



887/242: Copy of initial coherence information to be stored in series knowledge

Application: SEASABS - V 2.6

Status:

Awaiting Support Manager Assignment

Change Initiator

Name: **Frank Masci**

Requirement Details

Requirement Summary: **Copy of initial coherence information to be stored in series knowledge**

Impact: **New functionality**

Due Date: **15/12/2006**

Dependent Events
:

Related Requests: **887/180**

Background Information:

We need facilities within SEASABS to assist in managing time series which are "conceptually related" and to assist in assessing levels of consistency through diagnostics and summary measures. This project is broken into two phases:

- Phase 1: creation, storage and management of related series identifiers
- Phase 2: the development of a user-interface to compute and display consistency summary measures

Disclaimer: please avoid using words like "coherence", "coherent" or derivatives thereof in any of this work (eg. window titles, labels etc..). Words like "related" or "consistent" (the latter in the right context) should be used.

Requirement Description:

Phase 1: Creation and storage of related series in the series knowledge

a) perhaps record under series options, ability to enter in more than one time series as consistent.

b) group name and link to where group resides in SEASABS, ie. GROUP.SERIES ID

c) it's up to user to enter this information "manually" - preferred option is with a pick list from available time series across groups and owner groups.

d) a warning flag to be issued under a "Research Analysis" (not

concurrent) run that displays a message that this series may be related to other time series with a list of those particular time series. This warning should appear at the end of the reanalysis.

e) ability to click/select on a consistent time series and SEASABS opens this series (similar to the aggregation structure which allows clicking on a name aspect).

f) ability to store in the series knowledge information (metadata) about the level of consistency or what makes them consistent. For example, a description of the way in which highlighted series are consistent. Eg. "Series A comprises 90% of series B so movements should be similar, except in April due to seasonal rise in parakeet sales (parakeet sales not recorded in series A, but recorded in series B)".

g) ability to export a list of all related series names to DOS as a text/ascii file. More specifically, all series in a group that have been declared to be related to another series in either same group, other groups (including other owner groups) should be saved. The format of this output list should be for example:

```
<ownergroup>,<group>,<seriesname1a>,<ownergroup>,<group>,<seriesname1b>
```

```
<ownergroup>,<group>,<seriesname2a>,<ownergroup>,<group>,<seriesname2b>
```

```
<ownergroup>,<group>,<seriesname3a>,<ownergroup>,<group>,<seriesname3b>
```

