



# Running an Open Source National Research Data Registry

Amir Aryani  
Australian National Data Service

# ResearchData.ANDS.ORG.AU

Research Data Australia

About Collections Parties Activities Services Themes

ands


Search for Research Data

Browse by Subject Area

Browse by Map Coverage

Advanced Search

## What's in Research Data Australia

 **Collections (102928)**  
Research datasets or collections of research materials.

 **Parties (26905)**  
Researchers or research organisations that create or maintain research datasets or collections.

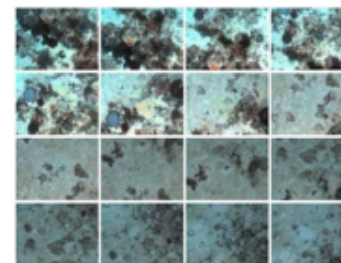
 **Activities (42046)**  
Projects or programs that create research datasets or collections.

## Spotlight on research data

### CATAMI:

#### Collaborative and Annotation Tools for Analysis of Marine Imagery and Video

In recent years, ecologists and biologists have increasingly relied on remote video and image-based methods to examine and monitor marine habitats, particularly those beyond easy diving depth. These data streams are currently being collected by the end user community at sites around Australia using modern image and video based tools including Baited Remote Underwater Video Systems (BRUVS), Autonomous Underwater Vehicle (AUV), Remotely Operated Vehicle (ROV) and Underwater Towed Video (UTV) systems. Transforming raw visual data into quantitative information useful for science and policy decisions requires substantial effort by human experts. The University of Sydney has led a collaborative project to develop common labelling standards, and to transform and collate existing datasets that support marine research across Australia. The



# Agenda

- Open source software
- Metadata
- Business of running a registry

What makes a good  
open source software package



# Make your code understandable / readable



# Make your API easy to use

The screenshot shows a web interface for the ANZSRC Field of Research widget. At the top, there is a green navigation bar with the 'ands Developer Toolbox' logo and links for 'Widgets', 'Web Services', 'Registry Software', and 'Developer Community Blog'. The main content area features a dark background with a code editor on the left and a widget preview on the right. The widget preview is titled 'ANZSRC Field of Research:' and contains a search input field with 'Biol' entered. Below the input is a list of research categories with their corresponding codes and URLs. The 'Soil Biology' category is highlighted with a mouse cursor. At the bottom of the widget, there is a 'Submit' button and a note about autocompletion.

**ands**  
Developer Toolbox

Widgets   Web Services   Registry Software   Developer Community Blog

```
SPACE);  
E);  
title),  
st_title), EXTRIF_NW  
, str_rep  
);  
utor[0]), EXTRIF_NW
```

## Controlled Vocabulary Widget

Instantly add Data Classification capabilities to your data capture tools

[Learn More](#)

### ANZSRC Field of Research:

<b>Biologically Active Molecules</b>	030401
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/030401">http://purl.org/au-research/vocabulary/anzsrc-for/2008/030401</a>	
<b>Characterisation of Biological Macromolecules</b>	030403
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/030403">http://purl.org/au-research/vocabulary/anzsrc-for/2008/030403</a>	
<b>Biological Oceanography</b>	040501
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/040501">http://purl.org/au-research/vocabulary/anzsrc-for/2008/040501</a>	
<b>Soil Biology</b>	050303
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/050303">http://purl.org/au-research/vocabulary/anzsrc-for/2008/050303</a>	
<b>BIOLOGICAL SCIENCES</b>	06
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/06">http://purl.org/au-research/vocabulary/anzsrc-for/2008/06</a>	
<b>BIOCHEMISTRY AND CELL BIOLOGY</b>	0601
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/0601">http://purl.org/au-research/vocabulary/anzsrc-for/2008/0601</a>	
<b>EVOLUTIONARY BIOLOGY</b>	0603
<a href="http://purl.org/au-research/vocabulary/anzsrc-for/2008/0603">http://purl.org/au-research/vocabulary/anzsrc-for/2008/0603</a>	

