Pure for research data: future developments and integrations

22th November 2016
RDMF16 - Edinburgh

Federica Fina, Data Scientist, University of St Andrews Library
ORCID: 0000-0003-2594-350X
Contents

• Pure at St Andrews
• Our datasets since June 2015
• The current workflow
• The good things about using Pure for datasets
• The things that are still missing for datasets
• The ideal workflow
  ○ The NOMAD-Pure project
• Integration with archival & preservation services
• Exposing content
Pure at St Andrews

- Introduced in **2010**
- Current Research Information System (CRIS)
- Research Outputs - catalogue (2010)
- Research Data - catalogue and ‘repository’ (2015)
Research at St Andrews

**Publications**
- WoS, arXiv, PubMed, Scopus...
- Bibtext, Refman
- Manual Input

**Activities**
- Award/recognition
- Dissemination/Engagement

**Impact**
- Case Studies
- Indicators
- Measures

**University Structure [HR]**

**Staff Records [Registry]**

**Projects, Grants, KT [Finance]**

**Full Text Repository**

**Research data sets** (local storage)

**Research data**
- Manual Input

**Research data pools**
- Pulling in
- Entered
- Uploaded

**Fed Out**
- CERIF

**Activities**
- Industry / SME’s Interface
- HEI – Strategic Planning, Benchmarking
- REF, RCUK, SFC, HESA

**Dissemination / Engagement**
- Award/recognition

**Impact**
- Case Studies
- Indicators
- Measures

**Collaborations Research Pools**

**Anna Clements**

**Research data Manual Input**
• The Portal

Research at St Andrews (https://risweb.st-andrews.ac.uk/portal/en/)
Data underpinning - Dynamic covalent assembly and disassembly of nanoparticle aggregates

Dataset

Associated persons
Stefan Borsley, (Creator)
Euan Robert Kay, (Creator)

Associated organisations
School of Chemistry
EaSTCHEM

Contact person
research-data@st-andrews.ac.uk

Description
Data for publication "Dynamic covalent assembly and disassembly of nanoparticle aggregates", DOI: 10.1039/c6cc00135a

The following file types are included:
- NMR spectroscopy files
- UV-Vis spectroscopy files
- Transmission electron microscopy images

Date made available 4 Apr 2016
Publisher University of St Andrews

Documents
Data_10.1039_c6cc00135a.zip
50 MB, multipart/x-zip

DOI
http://dx.doi.org/10.17630/463d9de9-a8c3-44b4-9eb2-bbc63f9364a6
Our datasets since June 2015

- 246 dataset records in Pure
- 153 dataset records with files and active St Andrews’ DOIs
- 32 datasets in progress
- 61 datasets deposited elsewhere but with metadata in Pure
The current workflow

Research activity

Active

Manual entry

Metadata

Pure

Files

Local storage
Private

Move

St Andrews Research Portal

Local storage
Public

User
The good things about using Pure for datasets

• Integrates with institutional systems
  o HR, students records, finance, ...

• Equipment module
  o Can provide additional information on how the datasets were generated (ScotCHEM project)

• Allows linking of content
  o Datasets – equipment – articles – projects - ... context and reporting on funder compliance

• Reporting
  o Scheduled reports

• Integrates with DataCite (!)
  o Before: 10-15 min to mint DOI - Now: click of a button

• Validation workflow

• Single interface
  o Researchers do not need to learn to use a new system
The things that are still missing for datasets

• **Metadata ingest**
  o From research tools such as ELNs, in-house developed systems
  o Other repositories
  o Data Centres

• **File ingest** (Integration with active research systems/storage)
  o From research tools such as ELNs, in-house developed systems
  o Storage systems (e.g. University cloud accounts)

• **Large files and sensitive files**

• **Integration with archival and preservation services**
  o Jisc Research Data Shared Services, Archivematica, Arkivum...

• **Exposing content for harvest**
  o Datasets model no OAI-PMH, content is not harvestable
  o Not OpenAIRE compliant (CRIS guidelines)
The ideal workflow

Research activity electronic tools

Active Storage solutions

Metadata

Large and sensitive files

Pure

St Andrews Research Portal

Any standard discovery service (e.g. Jisc UKRDDS, OpenAIRE)

Archival & Preservation services (e.g. Jisc RDMSS)

User

Any standard discovery service (e.g. Jisc UKRDDS, OpenAIRE)
Integration with active systems

• Example: GitHub - Zenodo

• What about Pure?
  o Hivebench – Mendeley Data – Pure

• Standard method of integration so to extend to other systems
  o ELNs
  o In-house systems
  o Cloud storage
  o …
At St Andrews: the NOMAD-Pure project

• NOMAD – NMR Online Management and Datastore
  o In-house developed system
  o Active development since 2012
  o Automates data collection
  o Active data management system
  o Heavily used in Chemistry (~500 experiments/day)

• NOMAD-Pure project
  o Started 2015
  o Aim: streamline the publication of NMR data for funders’ compliance
  o School of Chemistry, School of Computer Science, Library and IT
NOMAD – NMR Online Management And Datastore

Admin portal
- Centralised dashboard for traffic control & queue management (fair usage policy rules)
- Centralised database of users, research groups and experiments
- Usage and activity reports (accounting)
- E-mail notification about experiment status

Automation User Portal
- Collecting meta data
- Creating submission files for NMR instruments

Teaching User Portal
- Front end user portal customised for batch submission of NMR samples from teaching labs

Search portal
- Customisable instant search
- One click access to PDF prints and raw NMR data

Archive Data Storage
- MySQL
- ZFS

Manual user portal
- Upload of manually acquired NMR data

Dr Tomas Lebl - Senior Scientific Officer for Solution State NMR - tl12@st-andrews.ac.uk
• NOMAD – will act as a data repository for NMR data
• NOMAD public portal will allow for data visualisation and structure search
• Pure – data catalogue and validation workflow and DOI minting
• NOMAD will query Pure’s API to manage files visibility and publication
• Ideally, we would push/pull metadata from NOMAD to Pure (XML import?)
Integration with archival and preservation services

- Jisc RDM Shared Service Pilot

https://researchdata.jiscinvolve.org/wp/tag/rdm-shared-service/
Exposing content

• UK Research Data Discovery (Jisc) - https://www.jisc.ac.uk/rd/projects/uk-research-data-discovery

• Data harvested from 14 organisations
  o Uni. of St Andrews
  o ADS (Archaeological Data Service)
  o CCDC (Cambridge Crystallographic Data Centre)
  o NERC
  o UK Data Service
  o Uni. of Glasgow, etc....

• Problem: no OAI-PMH for Pure Datasets

• Workaround: for St Andrews the harvest is happening via DataCite
• DataSearch (beta), Elsevier - https://datasearch.elsevier.com/

• Harvest from data repositories such as Mendeley Data, Dryad...

• Will it be extended to institutional repositories (Pure portals?), Data centres etc.?
Research Data Management team

📞 01334 462343
✉️ research-data@st-andrews.ac.uk
🏠 http://researchdata.wp.st-andrews.ac.uk