

ALA IMPLEMENTATION PLAN FOR MICROORGANISM COLLECTION DATA

Author(s): Matt Branford
Version: v0.5
Date: January 2011
File: Microorganism data collection plan v0.5.docx

Revision history

Version	Date	Author(s)	Change description
0.5	January 2011	Matt Branford	Refine schedule dates, 'Shared Data Model' references
0.4	December 2010	Matt Branford	Minor amendments, clarifications from internal review
0.3	November 2010	Matt Branford	Detail "Aggregation Service" Adjust "Common Data Model" to base on information already provided Adjust "AMRiN Requirements" to base on information already provided General refinement
0.2	November 2010	Matt Branford	Title change Adjust activity/products for pilot Add "Program Alignment" Add "Project Operation"
0.1	October 2010	Bryan Kalms	Initial draft

Table of contents

1	Introduction	4
1.1	Purpose	4
1.2	Approach	4
2	Project Activities.....	7
2.1	Confirm AMRiN Data Requirements	7
2.2	Develop “Shared Data Model”	9
2.3	Implement BioloMICS Pilot Sites	11
2.3.1	Activity for each Pilot Site Implementation.....	12
2.4	Mobilise data into ALA/ AMRiN	14
2.4.1	BioloMICS data mobilisation.....	14
2.4.2	Non-BioloMICS data mobilisation.....	14
2.5	Implement Each CHACM Member Institution.....	16
3	Project Operation	17
3.1	Quality and Configuration Management.....	17
3.2	Communication Management	17
3.3	Reporting	17
3.4	Resources	17
3.5	Post-Project - Ongoing Operation of Delivered Systems.....	18

1 Introduction

1.1 Purpose

This document describes how ALA will implement deliverables agreed with CHACM through the *BioloMICS Implementation Proposal*. It outlines the approach to achieving:

Objective	...Met through Deliverable
Aggregate data held by microorganism collection owners, and make it available through ALA	ALA data collection/publication services for microorganism data
Deliver tools to assist members of the Australian microorganism community improve management of collection information	BioloMICS implemented for all CHACM member institutions that request it
Deliver tools and frameworks to assist members of the Australian microorganism community collaborate and share other information	AMRiN redeveloped as an ALA Hub

The deliverables enable:

- ALA to deliver in line with the *ALA Implementation Plan* and *ALA Business Plan 2010-11*
- Australian microorganism community progress towards objectives outlined in [Australian Microbial Resources](#)
- All parties to implement the [BioloMICS Implementation Proposal](#).

1.2 Approach

The approach for ALA microorganism data collection is to:

- develop a “Common Data Model” for microorganism information
- develop data collection and publication services so institutions can share information through ALA and AMRiN
- provide BioloMICS software to members of the Council of Heads of Australian Collections of Microorganisms (CHACM) who wish to use this software to manage their microorganism collections
- assist BioloMICS institutions install and set up the software, including migrating existing data into the new application and establishing protocols and mechanisms to share data through the ALA and AMRiN
- assist non-BioloMICS institutions to establish protocols and mechanisms to share data through the ALA and AMRiN
- redevelop a portal (AMRiN – Australian Microorganism Research information Network) to facilitate information sharing and collaboration.

Project activity is summarised at Table 1 and detailed in section 2.