(autocomplete; begin typing something (e.g. "BIOL"))

[researchdata.ands.org.au/developers](http://researchdata.ands.org.au/developers)



# Community blog



Enrich references across the registry

Link, relationship generation and all metadata generation are Enrich independent as of R12.2

Releases

Widgets

Web Services

Registry Software

Developer Community Blog

## Release 12.2 – Prefix Agnostic Harvester

Posted on April 22nd, 2014

Prior to Release 12.2 the only metadataPrefix for OAI was 'rif'. Even if coded with another prefix the harvester would over ride the provided prefix with 'rif'

As of release 12.2 the metadataPrefix is no longer limited to 'rif'. The prefix that is required when a harvest is created can be configured in the global\_config.php. This also allows the user to map prefixes to existing cross-walks. See example configuration code below.

```
// Provider types
// allow to add different metadataPrefix for OAI harvesting
// also can be configured to run cross walk if needed.

$ENV['provider_types'] = array(
'RIF-CS'=array('prefix'=>'rif','cross_walk'=>''),
'RIF-CS-2'=array('prefix'=>'rif2','cross_walk'=>''),
'OAI-DC'=array('prefix'=>'oai_dc','cross_walk'=>'Dublin Core to RIFCS')
);
```

Once the configuration for a provider has been set the available metadata prefix type will be available for selection in the Data Source admin settings eg 'OAI-DC'. If your cross\_walk is blank then no cross will be executed and will assume that the provided xml is RIF-CS. This will allow users to configure multiple metadata prefixes for RIF-CS.

Despite the fact that prefixes aren't required to be passed to the provider when requesting a feed using the GET method it can still used to identify cross-walks for GET feeds. This means that crosswalks can be defined for both GET and OAI-PMH feeds.

```
$ENV['provider_types'] = array(
'CSV2RIF'=array('prefix'=>'csv','cross_walk'=>'CSV 2 RIF-CS')
);
```

This configuration compliments the existing crosswalk handler and crosswalks can still be added as a class into the crosswalks directory.