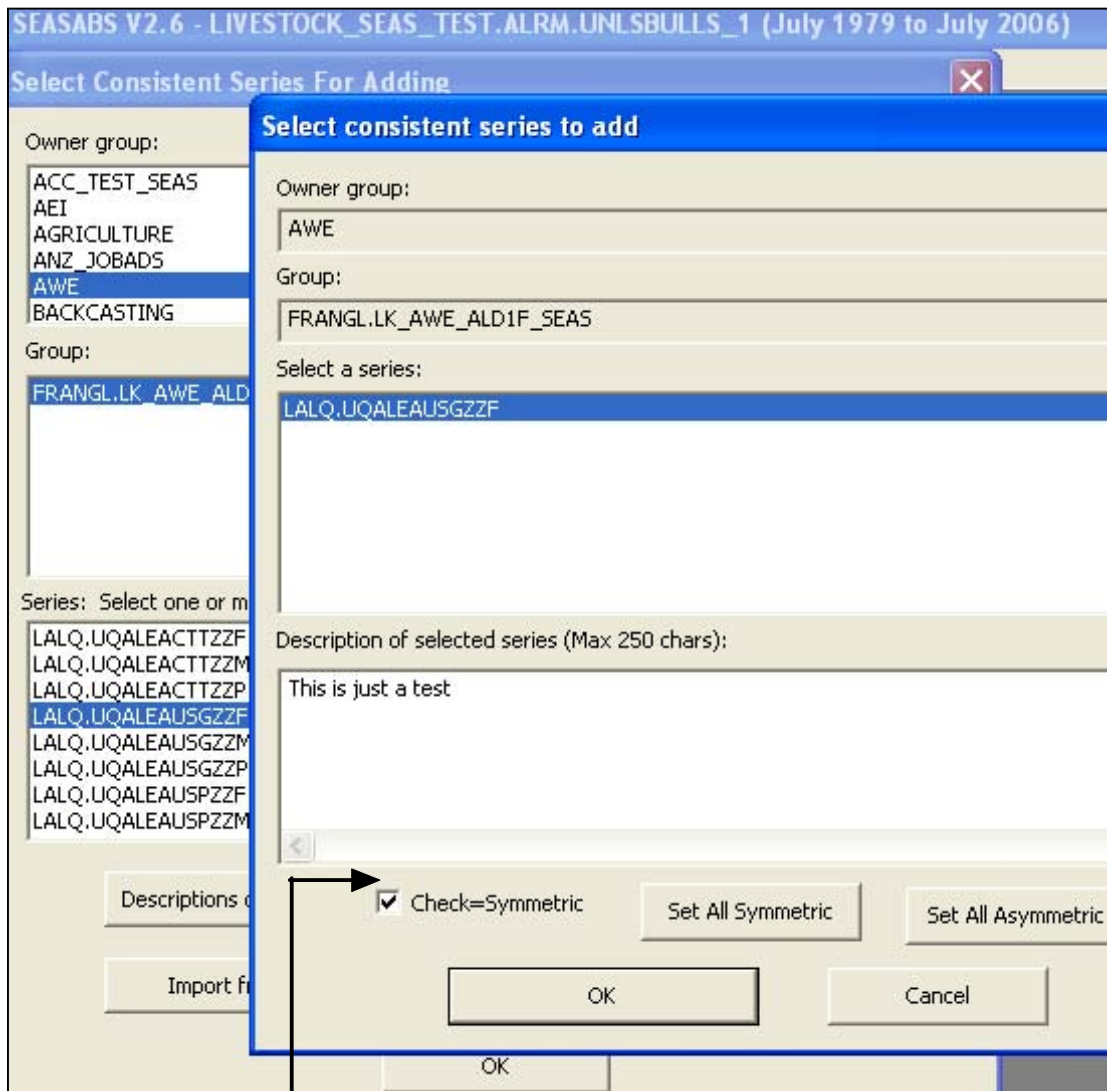
etc..

h) ability to import a text/ascii file containing a list of related series pairs as in the example in (h). This obviates the need for a user to manually select related series from a list within SEASABS. On import, this information should be stored directly into the series knowledge.

- The relationship between series would be symmetric. For example, if series A is related to series B this means that series B is related to A and so should be reflected in both series knowledge records. This pairing (two-way) association should be implicit in the series knowledge: ie. a user need not explicitly declare A -> B and B -> A separately. When A -> B is selected, then this immediately implies B -> A, ie. "A <-> B" is assumed.
- If A is related to B and B is related to C does that always mean A is related to C?: Possibly - but if we can't reflect this relationship within the series knowledge then we wouldn't be doing this automatically. ie. it would be up to the user to enter this information manually at this stage. The design should be based on explicit user-defined (paired-series) associations.
- We would like to automatically include who declared the association and when.
- If we copy a series or group with the production flag set, then the series knowledge about consistent time series should be retained. The production group is the "master" version from which deletion of prior declared associations is disallowed without sufficient privileges.

Some fixes/enhancements to PHASE 1 functionality following production release v2.6:

- Ability to import generic text file containing many different related series across any groups and/or owner groups, not just those series related to the single series loaded in a SEASABS session. See point (g) above for input format.
- Ability to assign attributes (or "dimension" information) to each series when a user declares a related series pair. This will enable a user to display all related series according to pre-selected attributes. This will later assist when generating consistency-diagnostic information by allowing users to focus on those related series that have a common theme or attribute. The specification for this enhancement can be built on the existing GUI. Here are the details:



Place new dialogue window here

Set series attributes (check all that apply for this pair):

Series 1:

LIVESTOCK_SEAS_TEST.ALRM.UNLSBULLS_1

Series 2:

FRANGL.LK_AWE_ALD1F_SEAS.LALQ.UQALEAUSGZZF

<u>Attribute</u>	<u>Series 1</u>	<u>Series 2</u>
Grand total	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Total industries/one state	<input type="checkbox"/>	<input type="checkbox"/>
Total states/one industry	<input type="checkbox"/>	<input type="checkbox"/>
General lower level	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Chain Volume Measure (CVM)	<input type="checkbox"/>	<input type="checkbox"/>
Current Price Value (CPV)	<input type="checkbox"/>	<input type="checkbox"/>
Monthly	<input type="checkbox"/>	<input type="checkbox"/>
Quarterly	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Conceptually related	<input type="checkbox"/>	<input type="checkbox"/>
Source series	<input type="checkbox"/>	<input type="checkbox"/>
Derived series	<input type="checkbox"/>	<input type="checkbox"/>

Current Price Value (CPV)	<input type="checkbox"/>	<input type="checkbox"/>
Monthly	<input type="checkbox"/>	<input type="checkbox"/>
Quarterly	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Conceptually related	<input type="checkbox"/>	<input type="checkbox"/>
Source series	<input type="checkbox"/>	<input type="checkbox"/>
Derived series	<input type="checkbox"/>	<input type="checkbox"/>

Notes:

- The above is purely illustrative, showing the boxes that a user may check when declaring a related series pair. By default, we would like no boxes checked on start-up. An error should be issued if a user did not check any boxes, i.e. if they inadvertently clicked OK in the above panel without checking any boxes. This error should be issued as a message on the screen, e.g: "no attributes were checked; please go back and tick all applicable attributes".
- The user is free to check more than one box for each series of a related pair. This information should obviously also be stored in the consistency metadata for later reference (see below).
- "**Series 1**" in the above panel should always refer to the series that is loaded in the current SEASABS session (in above example="LIVESTOCK_SEAS_TEST.ALRM.UNLSBULLS_1"). "**Series 2**" is then the series that is manually selected by the user in the "Select Consistent Series For Adding" panel (in above example="FRANGL.LK_AWE_ALD1F_SEAS.LALQ.UQALEAUSGZZF").
- For your information, here are the specific definitions for the above attributes. We envisage this will make its way into the SEASABS help.
 - **Grand total:** E.g. series representing national aggregate over all states and/or all industries within.
 - **Total industries/one state:** E.g. in the case of a 2D aggregation structure, this series could be the "marginal" sum over all industries for a particular state. Note that here "industry" and "state" are arbitrary attributes. These can stand in for other generic classification breakdowns, e.g. "number of juniors employed" versus "employment type".
 - **Total states/one industry:** E.g. in the case of a 2D aggregation structure, this series could be the "marginal" sum over all states for a particular industry or category. Note that here "industry" and "state" are arbitrary attributes. These can stand in for other generic classification breakdowns, e.g. "number of juniors employed" versus "employment type".
 - **General lower level:** Generic low level component (child) series that may be part of a larger aggregation structure. E.g. a series defined for a particular single industry and single state.
 - **Chain Volume Measure (CVM):** a series representing the "chained volume" equivalent of the CPV.
 - **Current Price Value (CPV):** a series representing the "current price" equivalent of the CVM.
 - **Monthly:** a series whose periodicity is monthly.
 - **Quarterly:** a series whose periodicity is quarterly.
 - **Conceptually related:** series that has an explicit conceptual relationship to another series (often declared by client).
 - **Source series:** series representing the "raw input" or progenitor of a collection that feeds into a series used/published by another subject matter area (e.g. input into National Accounts). See also "Derived series".
 - **Derived series:** series that is derived from an input series originating from

another collection (from the "Source series"). Most National Accounts series are "derived".

Phase 2: Facility/user-interface to compute and display consistency summary measures and diagnostics