Table 1 Summary of Activities, Schedule and Key Deliverables

Activity	Schedule	Comments	Key Deliverables
Confirm AMRiN requirements	November-January 2010	This information will inform the “Shared Data Model”, ALA aggregation services, and if ALA assistance is required to extend or redevelop http://www.amrin.org/ .	AMRiN Data Requirement Statement Rollout Schedule
Develop “Shared Data Model”	January-February 2011		Shared Data Model
Implement Pilot Sites	January-March 2011	Pilot objective is to optimise the BioloMICS installation/migration process	Site Implementation Plans Site Implementations Site-specific Data Mappings Change in response to Pilot “Lessons Learned”
Mobilise Data	April-May 2011	Aggregate data from all institutions so that it is available through AMRiN and ALA	Data Mobilisation function and assistance
Implement each CHACM member institution	As per Rollout Schedule	Implement BioloMICS or assist aggregation service use for each CHACM member institution. Up to 24 weeks, assuming 24 participants each taking up to 1 week. Involves: <ul style="list-style-type: none"> • mapping existing data fields to AMRiN data model • installing BioloMICS software if necessary • migrating data into BioloMICS • liaising and consulting with IT etc 	Site Implementation Plans Site Implementations Site-specific Data Mappings
Rebuild AMRiN as an ALA Hub	July-November 2011	Customisation of basic ALA Hub site to become “AMRiN”. May involve development of a specific project plan	Redeveloped AMRiN Portal