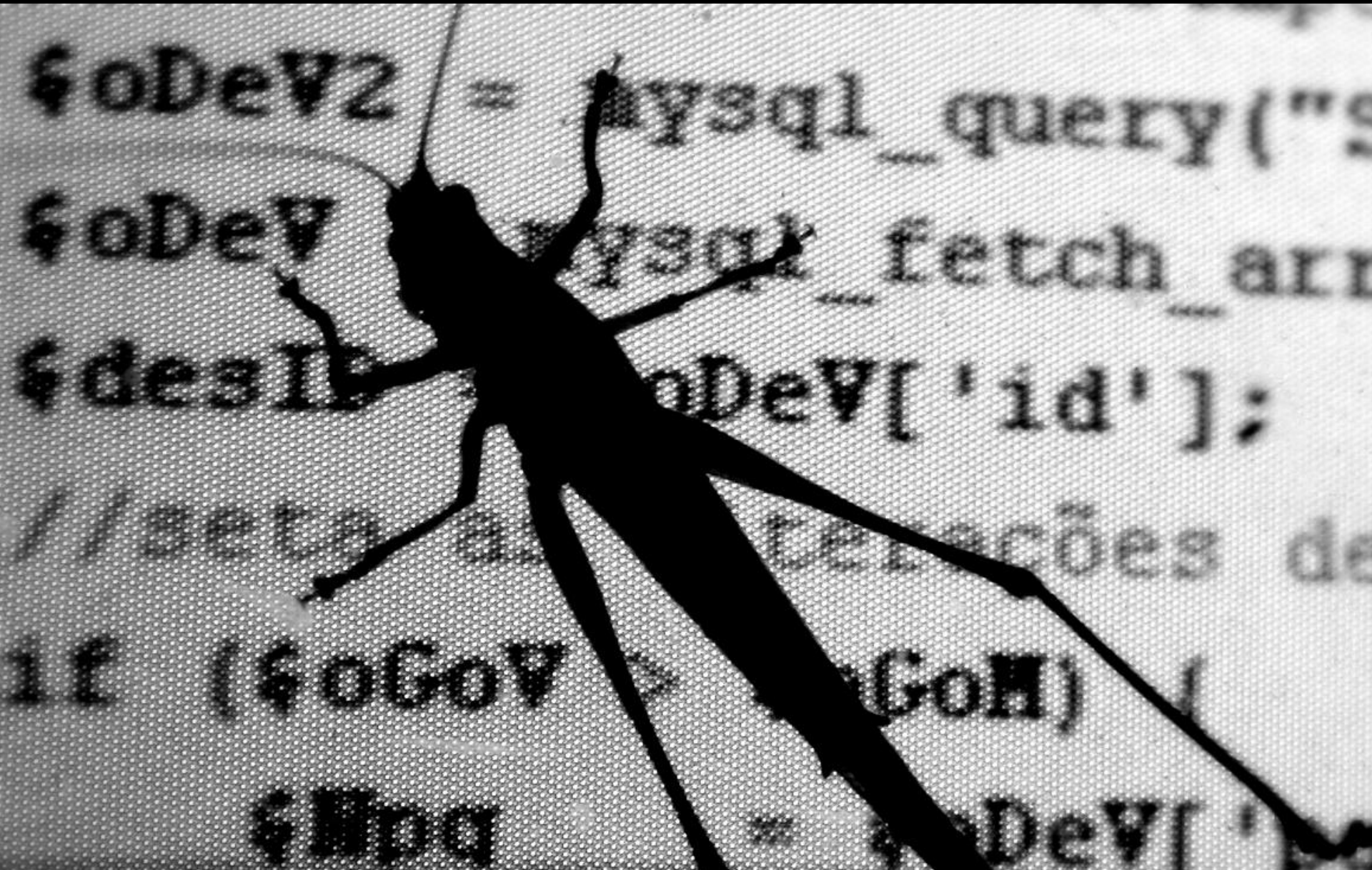
[developers.ands.org.au/](http://developers.ands.org.au/)

**Members of the community  
help each other**





# Software bugs are inevitable



**Software bugs are inevitable**

**How quickly  
can you deal  
with them?**



# Enabling creativity





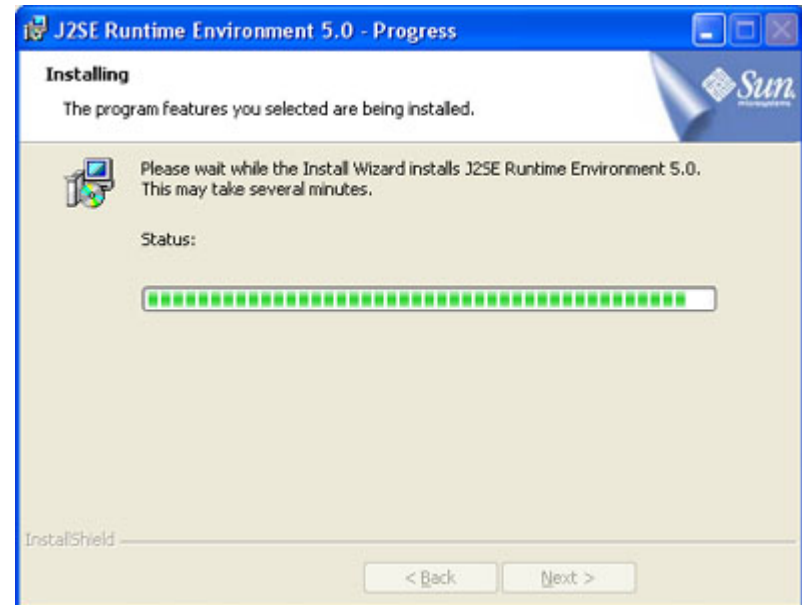
Modular code  
makes it easy to

- change,
- debug,
- fix,

... and enables  
creativity.