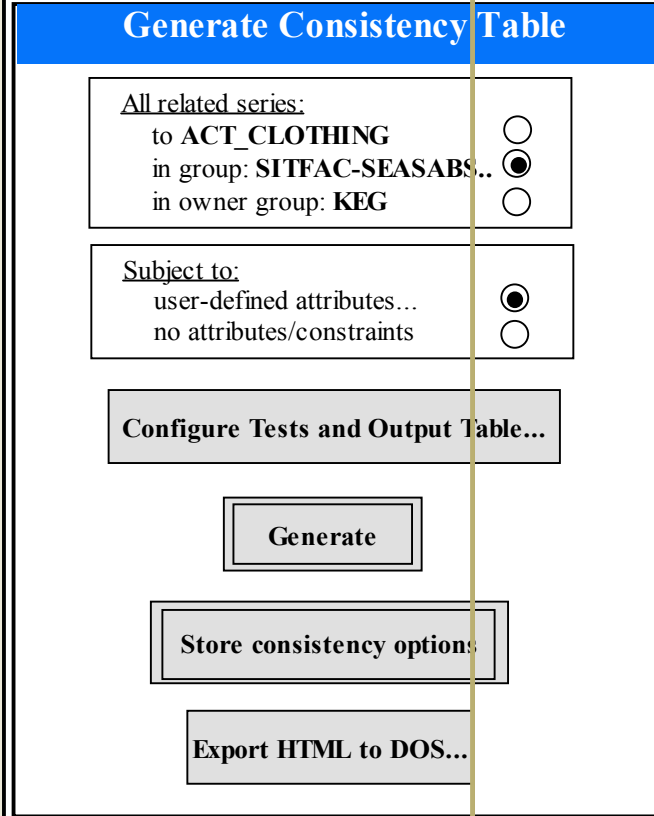
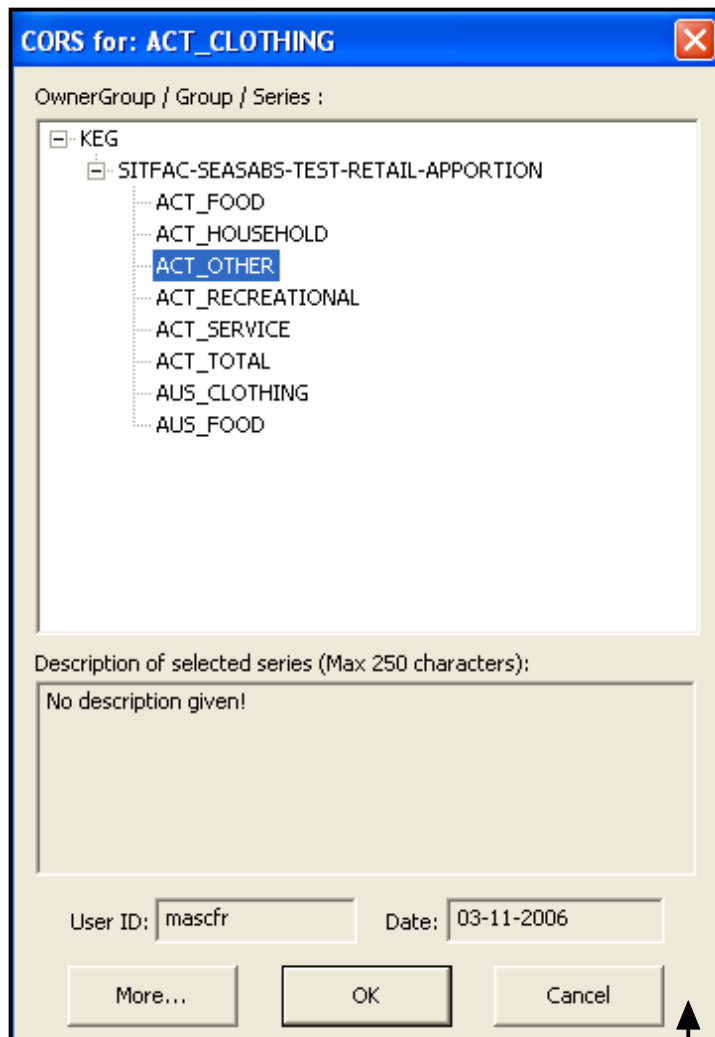
(1). We would like to generate consistency diagnostics for a list of related series pairs:

(a) either related to the single series open in a SEASABS session;

(b) for all series in the open group that are related to any other series across all other SEASABS groups/owner groups;

(c) for all series in the open owner group that are related to any other series across all other SEASABS groups/owner groups.

We envisage the following GUI for generating consistency diagnostics. This should be accessed from the existing "Related Series..." window in SEASABS version 2.6, e.g:



Consistency Table...

We'll leave it to the developer's imagination and creativity to come up with an esthetic interface.

