

# Using Pure as data catalogue and (optionally) as data repository

13 February 2015 – RDM Readiness workshop London  
Hardy Schwamm: [h.schwamm@lancaster.ac.uk](mailto:h.schwamm@lancaster.ac.uk)

# Overview

---

- CRIS landscape in the UK
- How can Pure help with RDM?
- Pure User Survey

# The CRIS Landscape

- 160 Universities and HEI organisations in the UK
- Of top 50 (by research income), 43 have a CRIS
- CRIS options
  - Elsevier - Pure
    - 24 Pure users in the UK & Ireland
      - 16 Pure Research Portals
  - Symplectic – Elements: 20 (?)users
  - Thomson Reuters – Converis: 13 (?) users
  - ePrints + extensions
  - Inhouse

## **Pure Users UK & Ireland**

Aberystwyth University

[Aston University](#)

Glasgow Caledonian University

[Heriot-Watt University, Edinburgh](#)

Institute of Education, London

[King's College London](#)

[Lancaster University](#)

[Queen's University Belfast](#)

[Royal Holloway, University of](#)

[London](#)

[University of Aberdeen](#)

University of Bath

University of Birmingham

University of Bristol

[University of Dundee](#)

University of East Anglia

[University of Edinburgh](#)

[University of Hertfordshire](#)