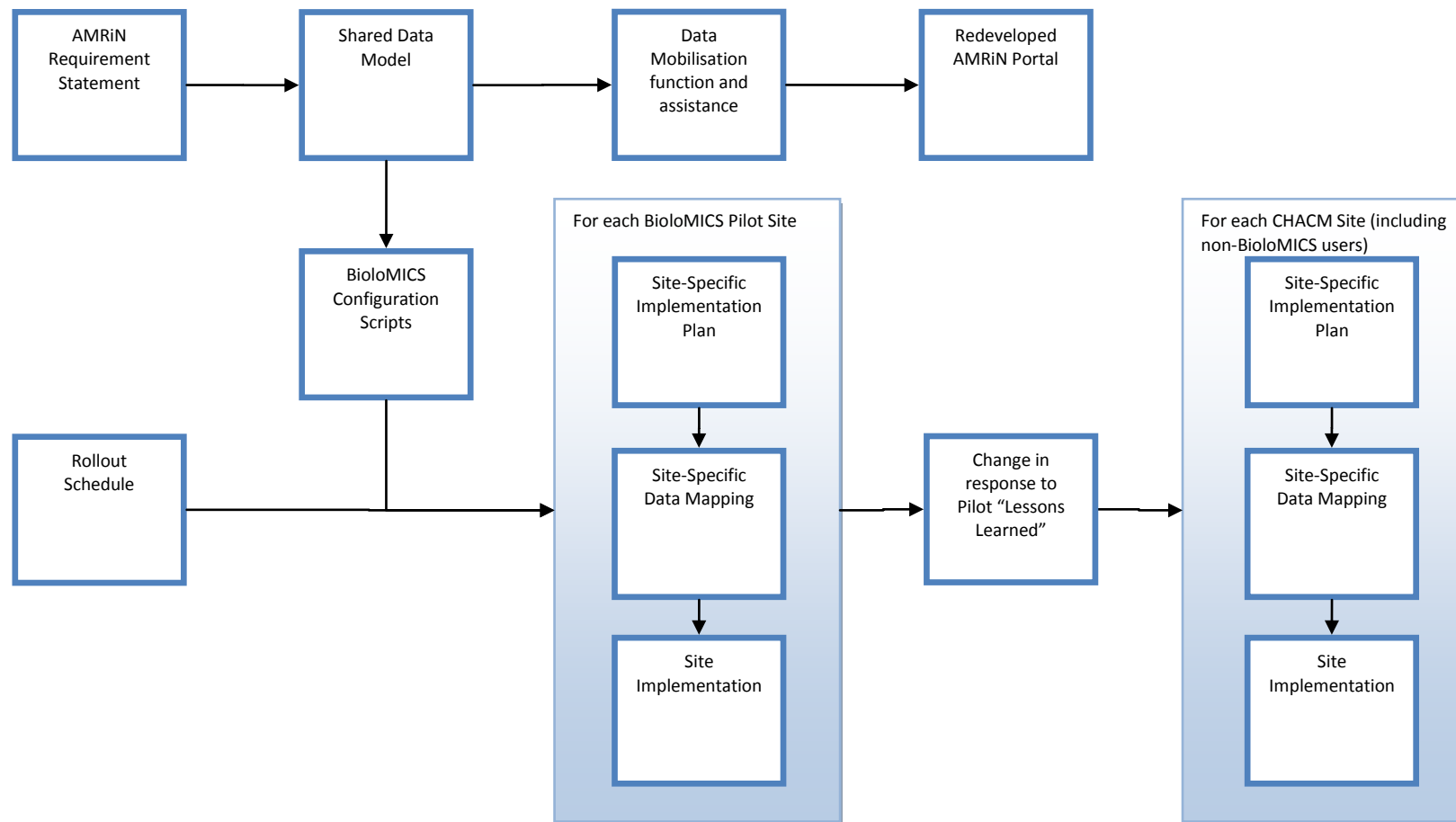


Figure 1 - Key Deliverables Flow Diagram

2 Project Activities

2.1 Confirm AMRiN Data Requirements

In March 2010, ALA proposed a scope for a rebuilt AMRiN to CHACM members. The requirements need be confirmed to reduce risk of:

- Stakeholders having different understandings of the portal's purpose and functions
- ALA delivering missing or unnecessary functionality.

The redeveloped AMRiN portal will be built on common "ALA Hub" infrastructure using shared modules and functions. AMRiN-specific functionality will be provided as appropriate, as for all other hubs.

From an ALA perspective, this activity is part of defining the requirements for all hubs.

The AMRiN Requirement Statement will inform:

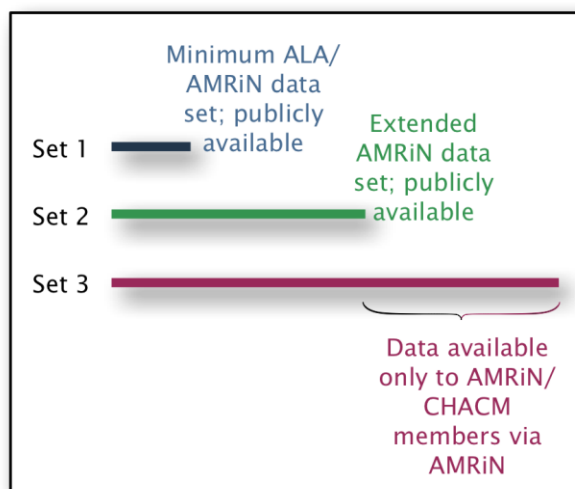
- the "Shared Data Model"
- content/payload for ALA aggregation services
- if ALA assistance is required to extend or redevelop <http://www.amrin.org/>.

Table 2 Confirm AMRiN Requirements

Stage	What	How	Output
1.	Discover: <ul style="list-style-type: none"> • what kind of questions CHACM members would like to be able to answer by querying an (AMRiN) aggregated database 	Collate information from material received/developed to date: <ul style="list-style-type: none"> • ALA AMRiN Proposal • ALA User needs analysis • General communication between ALA personnel, collection owners, CHACM and other members of the micro-organism community • Requirements identified during analysis of the Sensitive Data Service • Responses to previous questionnaires • Deliverables prescribed by NCRIS or EIS funding. 	Collection of raw input material
2.	Draft AMRiN Data Requirements Statement	ALA staff build document	Draft data requirement statement
3.	Review AMRiN Requirements Statement	Email-based review with CHACM Technical Group. If required:	Approved AMRiN Data Requirement Statement

Stage	What	How	Output
		<ul style="list-style-type: none">• Supplementary questionnaires• Site visits with practitioners.	
4.	Develop rollout schedule	Negotiate schedule with CHACM member institutions.	Rollout Schedule

2.2 Develop “Shared Data Model”



Data is expected to be of four types:

- A **core data set** to be shared amongst CHACM members and the wider microorganism community. This data will be published on AMRiN and is the minimum data expected to be shared. Each organisation may or may not already collect/record this data.
- An **extended data set (public)** also to be publicly available via AMRiN. May not be provided by all CHACM members.
- An **extended data set (private)** available via AMRiN only to members of CHACM. May not be provided by all members.
- **Organisation-specific data sets** containing data relevant only to a CHACM member such as the storage location of a specimen or strain. This data would not be shared via AMRiN.

