



FAIR and Accessible Software as a Reproducible Service

Geoffrey Squire

May 2019

Introduction

1. Why we need a Solution Centre
2. What does it do?
3. Working with Virtual Laboratories

Why?

- Scientific software and solutions need to be *useful*:
 - Make them FAIR
 - Trustable
 - Citeable
 - Usable
- Just publishing is easy (e.g. GitHub, PyPI, download)
- Needs to be easy to publish so it is easy to use
 - Make it easy for client software to automate

What does it do?

The Solution Centre makes published software more useful by improving FAIRness and enabling automation of scientific solutions

Information Model

- Description and other metadata
- Persistent URI
- Version
- Relationships between entries
- Provenance
- Execution environment / dependencies

Publication

- Authenticate
- Draft
- Optional review step
- Optional digital signature
- Publish



What does it do?

The Solution Centre makes published software more useful by improving FAIRness and enabling automation of scientific solutions

Information Model

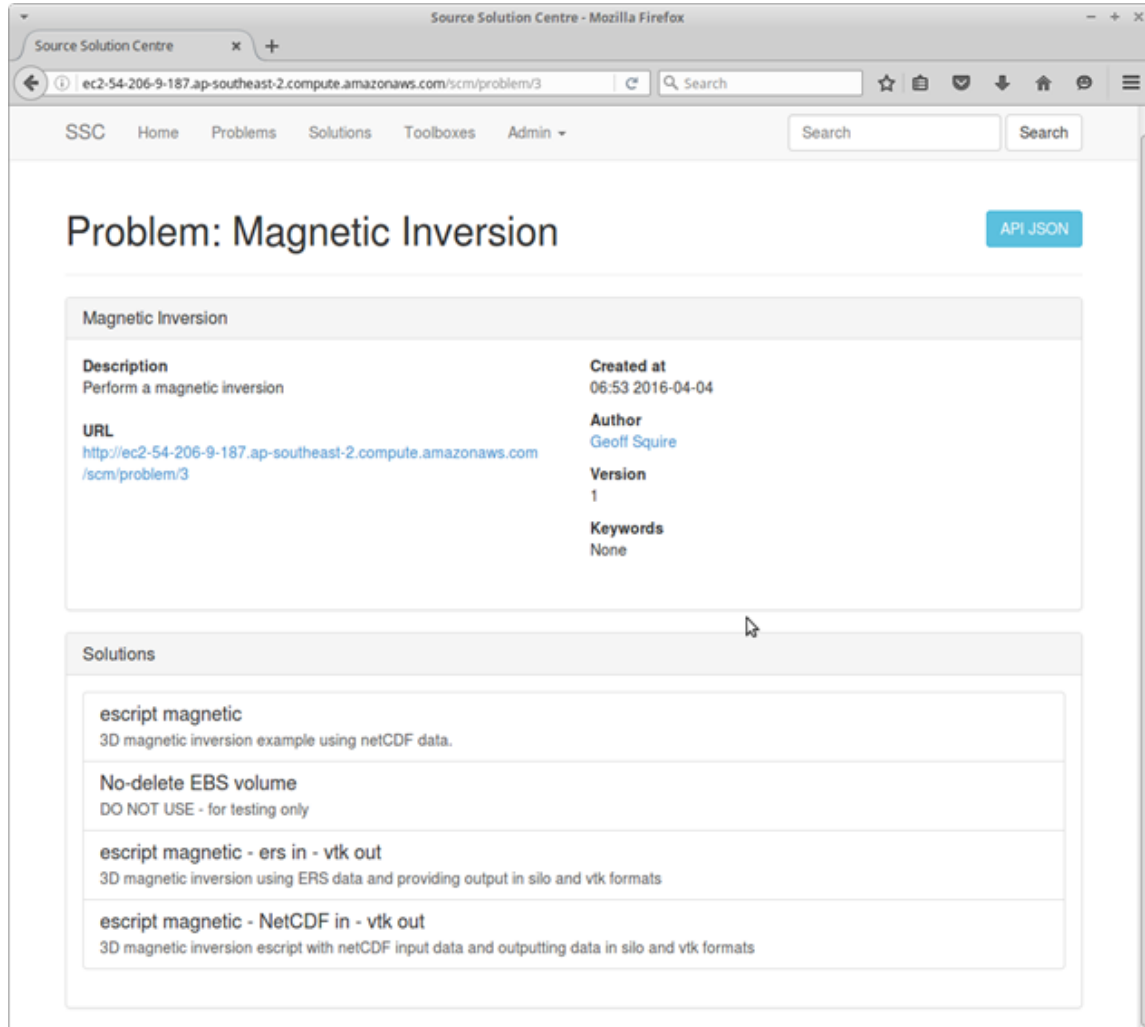
- Description and other metadata
- Persistent URI
- Version
- Relationships between entries
- Provenance
- Execution environment / dependencies

