

# ASReml moving forward

## Introducing Echidna

Arthur Gilmour

December 2018

# Outline

- Background
- State of Play
- Areas to address
- Echidna

# Background

- I wrote many programs for NSW DPI biometricians (1970-1995)
- These formed the basis for ASReml (1996-)
- Robin provided the basic algebra and many algorithmic ideas
- Brian Cullis raised many of the issues and validated the solutions.
- The intended scope was to analyse experimental data
  - spatial analysis of field trials
  - designed experiments
  - multi-environment trials
  - animal breeding data
  - multivariate data.

# Popularity

- ASReml quickly became internationally accepted because
  - it implemented the Average Information REML algorithm
  - using sparse matrix methods.
- ASReml can fit large sparse models quickly
- A variety of model functions and variance structures enable a wide range of models to be fitted.

# Software development is an ongoing process.

- Changing demands
- Hardware: memory and processors
- Compilers and libraries
- Better algorithms
- Models: HGLM and Genomic relationship matrices
- Syntax improvements
- Management:
  - DPI and Rothamsted (RR) contracted VSN to retail ASReml
  - DPI and RR sold ASReml to VSN in 2010 and 2012
  - ARG, BC and RT continued to support ASReml but

## So A fresh start.

- ARG with encouragement from David Butler and Brian Cullis
  - started a complete rewrite
    - under the name Echidna in March 2017.
  - The core code is 20 years old
  - Memory management style needed to change
  - Whole structure needed reorganization
- Clean foundation for implementing new algorithms for Genomic data
- ARG agreed to migrate advantageous innovations to ASReml

# Main issues

- User interface
  - more coherent organization of options.
- Memory organization
- Equation ordering for sparse processing
- Third party libraries
- Relationship of Echidna to VSN

# Equation ordering

- The pattern of filled elements in the mixed model equations:
  - pattern due to the design
  - pattern due to correlation pattern in relationship matrices
  - pattern due to spatial correlation of residuals
- These interact so the optimum equation ordering to minimise infill is not trivial.
- ASReml now
  - absorbs sets of dense equations without data
  - absorbs parallel equations as a set
  - determines an order by simulating the process choosing an equation with minimal infill at each step.
- This is an expensive process but performed better than Metis.



## Two sparse matrix storage modes

- ASReml half stores the matrix lower triangle row-wise
- as a Link-List
  - Each cell consists of the value, the Column Number and the address of the next cell
  - Easy to insert new cell
  - Performs poorly if many cells in the row.
- as an indexed list
  - Non zero cells stored sequentially with a parallel list of column numbers.
  - Can not insert new cells
  - Efficient

## Symbolic absorption (KM)

- Link list used when need to insert cells (form  $\mathbf{C}$  matrix)
- Symbolic absorption implemented in Echidna and ASReml 4.2
  - determines pattern of infill first and remembered
  - so that an indexed list can be used in the absorption step
  - is substantially faster for non-sparse problems

## Third party libraries

- Brian Cullis (UoW) has a student reviewing approaches

# Echidna

- A standalone program like ASReml 4
  - Most ASReml job files will run directly in Echidna
  - however there are some syntax changes
- Linux and Windows builds are now available
- The intention at the start was that DB would develop an R product based on the Echidna
- runs the model in 60% of the standard ASReml test suite
- has VPREDICT and basic PREDICT functionality
- has GLMM functionality
- Basic User Guide (pdf and html)

# Outstanding issues

- VSN: What do they own?
- Licensing of Echidna?
- Access to Source code?
- R version?

## Current Access and timeframe

- Users need to register
- Free for non-commercial use (education and public good research)
- Available to ASReml licencees
- Discussion with VSN continuing
- Development continuing on Stage 1: 80% ASReml functionality
  - Stage 2: match ASReml 4.2 performance
  - Stage 3: extend functionality to meet current new demands

# Conclusion

- ASReml 4.2 (standalone) is substantially faster for slow (large, dense) jobs than ASReml 4.1.
  - <http://www.asreml.co.uk>
- ASReml R: R code redeveloped; based on 4.1 code
- Echidna is available for trial and feedback.
  - It is not yet as fast as ASReml 4.2.
  - <http://www.EchidnaMMS.org>
- I acknowledge the grace of God through Jesus Christ in facilitating the development of ASReml and Echidna.
  - [arthur.gilmour@cargovale.com.au](mailto:arthur.gilmour@cargovale.com.au)
  - <http://www.cargovale.com.au>

# The Echidna

- The Australian marsupial spiny ant eater



- Arthur is contracted by VSN to support ASReml.