While the focus of the common data model is on the core and extended (public and private) data sets, ALA recognises that the model may need to cater for organisation-specific data sets as well. These elements will be included in the data model specific to each organisation.

The approach to developing the Shared Data Model is to:

- use information already received from participants to identify likely sets of shared data
- identify which MCL elements correspond to the likely sets of shared data
- ask participants to specify amendments through a review process.

Table 3 Common data model development process

Stage	What	How	Output
1.	Discover what data elements CHACM members are willing to share with each other and more widely.	ALA staff collate information from material received/developed to date: <ul style="list-style-type: none"> • ALA AMRiN Proposal • ALA User needs analysis • General communication between 	Collection of raw input material

Stage	What	How	Output
		<p>ALA personnel, collection owners, CHACM and other members of the micro-organism community</p> <ul style="list-style-type: none"> • Responses to previous questionnaires • Deliverables prescribed by NCRIS or EIS funding. 	
2.	Identify possible shared data elements against MCL	<p>ALA Data Analyst identifies MCL elements that correspond to:</p> <ul style="list-style-type: none"> • Core data set (public) • Extended data set (public) • Extended data set (private) 	Mapping against MCL
3.	Draft ALA/AMRiN Shared Data Model	ALA Data Analyst documents the MCL mapping in a format suitable for participant review	Draft ALA/AMRiN Shared Data Model
4.	Review ALA/AMRiN Shared Data Model	<p>Email-based review with CHACM Technical Group.</p> <p>If required:</p> <ul style="list-style-type: none"> • Supplementary questionnaires • Site visits with practitioners. 	Approved ALA/AMRiN Shared Data Model

2.3 Implement BioloMICS Pilot Sites

The pilot objective is to:

- Implement BioloMICS for pilot participants
- Optimise the BioloMICS installation/ migration process.

The pilot doesn't need to demonstrate the project's business case or every feature of its deliverables. As a result, all pilot participants will use BioloMICS. Pilot participants will be selected according to criteria:

- The set of participants represent a broad range of implementation complexity
- Each participant is willing to work with ALA according to the ALA schedule.

Table 4 Pilot process

Stage	What	How	Output
1.	Environmental scan	Initial familiarisation visit to a number of organisations to: <ul style="list-style-type: none"> • understand how they might use BioloMICS • understand the issues they have in implementing BioloMICS • identify the implications for how the ALA should approach implementation • identify potential pilot implementation sites. 	Improved ALA understanding of likely implementation issues
2.	Purchase BioloMICS licenses	ALA staff: <ul style="list-style-type: none"> • Contact each curator and confirm their intent to use BioloMICS • For each curator that intends to use BioloMICS, determine the number of licenses required • Procure the required number of licenses, following regular ALA program purchasing process 	BioloMICS licenses
3.	Develop "Standard Implementation Plan"	ALA staff: <ul style="list-style-type: none"> • Draft a general plan for implementing BioloMICS for a single participant • Conduct an email review with likely early adopters identified during the environmental scan. 	BioloMICS Standard Implementation Plan

Stage	What	How	Output
4.	Arrange preparation of configuration scripts which implement extended model (including core) in BioloMICS.	Liaise with Bio-Aware about translation of logical model to physical configuration; ensure configuration includes a view of core model elements only.	BioloMICS configuration scripts for Version 1.0 extended model
5.	Implement each pilot site	For each pilot participant, execute “Activity for each Pilot Site Implementation” (refer below)	BioloMICS implemented at pilot sites

2.3.1 Activity for each Pilot Site Implementation

The activities summarised below are repeated for each Pilot Site. Detailed activity is documented in a customised implementation plan, which is developed with each participant. The customised plan is based on the “Standard Implementation Plan”, available at <link>.