Publication

- Authenticate
- Draft
- Optional review step
- Optional digital signature
- Publish



Everyone has Problems



The screenshot shows a web browser window titled "Source Solution Centre - Mozilla Firefox". The address bar shows the URL "ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/problem/3". The page has a navigation menu with "SSC", "Home", "Problems", "Solutions", "Toolboxes", and "Admin". A search bar is present. The main content area is titled "Problem: Magnetic Inversion" and includes a blue "API JSON" button. Below the title is a table with the following details:

Magnetic Inversion	
Description Perform a magnetic inversion	Created at 06:53 2016-04-04
URL http://ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/problem/3	Author Geoff Squire
	Version 1
	Keywords None

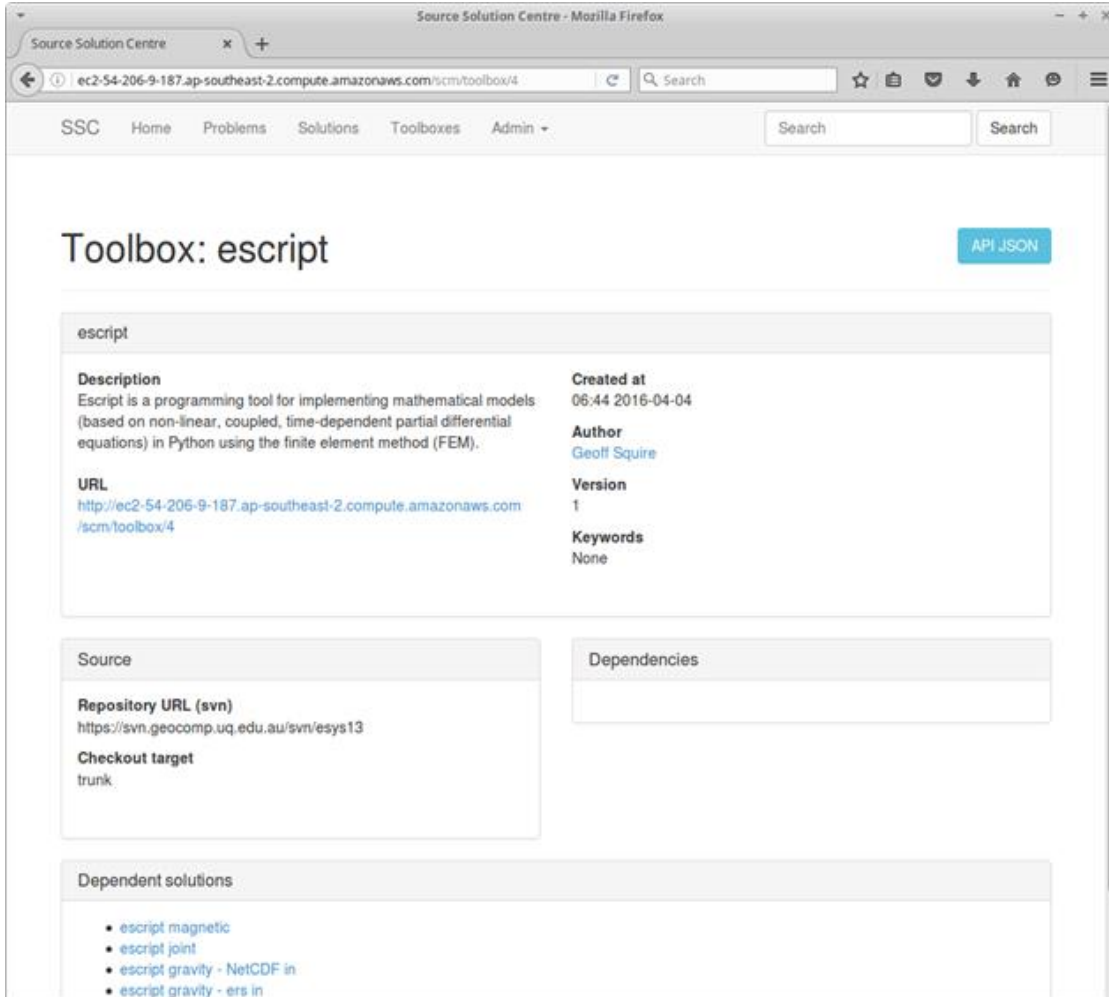
Below the problem details is a "Solutions" section containing four entries:

- escript magnetic**
3D magnetic inversion example using netCDF data.
- No-delete EBS volume**
DO NOT USE - for testing only
- escript magnetic - ers in - vtk out**
3D magnetic inversion using ERS data and providing output in silo and vtk formats
- escript magnetic - NetCDF in - vtk out**
3D magnetic inversion escript with netCDF input data and outputting data in silo and vtk formats

Entry type: Problem

- Description of a user's problem
- Classifies Solutions and Toolboxes
- Captures what users did

Someone writes software...



The screenshot shows a web browser window titled "Source Solution Centre - Mozilla Firefox" with the URL "ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/toolbox/4". The page header includes navigation links for "SSC", "Home", "Problems", "Solutions", "Toolboxes", and "Admin", along with search boxes. The main content area is titled "Toolbox: escript" and features a blue "API JSON" button. Below the title, there are several sections: "Description" (describing escript as a programming tool for mathematical models), "URL" (a link to the toolbox), "Created at" (06:44 2016-04-04), "Author" (Geoff Squire), "Version" (1), and "Keywords" (None). There are also sections for "Source" (Repository URL and Checkout target) and "Dependent solutions" (listing other toolboxes like escript magnetic, joint, gravity - NetCDF, and gravity - ers).

Entry type: Toolbox

- Describes a software environment
- Links to published software
- Describes implementation
 - Puppet module
 - Image snapshot in the cloud
 - HPC configuration
 - Dependencies
- Dependencies can be:
 - Toolbox(es)
 - Python modules on PyPI
 - Puppet modules on PuppetForge

...and someone solves the Problem

The screenshot shows a web browser window titled 'Source Solution Centre - Mozilla Firefox'. The address bar shows the URL 'ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/solution/4'. The page content includes a navigation menu with 'SSC', 'Home', 'Problems', 'Solutions', 'Toolboxes', and 'Admin'. Below the menu, the page title is 'Solution: escript magnetic' with an 'API JSON' button. The main content area is divided into three sections: 'escript magnetic', 'Problem', and 'Toolbox'. The 'escript magnetic' section contains a description, URL, created at date, author, version, and keywords. The 'Problem' section contains a name, description, and URL. The 'Toolbox' section contains a name, description, and URL.

escript magnetic	
Description 3D magnetic inversion example using netCDF data.	Created at 06:54 2016-04-04
URL http://ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/solution/4	Author Geoff Squire
	Version 1
	Keywords None

