Engaging Content
Engaging People

Engaging Information Professionals in the process of Authoritative Linked Data Interlinking

2018-11-28 at SWIB 2018

Lucy McKenna, Christophe Debruyne & Declan O'Sullivan

ADAPT Centre, Trinity College Dublin, Ireland

Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union
European Union
European Regional Development Fund
SFI Science Foundation Ireland

The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Library LD projects

- Increase in uptake & number of libraries implementing LD
- Mostly large institutions/organisations
  - Access to financial & technical resources
- Few implementations use multiple datasets
  - Often single institution initiatives
  - Limited interlinking across datasets
  - Mostly linked to large authorities/controlled vocabularies

Deliot (2014), Wang & Yang (2018), Vander Sande et al. (2018)

LD Survey for Information Professionals

Aims:

1. Explore Information Professionals' (IPs) knowledge & use of LD
2. Explore the challenges that IPs experience with LD
3. Explore how to overcome these challenges

- Online questionnaire - 50 Questions

- 185 participants
  - Primary Information Professionals from library domain
  - Majority had prior knowledge of the SW (84%) & LD (90%)

McKenna et al. (2018)

Key Findings - Experience with LD

Benefits

Improved data discoverability & accessibility

Cross institutional linking & integration – additional context for data interpretation

Enriched metadata & improved authority control

Challenges

Resource Issues: Dataset/provenance availability & quality, lack of guidelines & use-cases, funding & training, URIs

LD Tooling: Usability issues, unsuitable for needs of LAMs, immature software, technological complexity & learnability

Interlinking & Integration: Ontology & link-type selection, data reconciliation, vocabulary mapping

Key Findings - Potential Solution

- 89% rated LD Tooling specifically designed for IPs as **useful**
  - Reduce technical knowledge gap
  - Encourage increase of LD use in LAMs

- **Requirements**
  - Attuned and adaptable to LAM workflows
  - Hide LD technicalities
  - Aware of common LAM data sources & data quality

- **Importance Measure of Data Quality Criteria**
- Trustworthiness (66%), Interoperability (51%), Licensing (49%), Completeness, (41%), Understandability (40%), Provenance (39%), Timeliness (38%)

Research Focus

1. **Interlinking**
   - Limited interlinking across datasets & institutions
   - Area of particular difficulty in survey & literature
   - Limited guidelines on interlinking library resources
2. **Provenance**
   - Limited guidelines on LD provenance for LAMs
   - Adds to the authority & trustworthiness of LD
3. **LD tooling**
   - Usability issues – mostly designed for technical/LD experts
   - Often not suitable for library workflows or requirements
4. **Library Domain**
   - Majority of survey participants, Data access

Research Question

How can information professionals be facilitated to engage with the process of authoritative linked data interlinking with greater efficacy, ease, and efficiency?

What is Authoritative Interlinking?

- Interlinking – creating a link between two LD resources
- Authoritative - known to be reliable & trustworthy
  - LAMs are an authoritative source of information
  - Provision of provenance data
  - Quality of resources being interlinked

Why Information Professionals?

- Experts in metadata creation, knowledge discovery & authority control

Current Interlinking Frameworks

- RDF Refine, SILK, LIMES, MARiMbA, Catalogue Bridge
  - Majority require a technical knowledge of LD
  - Primarily support owl:sameAs links
  - RDF Refine & MARiMbA
    - Aimed at library domain
    - Access to large-scale datasets e.g. VIAF, LCSH

- **Further Requirements**
  - Additional link types e.g. dct:relation, schema:isPartOf
  - Interlink with datasets emerging from smaller authoritative institutions
  - Remove need for expert technical/LD knowledge

