The big challenge for the IRD: curating and maintaining data

• The lesson from OpenDOAR is that a centrally-managed dataset is too expensive to maintain properly. One or two people cannot add, maintain and periodically check so many repositories.

• So, the IRD has to find a way to ensure that the data is:
  
  • easy and quick to add
  
  • easy and quick to update (automated where possible)
  
  • easy and quick to verify (automated where possible)
  
  • trusted by users!
A better approach to data curation - the distributed model

- The IRD will be centrally hosted and managed by COAR

- However, the data will be added, updated and managed by the repository networks, in a distributed model of responsibility

  - Each network of repositories will have a designated Responsible Party

  - The Responsible Party is an organisation with responsibility for ensuring that the data about the repositories in their network is up-to-date and accurate
Responsible Parties and Authorised Curators

- The *Responsible Party* may delegate the actual checking and updating work to *Authorised Curators*.

- An *Authorised Curator* is a person with a user account on the IRD with the ability to update the IRD record for one or more repositories.

- In some cases, this could be the repository's manager, or it could be someone working directly for the *Responsible Party*.

- The *Authorised Curator* is not necessarily a person - it can be a software process run by the *Responsible Party* to automate curation of their IRD records.
Distributed curation model
Current status and next steps

• Software is 90% developed, but not yet properly tested

• Some "seed" data has been ingested from existing directories
  • This data has been checked and significantly improved through automated checking processes

• We have identified Responsible Parties for ~50% of the repository records

• Next steps:
  • Convene small group of early Responsible Parties and their Agents
  • Testing!
  • Identify more Responsible Parties for remaining repositories