**What do I need to do to  
install this?  
... should I press <next>?**



Do we have the  
right metadata schema





# RIF-CS: Registry Interchange Format - Collections and Services

## What's in Research Data Australia



### **Collections (102928)**

Research datasets or collections of research materials.



### **Parties (26905)**

Researchers or research organisations that create or maintain research datasets or collections.



### **Activities (42046)**

Projects or programs that create research datasets or collections.



### **Services (355)**

Services that support the creation or use of research datasets or collections.

[www.ands.org.au/resource/rif-cs.html](http://www.ands.org.au/resource/rif-cs.html)

# Does your model work for everyone?



# Versioning

## RIF-CS 1.5.0:

- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema
  - Activity Schema
  - Collection Schema
  - Party Schema
  - Service Schema
  - Type Definitions
- Controlled Vocabularies
- Schema Documentation

## RIF-CS 1.4.0:

- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema
  - Activity Schema
  - Collection Schema
  - Party Schema
  - Service Schema
  - Type Definitions
- Controlled Vocabularies
- Schema Documentation

## RIF-CS 1.3.0:

- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema

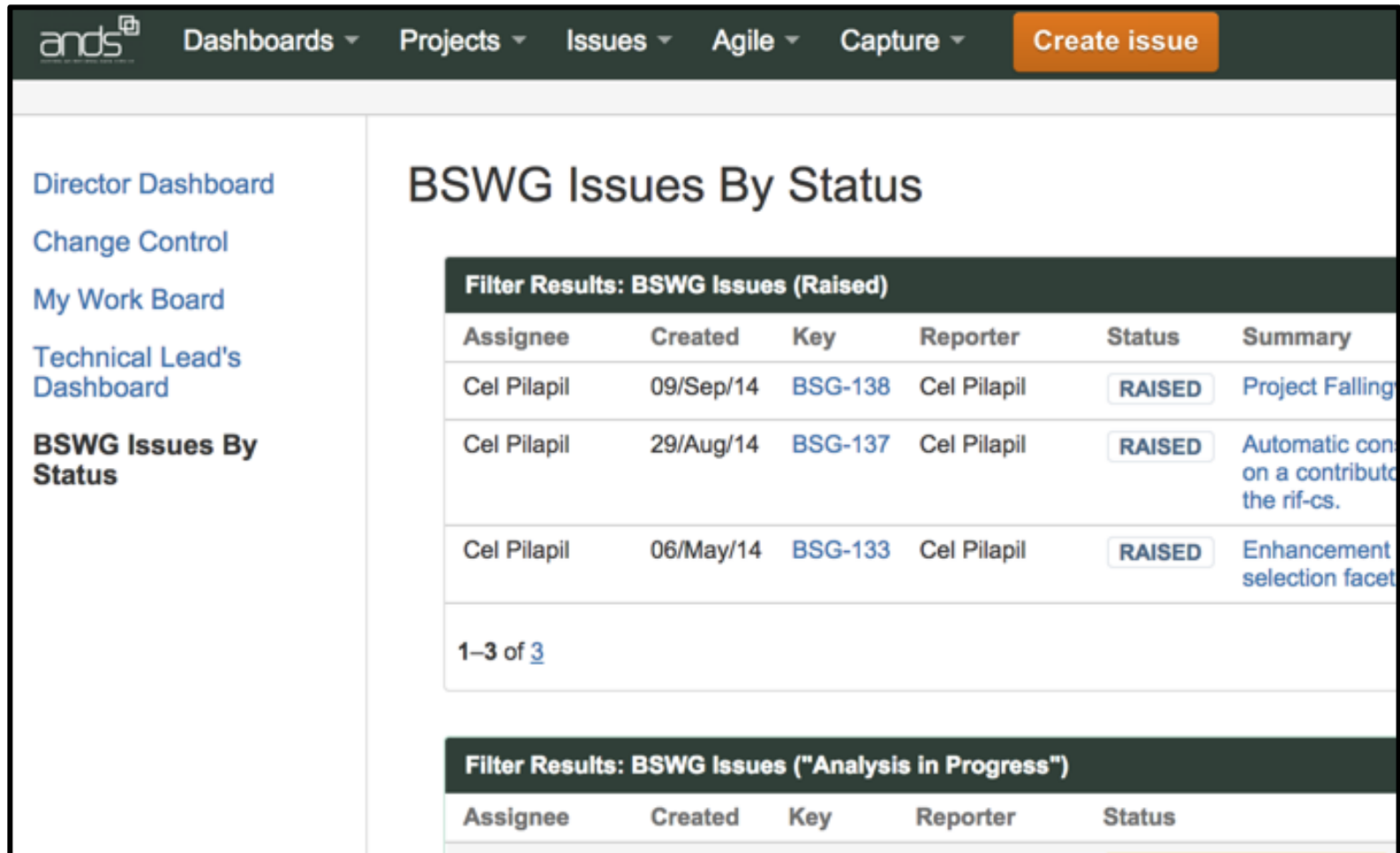
What are the top 3 things to  
know a national registry