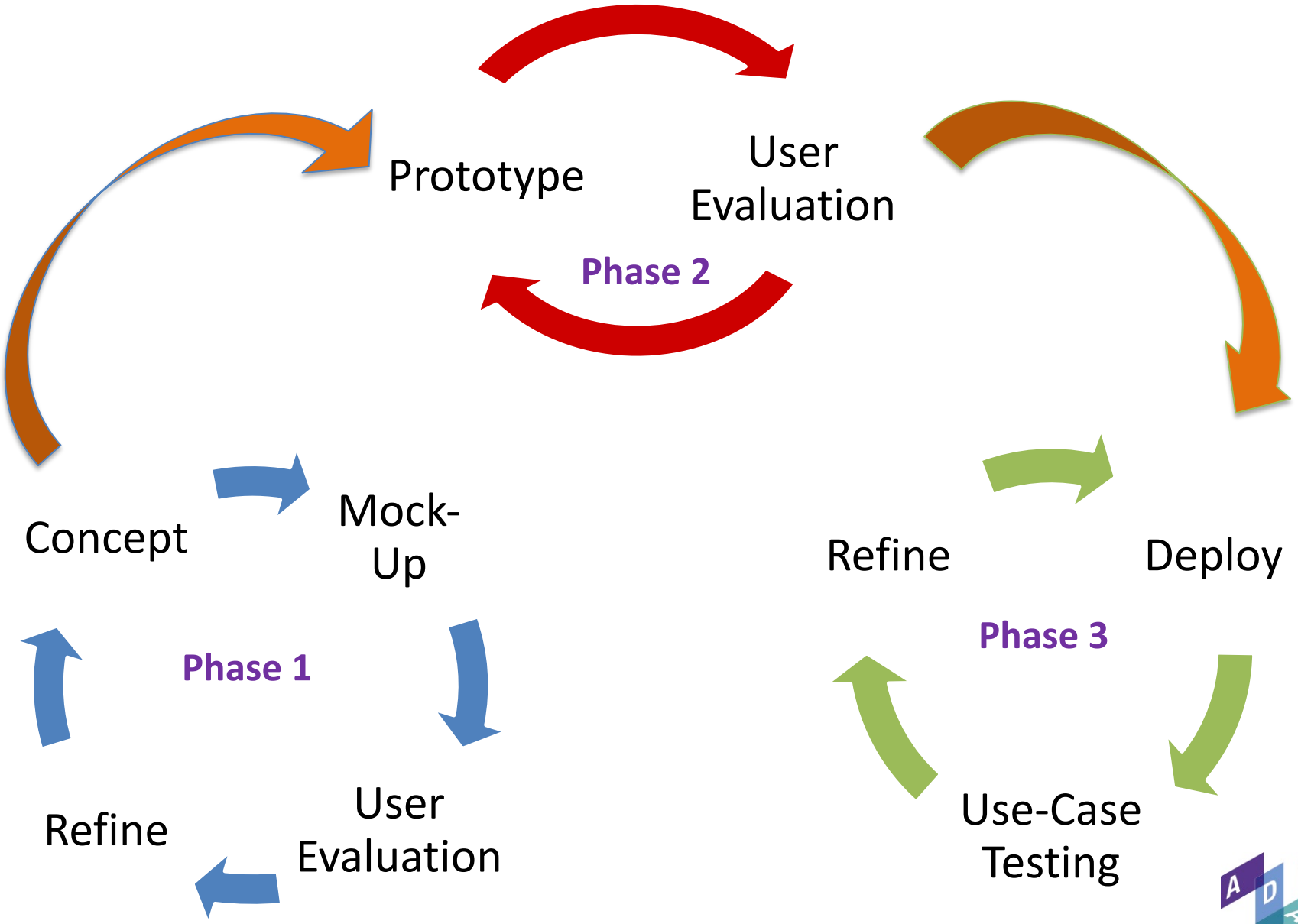
Research Aims

Develop an **authoritative interlinking framework** specifically designed with the workflows and expertise of IPs in the library domain in mind.