(2). Having picked the desired "**All related series:**" radio button option (first panel), the user can now fine tune the list of series pairs by only keeping those "**Subject to:**" either "**user-defined attributes**" or "**no attributes/constraints**". If "**user-defined attributes**" is selected, a window of all attributes with checkboxes should pop up (e.g. same attributes that were used to declare all related series from the phase 1 implementation):

Select desired series attributes:	
Grand total	<input checked="" type="checkbox"/>
Total industries/one state	<input type="checkbox"/>
Total states/one industry	<input type="checkbox"/>
General lower level	<input type="checkbox"/>
Chain Volume Measure (CVM)	<input type="checkbox"/>
Current Price Value (CPV)	<input type="checkbox"/>
Monthly	<input type="checkbox"/>
Quarterly	<input type="checkbox"/>
Conceptually related	<input type="checkbox"/>
Source series	<input type="checkbox"/>
Derived series	<input type="checkbox"/>


The user is free to select as many "attributes" as they wish. The initial list of series pairs will be filtered against these attributes. If a single series in the list (not a related pair) has any of the selected attributes set (i.e. as set when first declaring related series maps), then it, together with its related counterpart should be included in the 'final list'. This 'final list' of series pairs will be used for generating the consistency table. So in the above example, the checked box would mean that all series pairs where one of its members is a "Grand total" would be selected. If for instance "Grand total" and "Quarterly" were both selected above, then the final list will contain pairs where either member is a "Grand total" OR "Quarterly" series. Each series pair in the final list should be unique, i.e. no duplicates should exist prior to generating the consistency table.

If the user had selected "**no attributes/constraints**" in the second panel of the "Generate Consistency Table" window, then no filtering of the initial list of series pairs is needed.

(3). If the "**Configure Tests and Output Table...**" button is pressed, an interface of dialogue boxes should pop-up that allows a user to specify the following configuration parameters and output options:

- Compute and report consistency diagnostics in table? Yes / No; default = Yes. "Diagnostics" are defined in (4) below. Otherwise, only report SEASABS parameters and settings in the output HTML table (see 3 below).
- Report prior corrections in HTML table (namely TBs, SBs, LGEXs and AOs)? Yes / No; default = Yes.
- Time spans on which to apply all tests. Default = minimum common span of a pair.
- Time spans for temporal-movement plots; default = minimum common span of a pair.
- Critical P-values for tests 2,3,4,5,6 with global default = 5%.
- Outlier detection threshold (#sigma); default = 3 sigma.
- Movement difference threshold (test 5); default = 2 sigma.

- Max. identical movement tolerance; default = 98%

On this interface, there should also be an option to retrieve the options/settings that were set in a previous (historical) run of this tool (previously stored by pressing the "Store consistency options button"; see 11 below). We don't want to store the consistency measures themselves. We want to be able to re-compute them on the fly using any stored options (if they exist), the stored historical series knowledge and original series. We are currently not sure how to reference/label the historical runs. A version number of the consistency run and pointers to the original data and historical series knowledge (by end analysis-date) may be enough. We'll leave this to the programmer. This design is strongly recommended by Duncan McCaskill:  (Subject: SEASABS Consistency relationships, phase 2 - issues; Database: TA-Statistical IT Facilities WDB; Author: Duncan McCaskill; Created: 05/01/2007; Doc Ref: DMCL-6X69Z3).

(4). If the "**Generate**" button is pressed, a HTML table of SEASABS parameters, settings and diagnostics should be generated and spawned to the screen. Here's an example from our prototype tool:



results.html

The first series of a pair is in a yellow row and the corresponding second (related) series is in a subsequent blue row. "First" and "second" here is arbitrary. Note that if a series of a pair is repeated in any of the subsequent pairs of a list (e.g. if comparing an aggregate series against state components), then this series should only be shown once, i.e. in a yellow row in the output table. The series that are related (paired) to this single series should then be listed in subsequent blue rows. For example:



results_aggstruct.html

We request some modifications to this table which were not possible to program in our prototype tool. These are outlined in 5 and 6 below.

