A Framework for the Preservation of a Docker Container

Iain Emsley, David de Roure
12th International Digital Curation Conference (IDCC17), 20–23 February 2017
• Uses of Containers
• Issues with existing approaches
• A Framework
• Integration Prototype overview
• Providing a Provenance of Container Build
• Capturing the Dockerfile details
• Capturing Container Build Environment and Processes
• Capturing the Workflow
• Integrity
• Conclusion
Current uses of containers

• Being used as repositories for pipelines
  – Bio-sciences especially and Radio Astronomy
• Dockerfiles as a recipe
• Repeatable building of the Container
• Tagging containers in repository for versioning

© Loiol, CC BY ND NC 2.0, https://www.flickr.com/photos/lioil/1198857098
Issues with existing approaches

• Isn’t a container software itself?

• Linking this to a pipeline

• Repeatability and reproducibility issues
## A Framework

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software Composition</strong></td>
<td>• Description of the components in the container</td>
</tr>
<tr>
<td><strong>Provenance and Ownership</strong></td>
<td>• Processes used to build the container.</td>
</tr>
<tr>
<td></td>
<td>• The maintainer of the Dockerfile</td>
</tr>
<tr>
<td><strong>Container Environment</strong></td>
<td>• The environment used to construct the container</td>
</tr>
<tr>
<td><strong>Workflow</strong></td>
<td>• The workflow with the configuration details for the pipeline</td>
</tr>
</tbody>
</table>
SKA Integration Prototype

Developer → Github → Jenkins → Build → Cluster → Container Store
Providing a Provenance of Container Build

- Not all software can be archived for all time
- Linking artefacts and owners
- Capturing the environment and processes
Capturing the Dockerfile details
Capturing Container Build Environment and Processes
Capturing the Workflow
Integrity

• Build System
  Completeness and accurate processes used?

• Software Processes
  Conformance to project standards
Conclusions

• Capturing the Dockerfile, Environment and Workflow Processes
  
  Developing the Dockerfile as recipe

• Potential challenges in retrieving parameters

• Integrity
  
  Build Process
  Software Process

• Software Citation
Thanks for listening

Iain Emsley
Oxford e-Research Centre
iain.emsley@oerc.ox.ac.uk
@iainemsley
Acknowledgements

Benjamin Mort, Fred Dulwich, and Wes Armour, Oxford e-Research Centre
References

Emsley, I and De Roure, D, A framework for the preservation of a Docker container, International Digital Curation Conference, https://ora.ox.ac.uk/objects/uuid:f567f27a-4efb-431e-abcb-07b6e8c03ce2

Emsley, I, SDP Memo 29: Integration Prototype Build and Deployment Infrastructure Design