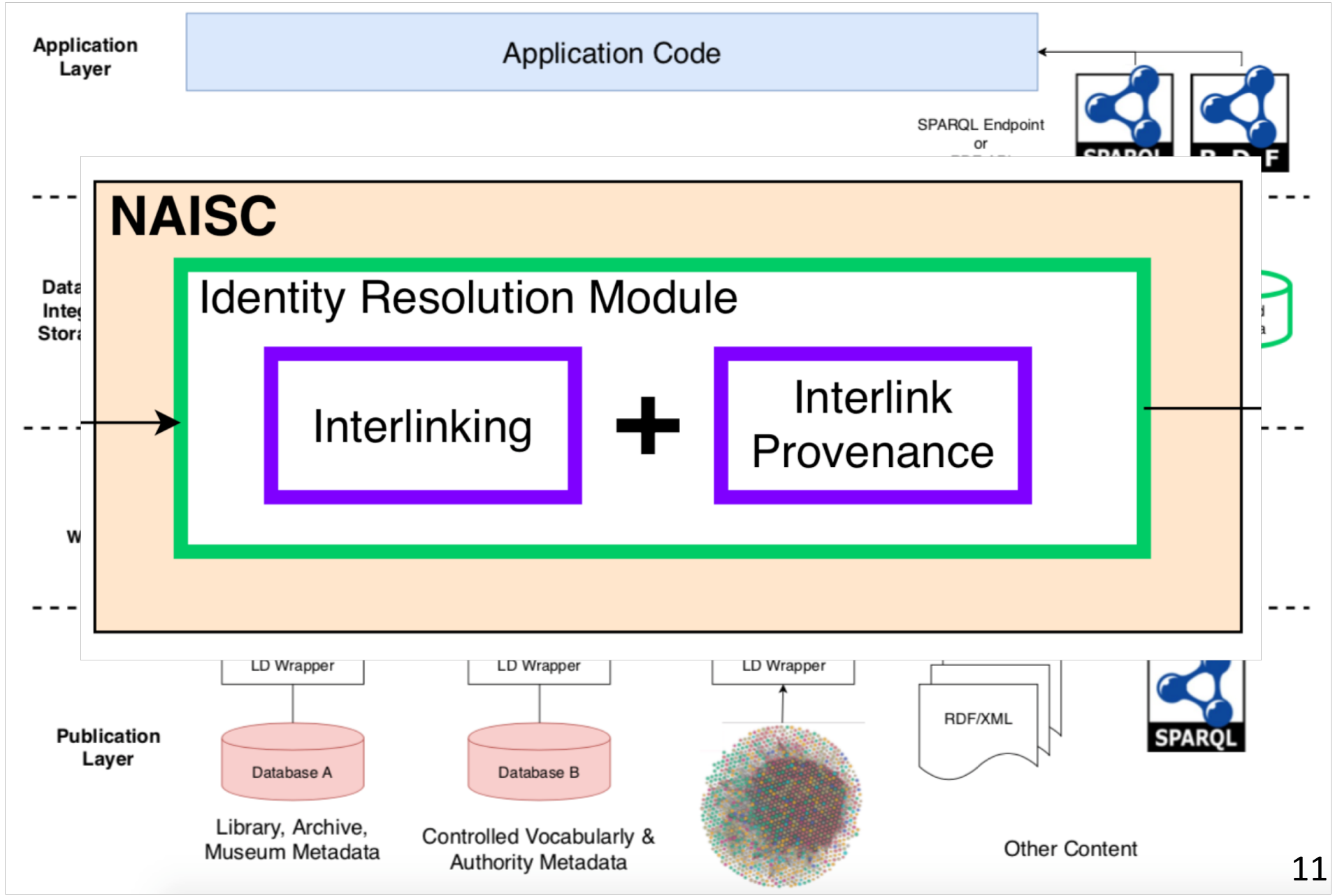
Develop a **provenance model** that expresses the required provenance of interlinks created by IPs.

Design an **interlinking interface** for IPs that guides users through the interlinking process including ontology and link type selection, and provenance data generation.