University of Manchester

University of Portsmouth

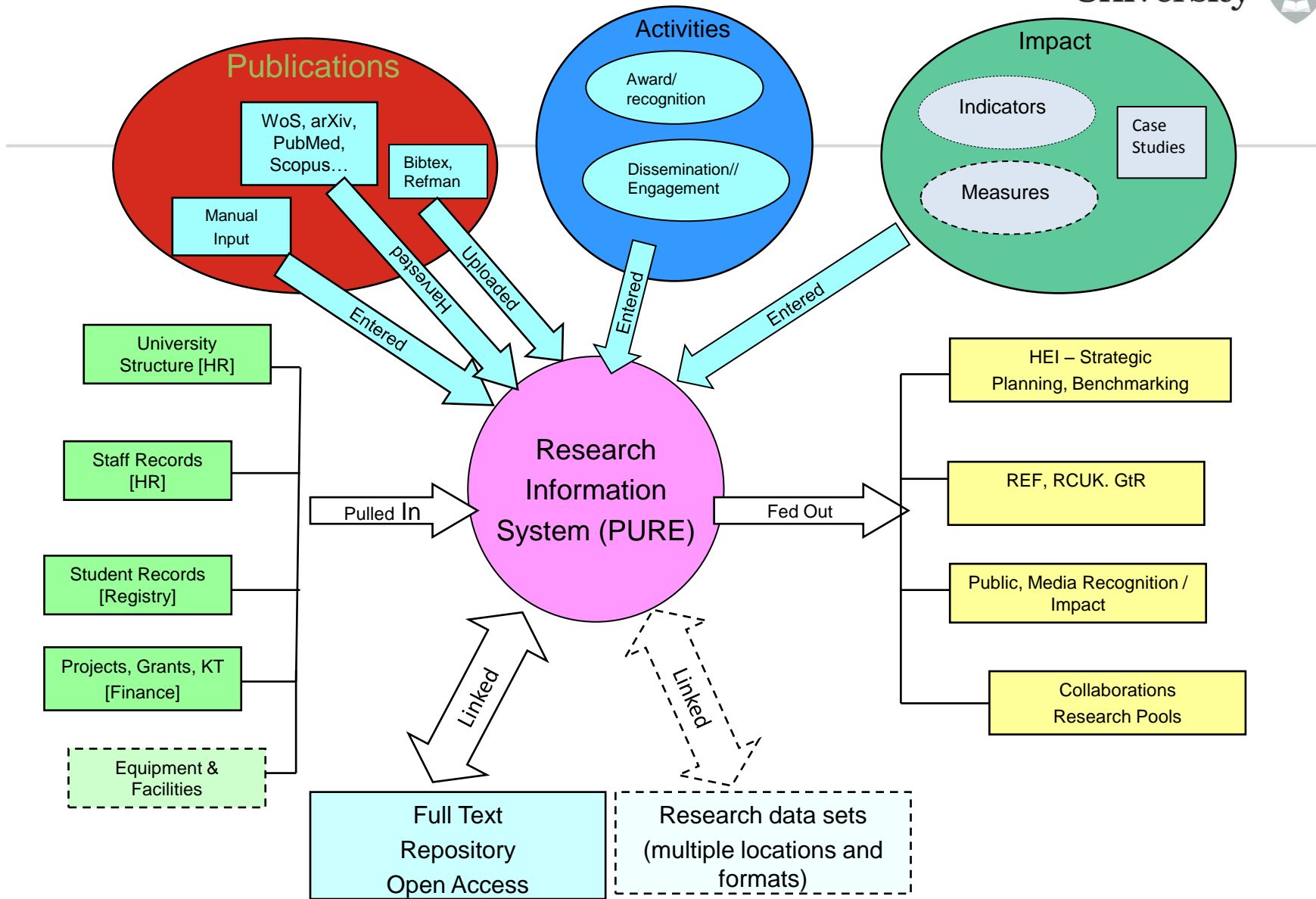
University of Southampton

[University of St Andrews](#)

[University of Strathclyde](#)

University of the Highlands & Islands

[University of York](#)



# How can Pure help with RDM?

---

- **Data catalogue**
  - New “Content Type” Dataset available in latest Pure release 4.20
  - Dataset metadata template developed by Pure User Group
  - Dataset metadata and files can be displayed on Pure Research Portal
- **Data repository**
  - Files can be uploaded into Pure and are stored according to a “mount point” (server address)
  - Embargoes and access restrictions can be added
  - Pure has API and connectors so that data can be synced with other systems

Bristol: Pure as data catalogue	Lancaster: Pure as data cat & repository
1. Researcher creates metadata record in Pure	1. Researcher creates metadata record in Pure. Researcher deposits data files in Pure
2. Metadata is automatically pulled into data.bris through the Rest API	2. Researcher receives automatic holding email that his data upload was successful.
3. Embryonic record created in with Title and putative DOI reference created	3. Data Librarian validates record and mints DOI by manually firing off request to DataCite.
4. Email is fired off to researcher to request that they login into to <a href="https://data.bris.ac.uk">data.bris</a> to link their data	4. Once DOI is minted, Data Librarian adds DOI to Pure record manually
5. Researcher links data and requests publication	5. Researcher gets email providing him with the DOI of his dataset in Pure
6. Data Librarian checks record and requests DOI from DataCite.	6. Data record goes live on <a href="#">Pure Portal</a> once the data is validated
7. Once DOI is minted, Data Librarian adds DOI to Pure record manually	

# Lancaster University Research

*Here you can find out more about Lancaster's world-class research activities, view details of publications, outputs and awards and make contact with our researchers.*



## Latest publications

### Review of Hong Kong documentary film by Ian Aitken and Michael Ingham

Bettinson, G. 1/06/2015 In : New Review of Film and Television Studies.

Contribution to journal > Book/Film/Article review

### An optimization framework for the development of efficient one-way car-sharing systems

Boyaci, B., Zografos, K. & Geroliminis, N. 1/02/2015 In : European Journal of Operational Research. 240, 3, p. 718-733 16 p.

Contribution to journal > Journal article

### Mapping the landscape of climate engineering

Oldham, P., Szerszynski, B., Stilgoe, J., Brown, C., Eacott, B. & Yuille, A. 02/2015 In : Philosophical Transactions of the Royal Society of London A. 372, 2031, p. 1-20 20 p.20140065

Contribution to journal > Journal article

## Latest news

### UK and India collaborate on future-proof crops

Lancaster University is working with researchers across the UK and India to provide improved crops for the future.

### Move over smart cities – the Internet of Things...

Lancaster University is about to take the concept of smart cities out of town. The Internet of Things - which enables object-t...

### Powering space craft of the future

Engineers at Lancaster University are working on powering future 'giant leaps' for mankind.

### Improved treatment for cancer patients

## Recently added publications

### Professor Garik Markarian Computing and Communications



### Incremental anomaly identification in flight data analysis by adapted one-class SVM method

Kolev, D., Suvorov, M., Morozov, E., Markarian, G. & Angelov, P. 2015 *Artificial neural networks*. Koprinkova-Hristova, P., Mladenov, V. & Kasabov, N. K. (eds.). Springer, p. 373-391 19 p. (Series in Bio-/Neuroinformatics; vol. 4)

**My research interests include:**  
Wireless Broadband Communications, Communication Systems for Aviation Security, Efficient Video Distribution, Wireless

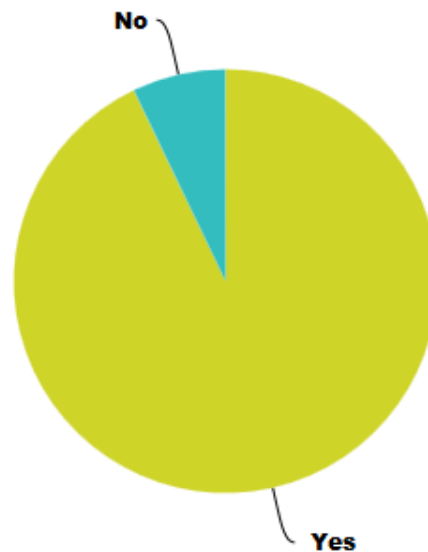
[Show more...](#)

# Pure User Group Survey

## January 2015

Are you or will you be using Pure as your data catalogue?