Problem
Name Magnetic Inversion
Description Perform a magnetic inversion
URL http://ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/problem/3

Toolbox
Name escript
Description Esript is a programming tool for implementing mathematical models (based on non-linear, coupled, time-dependent partial differential equations) in Python using the finite element method (FEM).
URL http://ec2-54-206-9-187.ap-southeast-2.compute.amazonaws.com/scm/toolbox/4

Entry type: Solution

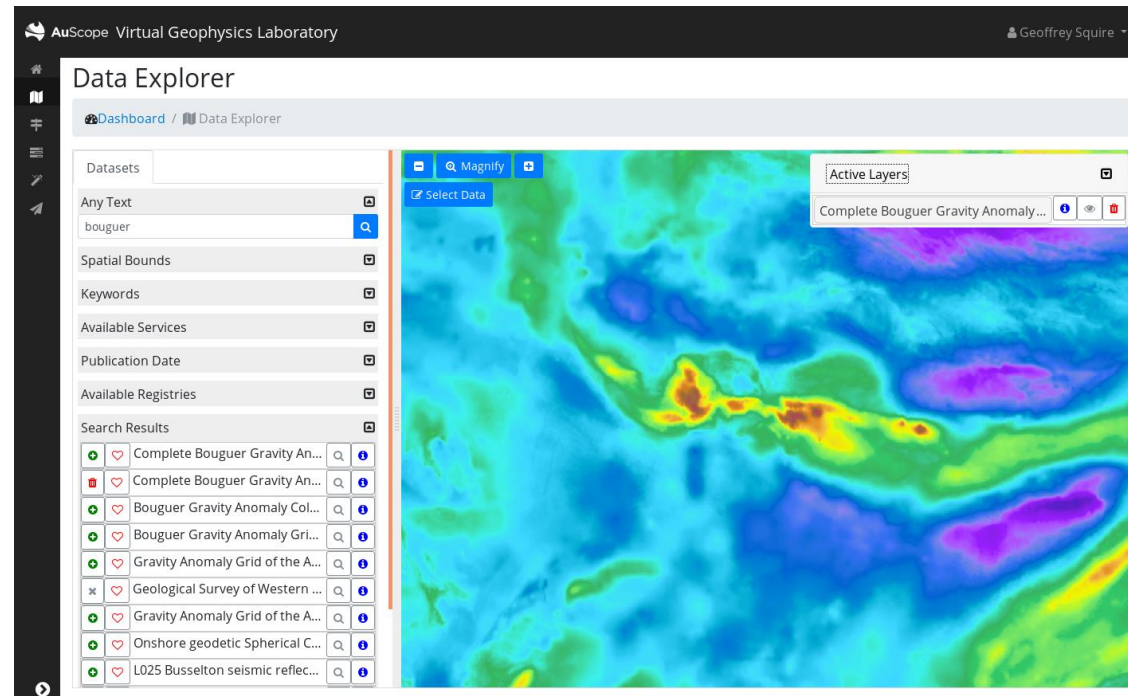
- Describes a specific workflow
- Solves a Problem
- Depends on a Toolbox
- Links to template
- Code that implements the workflow
- Published in the usual way
- Specifies inputs and outputs

Useful Solutions are the best kind

The Solution Centre enables clients to:

- Find relevant solution(s)
- Assemble software environment that can execute the chosen solution
- Identify and connect inputs required for the solution
- Execute the solution!

Clients can be users, but the machine readable information model allows for client *applications*.