# **Working groups: collective decisions are more expensive but more sustainable**



# Tip: use an issue tracking software



The screenshot shows the 'ands' issue tracking software interface. The top navigation bar includes 'Dashboards', 'Projects', 'Issues', 'Agile', and 'Capture', along with a 'Create issue' button. The left sidebar lists various dashboards: 'Director Dashboard', 'Change Control', 'My Work Board', 'Technical Lead's Dashboard', and 'BSWG Issues By Status'. The main content area displays 'BSWG Issues By Status' with a table of issues. The table has columns for Assignee, Created, Key, Reporter, Status, and Summary. Three issues are listed, all with a 'RAISED' status. Below the table, it shows '1-3 of 3' items. At the bottom, there is a section for 'Filter Results: BSWG Issues ("Analysis in Progress")' with a table header for Assignee, Created, Key, Reporter, and Status.

ands Dashboards ▾ Projects ▾ Issues ▾ Agile ▾ Capture ▾ [Create issue](#)

Director Dashboard  
Change Control  
My Work Board  
Technical Lead's Dashboard  
**BSWG Issues By Status**

## BSWG Issues By Status

**Filter Results: BSWG Issues (Raised)**

Assignee	Created	Key	Reporter	Status	Summary
Cel Pilapil	09/Sep/14	<a href="#">BSG-138</a>	Cel Pilapil	RAISED	Project Falling
Cel Pilapil	29/Aug/14	<a href="#">BSG-137</a>	Cel Pilapil	RAISED	Automatic con on a contribu the rif-cs.
Cel Pilapil	06/May/14	<a href="#">BSG-133</a>	Cel Pilapil	RAISED	Enhancement selection facet

1-3 of [3](#)

**Filter Results: BSWG Issues ("Analysis in Progress")**

Assignee	Created	Key	Reporter	Status
----------	---------	-----	----------	--------



**It is important to be agile  
as every contributor can **change** their  
systems at any time.**



**Final remark: “Running a national research data registry is a collaborative effort.”**



# Attribution for the photos

Code creativity: <https://www.flickr.com/photos/tmx-2/6145817075/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMB-chYpBy-51p7BJ-3AvBtV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JEY-9dUbHm-auWNfV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UfZRY-d9A4Eb-7d7ipd-9p77om-buRLHu-5WvBYq-aogX72-4SexF-8PNfWf-eY6Fn-8dpeZZ-8dsw7J-4mtkin-6Rh1p-vW8fo-3ijuU-fd3oYe-7q1fp-aMG5S4-kmhEbE-jmr8PB-azQhJy-icMHWD-4mtkj6-6SgG7q-3hkoRT-7HjkQa-dnQW4g-i3e73w-fprehD-5XzfyW/>