Table 5 BioloMICS implementation process

Stage	What	How	Output
1.	Plan	Initial planning meeting to: <ul style="list-style-type: none"> • identify implementation issues, eg firewalls • agree responsibilities, eg installation of software, data mapping • agree timings • agree BioloMICS architecture: number and type of licences, users • review data sets to be migrated to BioloMICS to scope data mapping and migration tasks • identify training needs • develop customised implementation plan. 	Participant-Specific Implementation Plan
2.	Map Data	Finalise data mappings in accordance with the Rollout Schedule and Participant-Specific Implementation Plan: <ul style="list-style-type: none"> • Advise and assist with data mapping 	Participant-Specific Data Mapping
3.	Implement	Implement in accordance with the Rollout Schedule and Participant-Specific Implementation Plan: <ul style="list-style-type: none"> • support development and testing activity • assist with data migration, • install software as necessary (local IT staff may chose to do this) • create database tables as necessary (local 	Participant Implementation

Stage	What	How	Output
		data manager may chose to do this) • provide training.	
4.	Document “Lessons Learned”	ALA and pilot participants identify issues and activities that should be treated differently for future BioloMICS implementations. ALA staff draft a summary document, and email to participants for review.	“Lessons Learned” document
5.	Change in response to “Lessons Learned”	ALA manages any changes required to: • data models • standard configuration scripts • standard implementation plan.	Managed changes

2.4 Mobilise data into ALA/AMRiN

Mobilisation is the process of exporting data from an institution's databases and transporting it to ALA.

Each institution is likely to require different mobilisation solutions.

Preliminary analysis suggests it may be possible to use an 'export' function from either BioloMICS or MySQL to export data. As a result, the activity stream "BioloMICS Aggregation Component" is listed separately. If the 'export' function cannot be used, the activity stream "Non-BioloMICS Aggregation Component" covers both participant classes.

2.4.1 BioloMICS data mobilisation

Stage	What	How	Output
1.	Define data flow and trigger event/schedule for BioloMICS participants	Collaborative effort involving: <ul style="list-style-type: none"> ALA Architect Bio-Aware ALA Data Analyst 	Activity Model
2.	Define interface for BioloMICS participants	<ul style="list-style-type: none"> Payload: ALA Architect selects out-of-the-box "export" parameters Wrapper: ALA Architect defines transport mechanism for the CSV implementation 	Message Definition
3.	Develop data mobilisation component for BioloMICS participants	ALA Technical Lead develops service that implements the Activity Model and Message Definition, and stores information ready for use in AMRiN and ALA.	AMRiN/ALA Data Mobilisation Component
4.	Implement BioloMICS configuration	<ul style="list-style-type: none"> Ensure Data Mobilisation component is enabled in the "standard installation" Update any installations that may already be complete 	Revised Standard BioloMICS Configuration Script Updated BioloMICS installations in institutions

2.4.2 Non-BioloMICS data mobilisation

Stage	What	How	Output
5.	Define data flow and trigger event/schedule for non-BioloMICS participants	Collaborative effort involving: <ul style="list-style-type: none"> ALA Architect Institution IT staff ALA Data Analyst 	Activity Model

Stage	What	How	Output
6.	Define interface for non-BioloMICS participants	ALA Architect, Institution IT staff and ALA Data Analyst define: <ul style="list-style-type: none"> • Payload • Transport mechanism • Any data conversion/mapping required to be performed by ALA during the mobilisation process 	Message Definition
7.	Develop data mobilisation component for non-BioloMICS participants	ALA Technical Lead develops service that implements the Activity Model and Message Definition, and stores information ready for use in AMRiN and ALA.	AMRiN/ALA Data Mobilisation Component

2.5 Implement Each CHACM Member Institution

The activities summarised below are repeated for each Pilot Site. Detailed activity is documented in a customised implementation plan, which is developed with each participant. The customised plan is based on the “Standard Implementation Plan”, available at [<link>](#).