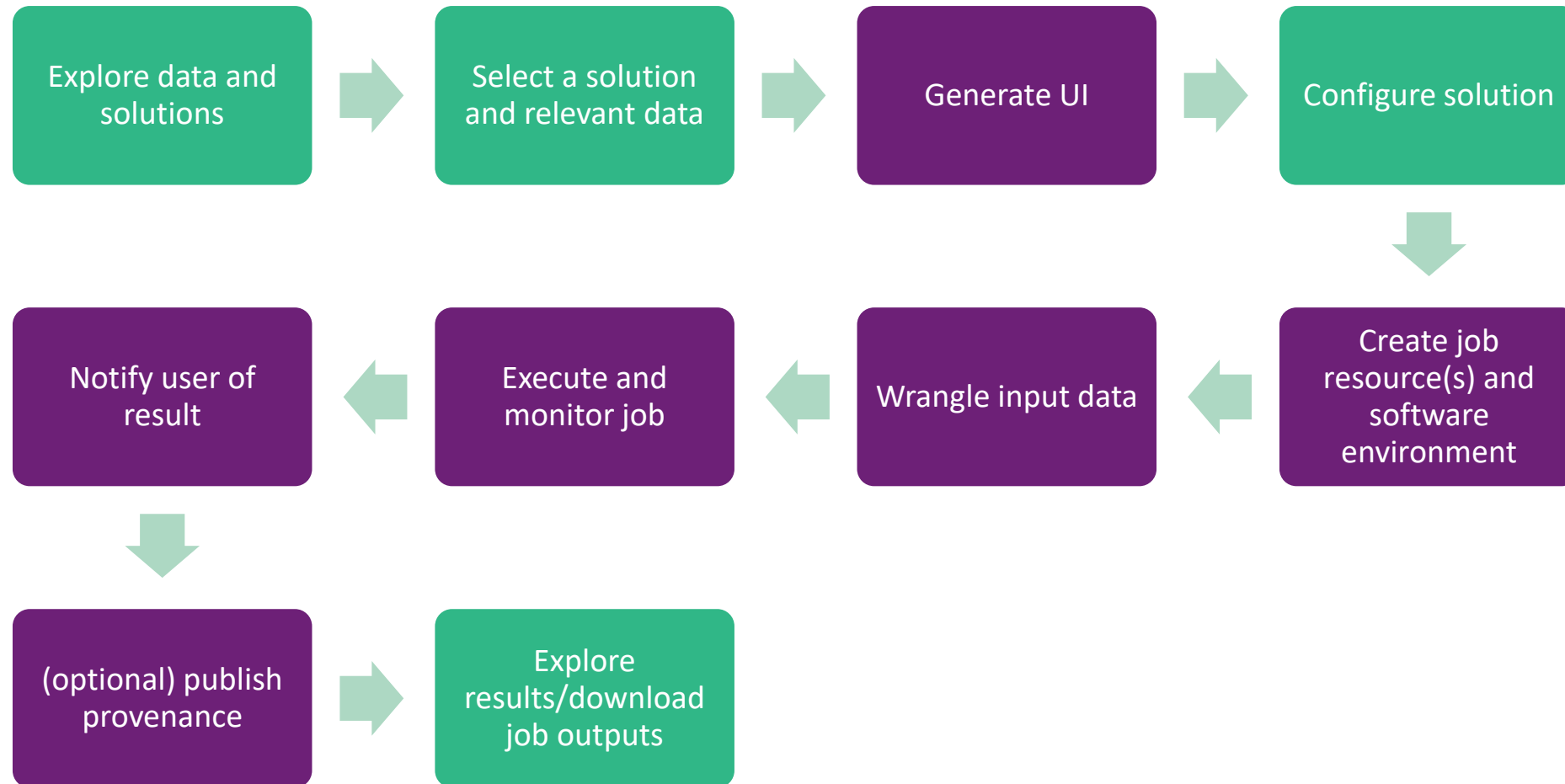


Working with Virtual Laboratories

- Virtual Laboratories (VLs) can be clients of the Solution Centre
 - Published solutions immediately available in VLs
 - Provides a task/domain appropriate UX
 - Integrates data services and software solutions
- Provenance is recorded
 - Solution Centre and VGL provide PROV-O descriptions
- Demonstrated in the Virtual Geophysics Laboratory
 - <https://vgl.auscope.org>
 - Run solutions on Nectar and Amazon clouds, or the NCI Raijin HPC facility



Virtual Laboratory workflow



- Home
- Data
- Solutions
- Jobs
- Create Job
- Contact us

Solutions

Problem

- Gravity Inversion
- Run python
- Joint inversion
- Magnetic Inversion
- AEM

Gravity Inversion

Perform a gravity inversion.