(5). The links labelled with "[Diagnostics tests failed: ...](#)" in the above HTML tables give a summary of the failed diagnostic tests. Clicking on these shows the diagnostics computed for the "ABS seasonally adjusted" component series only. We would like to additionally compute diagnostics for the "original" and "D11" seasonally adjusted components aswell. So in the end, there will be three separate links to the diagnostics for each series pair in this table. For example:

[Diagnostics tests failed for Originals: 1, 2, 4](#)
[Diagnostics tests failed for ABS-SA: 1, 3, 5, 6](#)
[Diagnostics tests failed for D11-SA: 1, 2](#)

(6). The summary table of "SEASABS parameters and settings" in the example HTML tables above is not complete. We would like to include the following additional parameters and values:

- adjustment type: "concurrent" or "forward factor";
- method: value can be either "direct", "indirect", or "not adjusted". Note: our prototype tool currently only reports "direct" or "indirect" (as available from TSA-download). "not adjusted" cases must be available from the series

knowledge, right? If so, we want to be able to report "not adjusted" cases in the method field too.


- Star value;
- I/C ratio;
- I/S ratio;
- If static trading day exists, the trading day weights (table C15 SEASABS outputs) for each series listed in the "Trading Day" row with an indication that these are static trading day weights;
- If moving trading day exists, the trading day weights (table C15 SEASABS outputs) for the last year available in each series listed in the "Trading Day" row. There should also be an indication that these are "year XXXX" moving trading day weights;

(7). It is requested that all prior corrections in the HTML table (namely trend breaks, seasonal breaks, large-extremes and additive outliers) be listed one per line and in time order (latest to earliest).

(8). Algorithms for computing all consistency diagnostics with plots have been implemented in a prototype tool written in "Perl" and "R": <S:\consistency\consistency.pl>. Complete specifications for each algorithm (for each diagnostic test) will be provided later. These may need to be re-programmed - a difficult undertaking since most algorithms are coded as complicated "R" function libraries. Other suggestions include: making calls to the R-libraries directly, or, invoking "R" as an executable from SEASABS. The latter will be slow and inefficient. This aspect is still on the drawing board and should not prevent implementation of items 1 - 7 above. The priority is to get the HTML table of parameters/settings coded up. The ancillary diagnostics that appear under the links mentioned in 5 can wait.

(9). In one of the diagnostic output plots (namely the movement versus movement plots) that are generically named "mvt_vs_mvt_pair_#.png" by TSA's prototype tool (generated under TEST 4), we would like to add another feature. For any outliers in this plot (currently labelled by a "X" symbol; see examples linked in HTML table above), we would like the ability to click on an outlier and a pop-up window would appear showing the time point to which the time-matched movements apply.

(10). In all plots and diagnostic outputs where we refer to time points, we would like these written in "long format" as used in SEASABS plots, e.g: Jan 1994, Qtr1 1994 etc..

(11). The button labelled "**Store consistency options**" allows the user to store the options and settings defined in item 3 above. These, together with the historical series knowledge and original series can be used to recreate a HTML table on the fly. This obviates storing any consistency measures. This design is strongly recommended by Duncan McCaskill:  (Subject: SEASABS Consistency relationships, phase 2 - issues; Database: TA-Statistical IT Facilities WDB; Author: Duncan McCaskill; Created: 05/01/2007; Doc Ref: DMCL-6X69Z3).

(12). The "**Export HTML to DOS...**" button allows a user to save the HTML table of SEASABS parameters and settings to a filename and directory of one's choice. The plots buried within the diagnostic links outlined in 5 above should

not be saved (since they will take up a lot of room and be rather cumbersome to manage). Only the textual contents of these diagnostic links should be saved. The user of course should be able to copy and paste any diagnostic plot from the SEASABS interface (like now) on a case-by-case basis for insertion into a Notes document.

Acceptance
Criteria:

Works as specified subject to programming and infrastructure limitations

Client Manager

Frank Masci mailed the Client Manager(s) seeking task approval on 03/11/2006
10:20:34 AM

Name: | **Frank Masci**

Additional
Reader(s):

Planning Details

Reason(s) for Change: **Client request**

Client Resource
Estimates:
(half days)

Priority: Critical High Normal Low

Priority Number: **0**

Test Release Date: **01/12/2006**

Production Release
Date: **15/12/2006**

Revised Date:

**Approval - New
Request**

Frank Masci approved this new request on 03/11/2006 10:24:35 AM

Approval Status: **New Request Approved**

Comments: **Phase 1 was accepted for the 2.6 release but it is incomplete. Further functionality is requested. This new SR covers the extra functionality for phase 1 and the future phase 2.**

**Approval -
Production
Release**

Approval Status:

Comments:

Support Manager

Name: | **Duncan McCaskill**

Comments:

Planning Details

[TA Resource](#)

[Estimate:](#)

(half days)

Cost: **\$0.00**

Note: The estimated cost is derived by using an average TA figure of **\$228** per half day.

