Reserved* or *Unreviewed* entries.

   3. Can Broadcast entries created by them.

   4. Cannot Broadcast entries created by other users unless the entry is *Reserved* or *Unreviewed*.

   5. Tier 3 Users:

      - Dirck Bradt

      - Harold Drabkin

      - David Hill

      - Ira Lu

      - Ken Frazer

      - Moya Lennon-Pierce

      - Pierre Vanden Borre

      - TB Reddy

      - Sophia Zhu

      - Connie Smith

4. **Tier 4**

   1. All editing privileges.

   2. Can process withdrawals in MGD.

   3. Can edit Marker Event Reasons controlled vocabulary in MGD.

   4. Tier 4 Users:

      - Lois Maltais

      - Rebecca Corey

      - Cathy Lutz

      - Cindy Smith

      - Debbie Reed

      - Li Ni

      - Bob Sinclair

- Richard Baldarelli

The following restrictions override the above permissions:

1. The *MGI Accession ID* is assigned by the system.

2. The *Created By* field is set by the system. This value will default to the user who is adding the entry into the database. This value cannot be changed by any user.

3. The *Modified By* field is set by the system. This value will default to the user who last modified the entry in the database. This value cannot be changed by any user.

4. The *Broadcast By*, *Broadcast* statuses and *Broadcast date* fields are set by the Broadcast process. These values cannot be modified by any user.

5. *Editorial Notes* can be modified by Tier 4, Tier 3 and Tier 2 users.

6. *Nomenclature Coordinator Notes* can be modified by Tier 4 users.

7. Tier 2 and Tier 3 users cannot modify the *Status* field.

8. Tier 4 users can modify the *Status* field of any entry to any value other than *Broadcast - Official* or *Broadcast - Interim*.

9. Tier 4 users cannot modify the *Status* field if it is equal to *Broadcast - Official* or *Broadcast - Interim*.

# 5   Nomenclature Statuses

1. **Broadcast - Official** - the symbol has been approved by Mouse Genomic Nomenclature Committee (MGNC) and/or the HUGO Gene Nomenclature Committee (HGNC) and transferred to MGI as an "official" symbol.

2. **Broadcast - Interim** - the symbol has not been approved by MGNC or HGNC at the time it was transferred to MGI. Symbol is temporary and awaiting further review until an official symbol is assigned. This allows data to be associated with the symbol and entered into MGI in a timely fashion.

3. **In Progess** - the symbol is awaiting additional work by the curator or confirmation from HGNC, authors or community experts.

4. **Reserved** - the symbol has been approved by MGNC and is to be kept confidential/private until publication of the paper or author/MGNC/HGNC requests that it be released.

5. **Approved** - the symbol is a locus that was *Reserved* or *In Progress* and is now in MGI as either the same symbol or a different symbol. The symbol can never be Broadcast to MGI since it already exists in MGI.

6. **Deleted** - the symbol is no longer needed for inclusion in MGI. The data associated with the symbol is needed for possible future use. Generally this is a reserved symbol, though it may have been a symbol in MGI that has been deleted in MGI.

7. **Unreviewed** - a BC###### symbol that was entered into Nomen via the LocusLink/MGC load. The data associated with this symbol have not been reviewed by a curator.

# 6   Nomenclature Events and Reasons

When a Nomen to MGD Broadcast occurs or a nomenclature withdrawal event occurs in MGD, the event is recorded in the Marker's History table. Within the controlled vocabulary list of nomenclature events, each defined *event* corresponds to specific data processing tasks within the database. For this reason, the addition or modification of these events must be coordinated between the Nomenclature Committee and the SE group:

1. assigned - the marker symbol has been created
2. withdrawn - the marker symbol is being changed
3. merged - the entire marker record is being merged with another marker record
4. allele of - the entire marker record is being merged with another marker record because the old marker is actually an allele of the other marker.
5. split - the marker record is being split into more than one new marker record
6. deleted - the marker record is being withdrawn but there is no new marker

Each nomenclature event must have one *reason* associated with it:

1. to conform w/Human Nomenclature
2. per gene family revision
3. per personal comm w/Author(s)
4. per personal comm w/Chromosome Committee
5. Not Specified (standard for a CV table)

When a marker symbol is added to Nomen, the default *Event* is *assigned* and the default *Reason* is *Not Specified*. The user can override the *Reason* default during data entry.