Answered: 14 Skipped: 0



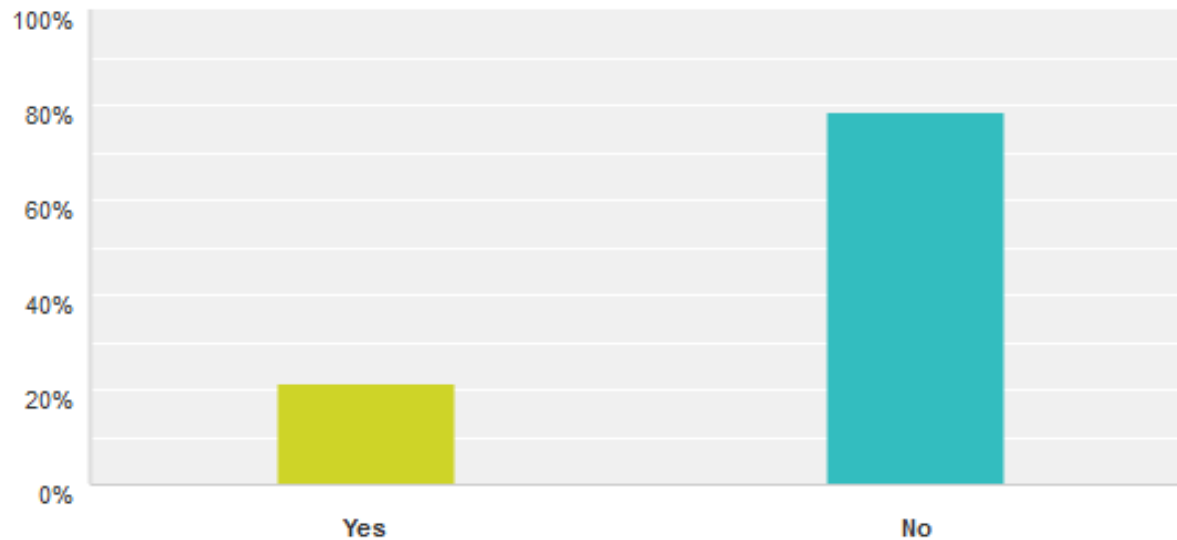
Answer Choices	Responses
Yes	92.86% 13
No	7.14% 1