[Comments:](#)

System Links

[Register of](#)

[Projects Entry:](#)

[Change](#)

[Management Entry](#)

:

[ABS Asset](#)

320152

[Number:](#)

[Application Type:](#) **Business Infrastructure Application**

Programmer

Name:

Actual Start

Date:

Actual Finish

Date:

Time spent on

this task:

(half days)

Solution

[Activity Type:](#)

Complexity: High Medium Low

Description of Action Taken: ➡

[Units Remediated/Built:](#)

No	Name	Version Number (if applicable)	Comments
.			
1.			

2.			
3.			
4.			
5.			

Unit and Integration Testing

Unit & Integration Testing Completed and Signed-off: **Yes** **No** **N/A**

Comments:

Documentation

Changes Documented: **Yes** **No** **N/A**

Comments:

System Testing

Does this change warrant System Testing? **Yes** **No**

System Tester(s):

Testing Instructions:

Expected Results:

Actual Results: **Passed** **Failed**

Comments:

Approval - System Testing

Approval Status:

Comments:

User Acceptance Testing

Does this change warrant User Acceptance **Yes** **No**

Testing?

Client Tester(s):

[Testing Instructions:](#)

[Expected Results:](#)

Actual Results: **Passed** **Failed**

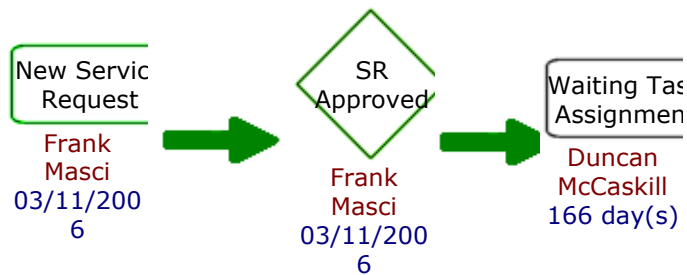
[Comments:](#)

Approval - UA Testing

Approval Status:

[Comments:](#)

Workflow



Audit Log

03/11/2006 10:24:40 AM Commencement Approval: Frank Masci approved task commencement
03/11/2006 10:20:34 AM Task Approval Sought: Frank Masci sought task approval from Frank Masci
03/11/2006 10:19:38 AM New Task created using Make Similar button (from Task ID 887/180) by Frank Masci

29/03/2007 08:11:06 AM: Frank Masci
29/03/2007 08:09:57 AM: Frank Masci
29/03/2007 08:09:49 AM: Frank Masci
27/03/2007 12:10:40 PM: Jack Shen
27/03/2007 09:07:56 AM: Frank Masci
27/03/2007 09:06:22 AM: Frank Masci
27/03/2007 09:04:13 AM: Jack Shen
26/03/2007 05:18:54 PM: Jack Shen
26/03/2007 05:15:53 PM: Jack Shen
26/03/2007 05:10:05 PM: Jack Shen
26/03/2007 03:46:25 PM: Frank Masci
26/03/2007 03:45:28 PM: Frank Masci
26/03/2007 03:44:26 PM: Frank Masci
26/03/2007 03:40:10 PM: Frank Masci
26/03/2007 03:34:05 PM: Frank Masci
26/03/2007 03:28:42 PM: Frank Masci
26/03/2007 03:26:45 PM: Frank Masci
26/03/2007 03:19:03 PM: Frank Masci