[More info...](#)

escript gravity - NetCDF in

escript gravity inversion for NetCDF data with output in silo format only. Inversion uses eScript solver and functionality from esys.downunder module.

[Select](#) [Details...](#)

Cart

eScript gravity - variable mu

escript gravity inversion for NetCDF data with output in and variable mu

[Select](#) [Details...](#)

- Home
- Data
- Solutions
- Jobs
- Create Job
- Contact us

Solutions

[Copy From Job](#) [Remote Download](#) [Upload File\(s\)](#)

Cart

escript gravity - NetCDF in

[Add a solution](#) [Edit template](#)
[Reset template](#) [View template](#)

escript gravity - NetCDF in

escript gravity inversion for NetCDF data with output in silo format only. Inversion uses eScript solver and functionality from esys.downunder module.

Dataset

gravity_complete_sph_cap_boug_anomaly

The path to a NetCDF input file.

Max Threads

1

The maximum number of execution threads to run (this job will have {0} CPUs).

Max Depth

40000

The maximum depth of the inversion (in meters).

Air Buffer

6000

Buffer zone above data (in meters; 6-10km recommended).

Z Mesh Elements

25

Number of mesh elements in vertical direction (approx. 1 element per 2km recommended).

X Padding



A Software Registry

- The Solution Centre can be used as a registry
- Enable FAIR publication of research software
- Complements existing registries
- Demonstrated in the AuScope Virtual Research Environment
<https://avre.auscope.org>.
- Describes Applications
 - “What solutions can this application run?”
 - “What applications can I use to run this Solution?”

The screenshot shows the 'Available Apps' section of the AuScope Virtual Research Environment. The header includes the AuScope logo, the text 'Virtual Research Environment', and an 'About' link. Below the header, there are six application cards arranged in a 2x3 grid. Each card features a thumbnail image, a title, a brief description, and 'LAUNCH' and 'DETAILS' buttons.

Application Name	Description
ANDS IGSN Minting Application	This service enables users to mint and assign International Geo Sample Numbers (IGSN) identifiers. The IGSN is a resolvable identifier for physical samples from our natural environment and related sampling features associated with research.
AuScope Portal Application	AuScope GRID – Web Based Interface for searching and accessing data, information, imagery, services and applications connected to the Grid.
AusPASS Application	AusPass is a service dedicated to the acquisition, management, and distribution of passive seismological data in Australia.
MT Training Application	(Description not visible in the image)
Underworld2 Training Application	(Description not visible in the image)
VGL Application	(Description not visible in the image)

Describing deployable software

- A Toolbox can describe a deployable service
 - More than just a software environment for batch-style jobs
 - Can be used to specify a long-running service
 - Includes a description of parameters for configuration
- Example 1: Amazon AMI running a web service
- Example 2: Helm chart for running a JupyterHub instance on Kubernetes
- Similar workflow to before:
 1. Choose Toolbox to deploy
 2. Set parameters based on Toolbox specification
 3. Spin up instance and pass configuration
 4. Monitor/control running instance



THANK YOU

Data61

Carsten Friedrich
Team Leader

t +61 2 9123 4567

e carsten.friedrich@data61.csiro.au

w www.data61.csiro.au

Data61

Geoffrey Squire
Software Engineer

t +61 2 6216 7064

e geoffrey.squire@data61.csiro.au

w www.data61.csiro.au

www.data61.csiro.au

