How close are we to a complete view of research activity?

Karen Kay | University of Birmingham
Nov 1, 2018
The city of Birmingham

- 34 Million visitors a year
- One of the youngest cities in Europe (40% of people under 25)
- Home to 5 universities (73,000 students)
- 90% of UK is within 4 hours of Birmingham
- Birmingham’s Bullring is one of the largest shopping centres in Europe
- ‘The Hobbit’ and ‘The Lord of the Rings’ were inspired by places and people in Birmingham
- Birmingham is the birth place of heavy metal
The University of Birmingham

- Campus a few miles outside of the city centre home to approx. 29,000 students
- 1390 academic staff with research responsibility
- And........
......home to one of the actual two towers!
Who am I?

Deputy Head of Research Planning and Research Information Manager

- Project managed Pure implementation in 2012
- Chair of the Pure UK User Group
- Research Planning team responsible for impact development and the REF submission
- Work closely with Library and other research support functions at the institution to ensure research information collected and used responsibly and with purpose
A little about me

What does my name mean.com says:

Karen

Gender: Girl  Origin: Scandinavian  Meaning: Pure

So Pure and I were destined to be together!
The UK research landscape

• **Research Excellent Framework (REF)** is the system for assessing the quality of research in UK higher education institutions

  - Large data collection exercise
  - Staff, outputs, research income, PhD awards.
  - Quantitative evidence for the distribution of core research funding

• **Open access** for research publications has been mandated by research councils and for REF

  - Depositing of author accepted manuscript within 3 months of acceptance for publication (REF)
  - Funding available to pay article processing charges (funders)
Beginning with Pure

- Quick 6 month implementation project in 2012
- Rolled out to the Library and REF teams within weeks of technical implementation
- 6 years on it is embedded as a core system of the university.
- Majority of academic staff engage with Pure

Drivers for uptake include:
- Open access requirements
- REF preparations
- Updates to the website
- Reporting for promotions, PDRs etc
- Easy view of research funding
The amount of data in Pure is definitely going up!

- Outputs
- Impact
- Activities
- Media
- Linking content together
We are collecting so much information about the research that we do. But what are we actually doing with it, and does it even serve the needs that we have as an institution?
The research information landscape at UoB
What are we being asked to do?

What other countries do we publish with?

What does our activity in China look like?

Who is working in this area?

How do we look against our competitors?

What have we published in the past month?

How much do our academics collaborate internally?

Can I have an overview of their research activity over the past 2 years?
Easy!

- Quick, easy and self contained questions
- Whether Pure or another system we can get this quickly and easily
- Self serve most of the time

- What other countries do we publish with?
- What have we published in the past month?
- Who is working in this area?
How much do our academics collaborate internally?

What does our engagement with this partner look like?

Technically the data we need are in Pure
Widget available but not configurable
Can not drill up or down to understand data

CRMs hold data about engagements and activities
Pure holds grants and outputs and activities
External organisations and activities fields do not match
Nope… not a chance

What does our activity in China look like?

- Not totally sure they knew what they wanted!
- The way the data are held means can’t identify finding from China
- Have to source from several different systems

Can I have an overview of their research activity over the past 2 years?

- How to calculate awards/income?
- No system has a complete overview
- How can this be linked together into one overview
Barriers to answering these questions

- Data quality – manual input
- Lack of awareness of the range of systems we have
- Always an incomplete picture
- Aligning different data sets
- Time and resources required for longer term goals
- Data not collected for this purpose
- Partial integration via a data warehouse
"it's is my life"

-Jon Bovi
Don’t have the answer?

… just go buy another system!

“After analyzing all your data, I think we can safely say that none of it is useful.”
What are we doing?

- Data Management Group established

- Management of all data must be improved to be able to feed into the institution’s strategic goals

- New BI strategy to redevelop the data warehouse and link together disparate datasets

- Balancing the strategic demands and investment/priorities are an issue
Translating the strategy

- Data cleaning in Pure needs time and resource
- It is continuous, it is laborious
- Cleaning Scopus data is funded, because the feeling is that it feeds into global league tables data
- Challenge is how to articulate the benefit of cleaning Pure data
Look after your data and your data will look after you

- Pure does have some tools for managing master data
- We haven’t always given this the time needed and are now playing catch up

Is it possible to fully manage the data in Pure and does it do what we need to join together different data sets?
Thomas Gurney, Elsevier product manager of Datamodels, explains how their CCIC Data mantra, "Complete, Correct, Inter-connected, Current", is essential to the value and usability of Pure and a central component in the future development. #prcn2018

4:16 PM - 31 Oct 2018
3 out of 4 ain’t bad?
Who has a stake in collecting the data?

Activities

Press Office
Public Affairs
Research Support team
International Relations
Impact development and REF
Funding teams

Pure
CRM
Local databases
Personal CVs
Moving from many to one

- Dedicate some time to understanding what we need the data in Pure to look like

- How to link together different internal data sets

- APIs from Scopus and SciVal could help in the data warehouse
We can’t do it all and we can’t do it alone

• We can not solve this as one institution

• Information and metadata often are not just ours, they are shared

• The big questions are about competitors and collaborations
Are we storing information in the right way?

• Shared taxonomies – the ever elusive shared external orgs structure

• CRUD APIs – would these allow us to do more?

• Should we have internal and external persons?

• How get metadata providers to improve with us?

• Can initiative like ORCID really help?
Pure is not the one ring!

- Pure exists within a landscape of systems
- Local preferences for how to report certain data are always going to exist
- Pure is not a CRM, but also, CRMs are not CRISs!
The future – what do we need
Any questions….

Any quick solutions?