26/03/2007 03:04:02 PM: Frank Masci
26/03/2007 02:53:00 PM: Frank Masci
26/03/2007 02:46:27 PM: Frank Masci
26/03/2007 02:38:13 PM: Frank Masci
26/03/2007 02:22:40 PM: Frank Masci
26/03/2007 02:20:52 PM: Frank Masci
26/03/2007 02:04:41 PM: Frank Masci
26/03/2007 02:01:46 PM: Frank Masci
26/03/2007 01:59:40 PM: Frank Masci
26/03/2007 01:51:14 PM: Frank Masci
26/03/2007 01:40:38 PM: Frank Masci
26/03/2007 01:34:31 PM: Frank Masci
26/03/2007 01:20:30 PM: Frank Masci
26/03/2007 01:09:59 PM: Frank Masci
26/03/2007 01:05:51 PM: Frank Masci
26/03/2007 12:39:12 PM: Frank Masci
26/03/2007 11:21:11 AM: Frank Masci
26/03/2007 11:18:02 AM: Frank Masci
15/03/2007 10:15:29 AM: Frank Masci
15/03/2007 10:15:11 AM: Frank Masci
15/03/2007 09:55:28 AM: Frank Masci
15/03/2007 09:54:55 AM: Frank Masci
15/03/2007 09:53:08 AM: Frank Masci
15/03/2007 09:47:23 AM: Frank Masci
15/03/2007 09:45:36 AM: Frank Masci
15/03/2007 09:17:51 AM: Frank Masci
15/03/2007 09:02:47 AM: Frank Masci
14/03/2007 04:15:47 PM: Frank Masci
14/03/2007 02:27:50 PM: Frank Masci
14/03/2007 02:27:04 PM: Frank Masci
14/03/2007 02:04:34 PM: Frank Masci
14/03/2007 12:43:51 PM: Frank Masci
13/03/2007 12:07:02 PM: Frank Masci
13/03/2007 09:43:51 AM: Frank Masci
13/03/2007 09:43:33 AM: Frank Masci
13/03/2007 09:24:14 AM: Frank Masci
13/03/2007 09:15:47 AM: Frank Masci
13/03/2007 09:11:36 AM: Frank Masci
13/03/2007 09:01:42 AM: Frank Masci
13/03/2007 08:56:59 AM: Frank Masci
05/03/2007 10:26:34 AM: Frank Masci
05/03/2007 10:16:36 AM: Frank Masci
15/01/2007 10:52:38 AM: Frank Masci
15/01/2007 10:52:25 AM: Frank Masci
12/01/2007 03:55:42 PM: Duncan McCaskill
04/01/2007 10:07:44 AM: Duncan McCaskill
17/11/2006 12:47:57 PM: Frank Masci
17/11/2006 12:44:50 PM: Frank Masci
17/11/2006 12:40:38 PM: Frank Masci
17/11/2006 12:03:04 PM: Frank Masci
17/11/2006 11:47:34 AM: Frank Masci
17/11/2006 11:45:27 AM: Frank Masci
17/11/2006 11:43:19 AM: Frank Masci
17/11/2006 11:16:08 AM: Frank Masci
17/11/2006 11:14:39 AM: Frank Masci
17/11/2006 11:12:07 AM: Frank Masci
17/11/2006 11:01:57 AM: Frank Masci
17/11/2006 10:49:28 AM: Frank Masci
17/11/2006 10:48:36 AM: Frank Masci
17/11/2006 10:47:10 AM: Frank Masci
17/11/2006 08:52:23 AM: Frank Masci
17/11/2006 08:33:00 AM: Frank Masci
17/11/2006 08:19:48 AM: Frank Masci
16/11/2006 05:10:50 PM: Frank Masci
16/11/2006 05:10:39 PM: Frank Masci
16/11/2006 05:05:19 PM: Frank Masci
16/11/2006 05:05:11 PM: Frank Masci
16/11/2006 05:02:35 PM: Frank Masci
16/11/2006 04:54:58 PM: Frank Masci
16/11/2006 04:45:05 PM: Frank Masci
16/11/2006 04:32:15 PM: Frank Masci
16/11/2006 04:22:04 PM: Frank Masci

16/11/2006 04:11:58 PM: Frank Masci
16/11/2006 04:09:26 PM: Frank Masci
16/11/2006 03:56:01 PM: Frank Masci
16/11/2006 03:41:49 PM: Frank Masci
15/11/2006 03:08:18 PM: Frank Masci
15/11/2006 03:06:52 PM: Frank Masci
15/11/2006 02:58:05 PM: Frank Masci
15/11/2006 02:29:23 PM: Frank Masci
03/11/2006 10:24:49 AM: Frank Masci
03/11/2006 10:24:40 AM: Frank Masci
03/11/2006 10:21:23 AM: Frank Masci
03/11/2006 10:20:34 AM: Frank Masci
03/11/2006 10:20:26 AM: Frank Masci

Total document updates: 103