When a marker symbol is withdrawn in MGD, the default *Event* is *withdrawn* and the default *Reason* is *Not Specified*. The user can override both defaults during data entry.

When an *Event* term is concatenated with a *Reason* term, the result is a "sentence" which explains the event. If the *Reason* is *Not Specified*, then the *Reason* term will not be included in the concatenation. Examples:

- "withdrawn" + "per gene family revision" = "withdrawn per gene family revision"
- "deleted" + "Not Specified" = "deleted"
- "merged" + "per personal comm w/Author(s)" = "merged per pers comm w/ Author(s)"

Tier 4 users can edit the list of Reasons (add new items, modify existing items, delete unwanted items) in the same way that they can edit other controlled vocabulary lists.

Nomen shares these CV tables with MGD.

# 7   Nomenclature Assignments

Tier 2, Tier 3 and Tier 4 users enter new gene symbols into Nomen.  When the gene symbol is ready to be publicized, it is Broadcast to MGD.

The following information is required:

1. marker type (gene, DNA segment, QTL, etc.)
2. symbol
3. name
4. chromosome
5. primary reference

When a user enters a symbol into the symbol field and hits TAB, a verification routine:

1. Informs the user if the symbol already exists in Nomen.
2. Informs the user if the symbol already exists as an official symbol in MGD.
3. Informs the user if the symbol already exists as an interim symbol in MGD.
4. Informs the user if the symbol already exists as a withdrawn symbol in MGD.

The following additional information is optional (that is, not required by the editorial interface):

1. related references
2. synonyms
3. sequence ids and references, where references are required.
4. editorial notes

When the user adds the Nomen record, the following defaults will automatically be assigned (note that the Event and Event Reason fields can be set by the user prior to clicking the button, but if they are left unset, then the defaults will be used):

1. Event = *assigned*
2. Status = *In Progress*
3. Chromosome = *UN*
4. Created By =  User who is submitting the data
5. Modified By =  User who is submitting the data
6. Broadcast By = null
7. Creation Date = today
8. Modification Date = today
9. Broadcast Date =  null
10. MGI Accession ID = system-generated

When the user hits the *Add and Broadcast - Official to MGD* button, the following defaults will automatically be assigned (note that the Event and Event Reason fields can be set by the user prior to clicking the button, but if they are left unset, then the defaults will be used):

1. Event = *assigned*
2. Status = *Broadcast - Official*
3. Chromosome = *UN*
4. Created By =  User who is submitting the data
5. Modified By =  User who is submitting the data
6. Broadcast By = User who is broadcasting the data
7. Creation Date = today
8. Modification Date = today
9. Broadcast Date = today
10. MGI Accession ID = system-generated

If the Broadcast to MGD fails, then the defaults for the *Add* button will be used instead.

# 8    Broadcast to MGD

There are two methods by which Nomen symbols are Broadcast to MGD:

1. Official Symbol - immediate Broadcast of one Nomen symbol to MGD.  The symbol cannot have a status of *Broadcast*, *Approved* or *Deleted*.  Tier 3 and Tier 4 users only. Tier 3 users can only Broadcast symbols which they have created in Nomen and which have a status of *In Progress*.  The status of the symbol in MGD is set to *official*.

2. Interim Symbol - immediate Broadcast of one Nomen symbol to MGD.  The symbol cannot have a status of *Broadcast*, *Approved* or *Deleted*.  Tier 3 and Tier 4 users only. Tier 3 users can only Broadcast symbols which they have created in Nomen and which have a status of *In Progress*.  The status of the symbol in MGD is set to *interim*.

During a Broadcast, the following information is transferred to MGD.  Note that the *Chromosome* field is not Broadcast but is always set to *UN* in MGD.

1. Event = *assigned*
2. Event Reason
3. Marker Type
4. Marker Status
5. Symbol
6. Name
7. Primary Reference
8. MGI Accession ID

9.   Related References

10. Synonyms

11. Other Accession IDs and References (Sequence,etc.)

# 9   Nomenclature Withdrawals

The withdrawal events are:

1.   Simple; occurs when the one new symbol does not exist in MGD.  This method corresponds to the event *withdrawn*.

2.   Merge; occurs when the one new symbol does exist in MGD.  This method corresponds to the event *merged*.

3.   Allele Of; occurs when the one new symbol does exist in MGD and the old symbol is an allele of the new symbol.  This method corresponds to the event *allele of*.