# Pure as data catalogue

## Is your data catalogue live?

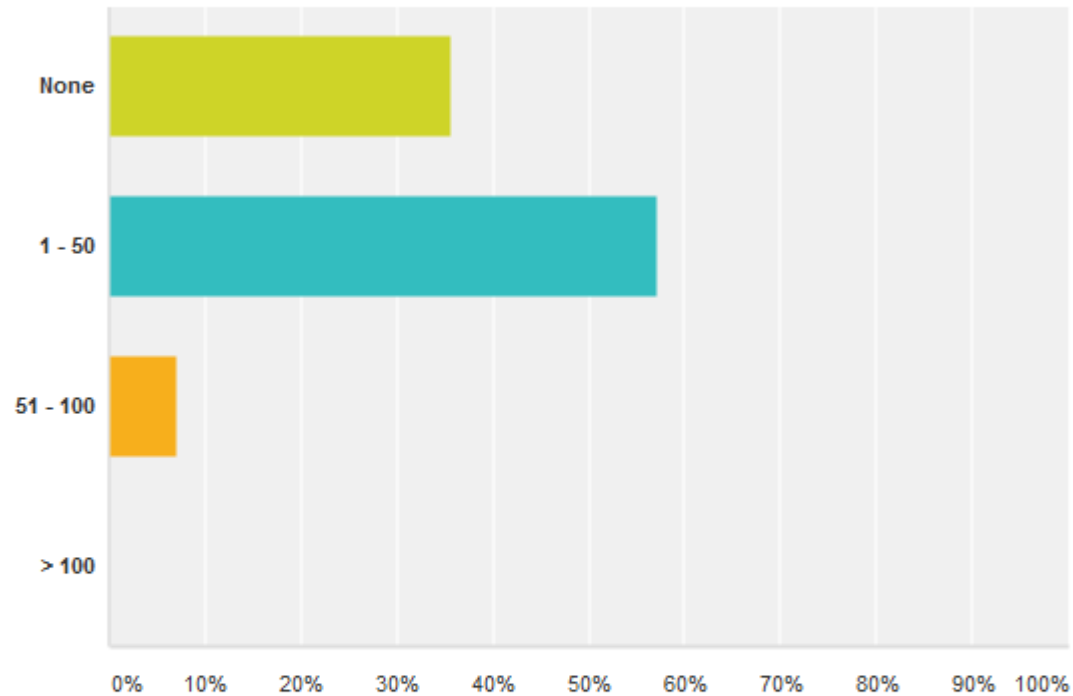
Answered: 14 Skipped: 0



Answer Choices	Responses
Yes	21.43% 3
No	78.57% 11
Total	14

## How many items do you have in your data catalogue?

Answered: 14 Skipped: 0

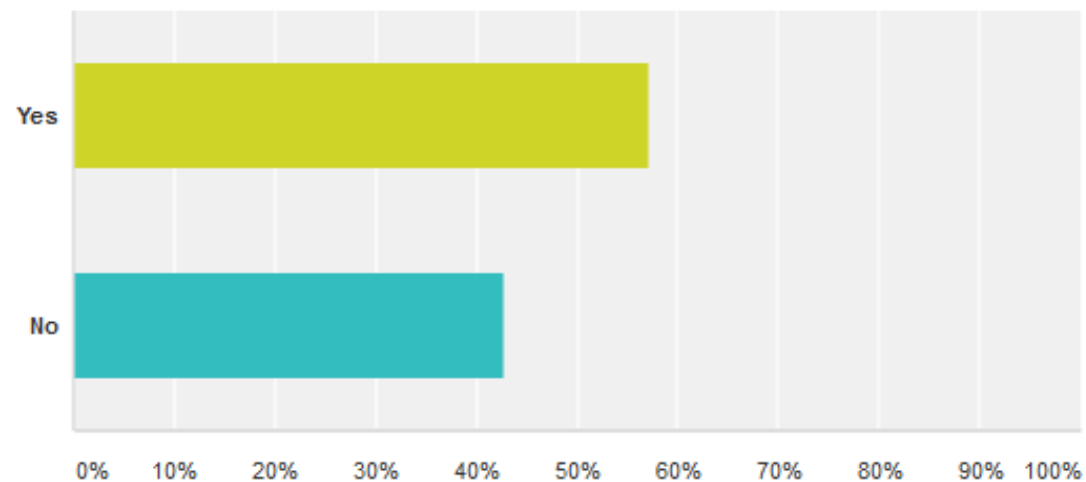


Answer Choices	Responses
None	35.71% 5
1 - 50	57.14% 8
51 - 100	7.14% 1
> 100	0.00% 0

# Pure as data repository

Are you or will you be using Pure as your data repository?

Answered: 14 Skipped: 0



Answer Choices	Responses
Yes	57.14% 8
No	42.86% 6
Total	14

# What do you think Pure does well with respect to RDM?

---

- “One stop shop for academics “
- “Familiar to researchers“
- “Relationships between datasets, outputs, projects etc”
- “Good metadata schema”
- “Supports compliance i.e. creates the link between the funding, the publications and the data”

# What are your biggest concerns about Pure and RDM?

---

- “Not open source, new functionality dependent on Elsevier priorities”
- “How do we get buy-in from our academics? They dislike Pure already”
- “User interface is muddled and inconsistent metadata fields”
- “How can Pure interact with other University systems?”
- “One size fits all metadata isn't suitable for all types of datasets”
- “Lack of curation functionality”

# How are you/ will you be handling digital preservation?

---

- “Not sure! Data will be held with Arkivum”
- “Possibly Archivemata supporting a manual workflow.”
- “We intend to copy out the files, re-label the originals within Pure, carry out file migrations as required and put the various version back into Pure”
- “Some level of preservation/services with ePrints, looking at links to long term dark archival storage options.”
- “We are using Arkivum 100 very long term preservation storage. We don't as yet have any curation policies.

# Lessons learned so far

---

- Pure can act as data catalogue and data repository
- Currently there are limitations in the interoperability of Pure with existing systems
- Further development is dependent on vendor. Pure User Group represents community interests
- Preservation is an issue that needs further investigation. Pure is not a preservation tool