Software bug: <https://www.flickr.com/photos/guitavares/1703252007/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMB-chYpBy-51p7BJ-3AvBtV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JEY-9dUbHm-auWNfV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UfZRY-d9A4Eb-7d7ipd-9p77om-buRLHu-5WvBYq-aogX72-4SexF-8PNfWf-eY6Fn-8dpeZZ-8dsw7J-4mtkin-6Rh1p-vW8fo-3ijuU-fd3oYe-7q1fp-aMG5S4-kmhEbE-jmr8PB-azQhJy-icMHWD-4mtkj6-6SgG7q-3hkoRT-7HjkQa-dnQW4g-i3e73w-fprehD-5XzfyW/>

Community: <https://www.flickr.com/photos/purrrr/126597849/in/photolist-cbR8x-iRL2Xc-cxhBwN-6Lof3x-njiND1-9SAJAG-K83wP-7uqhbU-BvU6G-6KjXVN-28VCp4-4fTR98-nhDCqj-4j8w4p-hTrnRM-ikTP1f-5SzsQ-bkkuEB-8qcf1w-a26aqX-5TQ6Bq-8fBCEC-7R58b8-ogFetv-bzy1PV-7Y7wwk-fGazWu-y59xh-hYY15-bzd5ym-5XudVL-cEwUnU-jaf3S1-bFQKS6-njiRt8-34Wub7-nAyTys-GyMse-3b8udo-hgLxR5-5R5ebY-aZk2Zz-2WxQ3n-7w9QMB-ec7f5T-4Fh97Y-3mzwwqk-7yraFx-8PFLQb-qA3pC>

Lego: <https://www.flickr.com/photos/defjux/8152400914/in/photolist-kfoWq7-aB6Zno-aB6Zs7-dXFmeW-nyUx8B-5YoNCK-hJt7jp-bAuBzn-bPpgo4-bAuBFu-bPpgUr-bAuBjY-bAuBxW-bPpfUH-bAuAV1-bPpfWB-bPpfLg-bPpgsT-bAuCwJ-bPpgQF-bPpgYp-bPpgCk-bAuC1j-bAuB9o-bPpfHR-5gwDbS-bAuCn5-8M2XsG-bAuCrA-bAuBNU-bxU3LK-dqpbLs-ddM4bP-ddM2qi-ddM4GP-ddM6M2-ddM4wM-ddM6Xp-ddM2Hp-ddM4kV-ddM3g1-ddM5eV-ddM8YS-ddM8RW-ddM5pH-ddLZR4-ddM4qR-ddM6my-fGgEFR-7g8DPy/>

Dancers:

<https://www.flickr.com/photos/lecercle/845619594/in/photolist-cKFQSE-jorofj-tpG66-694yuv-nn1mM9-2hJ2fb-4YZNM-p9NuYH-5Sm8G5-6524K8-94NaBt-4oWJP4-9h9RZt-4GJ2tJ-aPVock-f7VPyF-8TiW7t-aCKZKC-ddqWc6-5XwBMM-7TK97G-bB5JDB-99rwUF-4HUZpD-6c3896-cqhbtf-cqhbXq-cqhaRU-4HUZvV-2jcZhz-5LPrLJ-aXVac-7tikoY-6z4PwQ-9fA2ZV-2JmKvD-5ZusNy-4q6kaD-8BEWeW-edfMiF-9taSq6-58gpGk-jdQLuV-8A7cs7-fDntzY-ddqW9T-sp4g2-AL6RM-2HvfzL-4Z5FMC>