4.   Split; occurs when there are multiple new symbols which do not exist in MGD. This method corresponds to the event *split*.

5.   Deletion; occurs when no new symbol is specified.  This method corresponds to the event *deleted*.

## 9.1  Simple Withdrawal

A *simple withdrawal* is one where the new symbol does not exist in MGD.  This is a basic symbol change.

The following data changes occur for the old symbol:

1.   marker status = *withdrawn*

2.   name = *withdrawn, new symbol*

3.   current symbol = *new symbol*

4.   offset = *-999.0*

5.   all marker history is migrated to new symbol

6.   all allele symbols are converted to the new symbol

7.   all other data (mapping, homology,  etc.) is migrated to the new symbol

## 9.2  Merge Withdrawal

A *merge withdrawal* is one where the new symbol does exist in MGD.  All information associated with the old symbol is migrated to the new symbol.

The following data changes occur for the old symbol:

1.   marker status = *withdrawn*

2.   name = *withdrawn, new symbol*

3.  current symbol = *new symbol*
4.  offset = *-999.0*
5.  all marker history is migrated to new symbol
6.  all allele symbols are converted to the new symbol
7.  all other data (mapping, homology,  etc.) is migrated to the new symbol

## 9.3  Allele Of Withdrawal

An *allele withdrawal* is one where the new symbol is an allele of the new symbol and the new symbol does exist in MGD.  All information associated with the old symbol is migrated to the new symbol.  It is a special case of a *merge*.

The following data changes occur for the old symbol:

1.  marker status = *withdrawn*
2.  name = *withdrawn, allele of new symbol*
3.  current symbol = *new symbol*
4.  offset = *-999.0*
5.  all marker history is migrated to new symbol
6.  all allele symbols are converted to the new symbol
7.  all other data (mapping, homology,  etc.) is migrated to the new symbol

## 9.4  Split Withdrawal

A *split withdrawal* is one where there is more than one new symbol and all new symbols do not exist in MGD.  In a split withdrawal, all information associated with the old symbol remains associated with the old symbol and must be manually copied to one of the new symbols (this includes Accession Numbers, Alleles, MLC, Homology, Mapping, etc.) and manually deleted from the old symbol.  The preferred MGI accession ID of the old symbol also remains associated with the old symbol (the rules of accession numbers state that an accession number can be associated with one and only one object).

The following data changes occur for the old symbol:

1.  marker status = *withdrawn*
2.  name = *withdrawn, = new symbol 1, new symbol 2, ...*
3.  current symbol = *new symbol1, new symbol2, ...*
4.  offset = *-999.0*
5.  all marker history is migrated to all new symbols

## 9.5  Deletion Withdrawal

A *deletion withdrawal* is one where there is no new symbol.  In a deletion withdrawal, all information associated with the old symbol remains associated with the old symbol because there is no new symbol.

The following data changes occur for the old symbol:

1. marker status = *withdrawn*
2. name = *withdrawn*
3. offset = *-999.0*

## 9.6  Allele Conversions

During a *simple* , *merge* or *allele of* withdrawal, allele symbols of the old symbols are converted using the following rules:

1. *oldsymbol<allele>* is converted to *newsymbol<allele>*.
2. *allele* is converted to *newsymbol.*
3. *allele<+>* is converted to *newsymbol<+>*.

For example, if symbol A is withdrawn to symbol B:

1. Allele *A<a1>* is converted to *B<a1>*.
2. Allele *A* is converted to *B*.
3. Allele *A<+>* is converted to *B<+>*.

During an *allele of* withdrawal, allele symbols are converted using the following rules:

1. *oldsymbol<allele>* is converted to *newsymbol<oldsymbol-allele>*.
2. *allele* is converted to *newsymbol<allele>*.
3. *allele<+>* is converted to *newsymbol<+>*.

For example, if symbol A is withdrawn as an allele of B:

1. Allele *A<a1>* is converted to *B<A-a1>*.
2. Allele *A* is converted to *B<A>*.
3. Allele *A<+>* is converted to *B<+>*.

No allele conversions occur during a *split*.  Alleles must be manually moved to the appropriate new symbol after the split has been processed.