What is CASRAI?

● nonprofit global network of organizations
● mandate of reducing burden while improving information quality and access
● projects commissioned to develop common reusable agreements for interoperability
● decentralized network of networks
● institutions, funders, vendors, libraries
Why CASRAI?

- Lack of interoperability a root cause of admin burden and poor data quality
- Resources wasted; assessments blinded
- Need a sustainable agreement mechanism
- Avoid forcing agreement at the wrong layer
Common Interop Patterns

● A > B
  ○ information stored with one database needs to get into another database (so a local step can occur)

● A + B
  ○ information in multiple databases needs to be aggregated (so analysis can occur)
How we work

- Common open dictionary of data profiles and underlying terminology
- Policy leads, technology follows
- Decentralized, distributed projects use and iteratively evolve the dictionary
- Compliant systems signal other systems which data profiles they can support
Applied to Open Access Interop

● a lifecycle with multiple interop cases
  ○ article submission \((A > B)\)
  ○ APC reporting \((A > B)\)
  ○ OA Policy Monitoring \((A + B)\)
  ○ Others?

● a community with diverse needs
  ○ repositories
  ○ funders
  ○ research offices
CASRAI-UK Open Access Profile

- Glasgow (Chair)
- ESPRC (Chair)
- Jisc (Sponsor)
- MRC
- NERC
- Wellcome Trust
- HEFCE

- St. Andrews
- Kings
- Edinburgh
- Elsevier
- Eprints (new)
CASRAI-UK Open Access Profile

- Includes elements describing:
  - Output (ID; title; authors; version; etc)
  - Access (license type; freedoms; embargos)
  - Compliance (first/subsequent deposits; checks)
  - Underlying Research (materials statement)
  - Financing (funder; APC; EU component; etc)
CASRAI-UK Open Access Profile

- single profile supporting key UK cases:
  - REF Post-2014 (A > B)
  - APC Intake (A > B)
  - Horizon 2020 (A > B)
  - Internal Institutional Workflows (A > B / A + B)
  - Others??
Next Steps

- Finalize the policy view of the OA Data Profile
- Create the technology view
- Public review/comment
- Implementation pilots / phase 2 work
Thank-you

info@casrai.org
Researchers are wasting time and focus on a growing set of administrative burdens. Two consecutive questionnaires over six years in the US found teams spent up to 42% of their time on administration vs research. The findings point to ‘death by a thousand cuts’ of chronic duplication across multiple funders.
Problem 2

Research units at higher education institutions and the scholarly publishing units across the sector are increasingly frustrated by obstacles to smooth information flow and sharing in their common administration activities (reporting, partnering, evaluation).
Problem 3

Research funders are increasingly unable to grasp and understand the impacts of their investment as the evidence (information) they require is fragmented across multiple silos in multiple organizational units (CRIS, repositories, HR, Finance, etc.). This affects government as well as foundation funders.
Common Root Cause

Lack of information flow and reuse. If researchers could reuse their admin info they could reduce duplication. If organizations could combine and compare admin data regardless of the home silo they could gain better insight for evaluation and trend analysis.
Solution

Interoperability. Adopt new design goals and operational disciplines into our local and national infrastructure for digital administration. Include interoperability best practices in our systems requirements. Include interoperability in the design of our business processes.
Interoperability Option 1

Everyone maps to everyone else. Interoperability would be achieved if every HE research unit, library, repository, publisher, funder, industry partner agreed to write and maintain a technological map each time two or more systems needed to interoperate. This is currently the most common strategy.
Interoperability Option 2

Everyone use the same product or technology. Interoperability would be achieved if every HE research unit, library, repository, publisher, funder, industry partner agreed to adopt the same commercial product or a single underlying technology/database model.
Interoperability Option 3

Everyone adopts a common dictionary. Interoperability would be achieved if every HE research unit, library, repository, publisher, funder, industry partner agreed on common terminology and implements ‘purpose-built’ exchange profiles based on that common terminology. A hybrid of the other two options.
Business Requirements
Subject Matter Experts

Policy Domain

Tech Domain

Technology Implementations
Modelling, Software Experts

Project Interactions
Meetings, Documentation

Build & Maintain Agreements

Build & Maintain Mappings
Status

● what we have
  ○ large profiles for describing personnel & projects
  ○

● what we need
  ○ specific implementations with both sides of the bridge
  ○
Standards Consortia - membership

● large membership
  ○ larger network
  ○ lower costs

● tiered membership
  ○ flexible involvement

● representative membership
  ○ by discipline
  ○ by function
Which Interop Strategy?

- Build first, then invest in mapping
- Harmonize first, then
# Information Flow

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Library</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Closed vs Open Sharing

Not all of your sharing will be ‘open’. In many cases you will share information privately with one or more organizations with whom you are working.
Data Custodian Risk

How do we deal with the risks inherent in granting custodial role to unaccountable and non-transparent data hosts (Google, Twitter, etc.)
CASRAI Friction Points

● software-centricity
● silo worship
● knee-jerk decisions
Define Standards

● formal vs informal
● format vs tool
The Programmable Organization

If we all had an API?
Key Drivers for CASRAI

- **Reduce administrative burden**
  - researchers can’t reuse information
  - research offices - duplication
  - funders develop policies solo

- **Improve administrative data**
  - data about: people, orgs, equipment, projects, funding
  - strengthen data practices
  - enable cost-effective sharing
Sharing Scenarios

private

performers

funders

public