Table 6 Repeated Implementation process

Stage	What	How	Output
1.	Plan	<p>Initial planning meeting to:</p> <ul style="list-style-type: none"> • identify implementation issues, eg firewalls • agree responsibilities, eg installation of software, data mapping • agree timings • develop customised implementation plan. <p>If the participant is using BioloMICS:</p> <ul style="list-style-type: none"> • agree BioloMICS architecture: number and type of licences, users • review data sets to be migrated to BioloMICS to scope data mapping and migration tasks • identify training needs. 	Participant-Specific Implementation Plan
2.	Map Data	<p>Finalise data mappings in accordance with the Rollout Schedule and Participant-Specific Implementation Plan:</p> <ul style="list-style-type: none"> • Advise and assist with data mapping 	Participant-Specific Data Mapping
3.	Implement	<p>Implement in accordance with the Rollout Schedule and Participant-Specific Implementation Plan:</p> <ul style="list-style-type: none"> • Support development and testing activity <p>If the participant is using BioloMICS:</p> <ul style="list-style-type: none"> • assist with data migration, including data mapping • install software as necessary (local IT staff may chose to do this) • create database tables as necessary (local data manager may chose to do this) • provide training. 	Participant Implementation

3 Project Operation

The project will use the same tools and processes as other ALA components.

3.1 Quality and Configuration Management

- Final versions of all products are published through <http://www.ala.org.au/>
- Plans and related material managed through Google docs and CSIRO fileshare
- Design, code and test material managed through the AtlasLivingOz wiki, using the “Development” processes
- Production issues managed through the AtlasLivingOz wiki, using the “Service Desk” processes.
- Documents and other configurable items that need to be distributed with stakeholders will be emailed directly, or made available through Google Docs.

3.2 Communication Management

Document and information exchange:

- within the ALA team will rely on email, AtlasLivingOz wiki, Google docs and CSIRO fileshare.
- between ALA and CHACM will rely on email
- between ALA and participant institutions will rely on email, telephone and site visits.
- between ALA and Bio-Aware will rely on email and telephone.

3.3 Reporting

- Project Manager will make monthly Status Reports using standard ALA reporting processes.
- Project Manager will make Exception Reports immediately to the ALA Program Director.
- ALA Communications Officer will manage Media Reports and requests and information about ALA generally.

3.4 Resources

ALA will allocate the following resources at least to implementing BioloMICS:

- Bryan Kalms – project oversight, implementation analysis, resource management
- Nathalie van de Wiele – BioloMICS specialist, implementation analysis, software installation, data mapping, data migration, user training, user support
- Tania Volk – IT specialist, resolve IT issues, implementation scheduling and management
- Bryn Kingsford/Miles Nicholls – data specialists, development of data migration scripts
- Matt Branford – business analysis, project management, scheduling, coordination.

3.5 Post-Project - Ongoing Operation of Delivered Systems

ALA will provide ongoing support until funding expires June 2012 for:

- data mobilisation
- AMRiN (infrastructure and functionality only)
- ala.org.au.

Please note that part of ALA Program Activity in 2011 is to recommend governance and operation arrangements for operations after 2012.

Bio-Aware will provide support under licensed user agreement with each institution for BioloMICS during the currency of those licences.

CHACM will provide ongoing support for the content of AMRiN, management of any access controls and all other non-infrastructure components.

Individual institutions will remain responsible for future licensing of BioloMICS and the ongoing support of their installation, including ongoing supply of data to ALA/AMRiN.