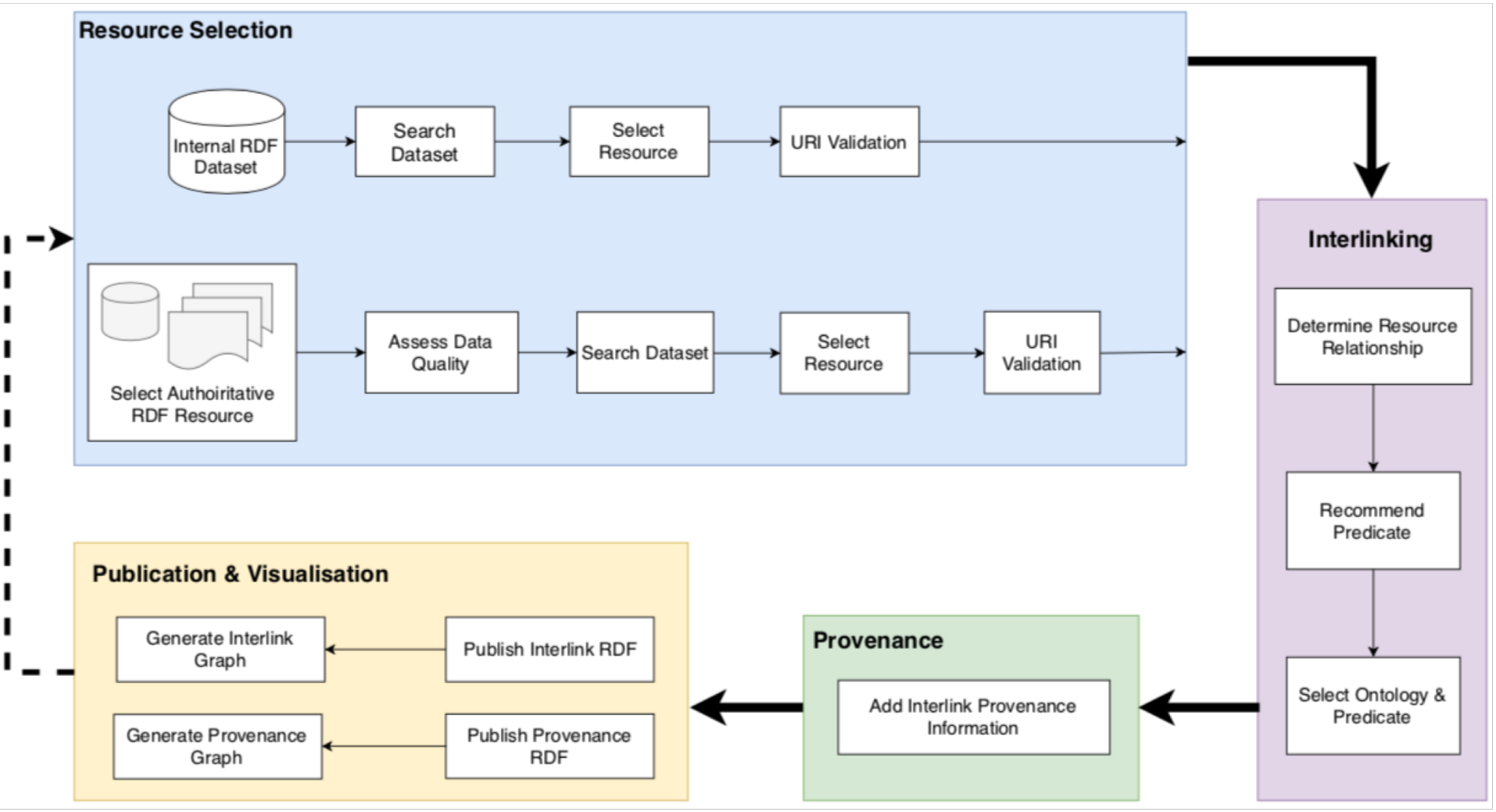
Methodology



NAISC – Novel Authoritative Interlinking of Schema & Concepts

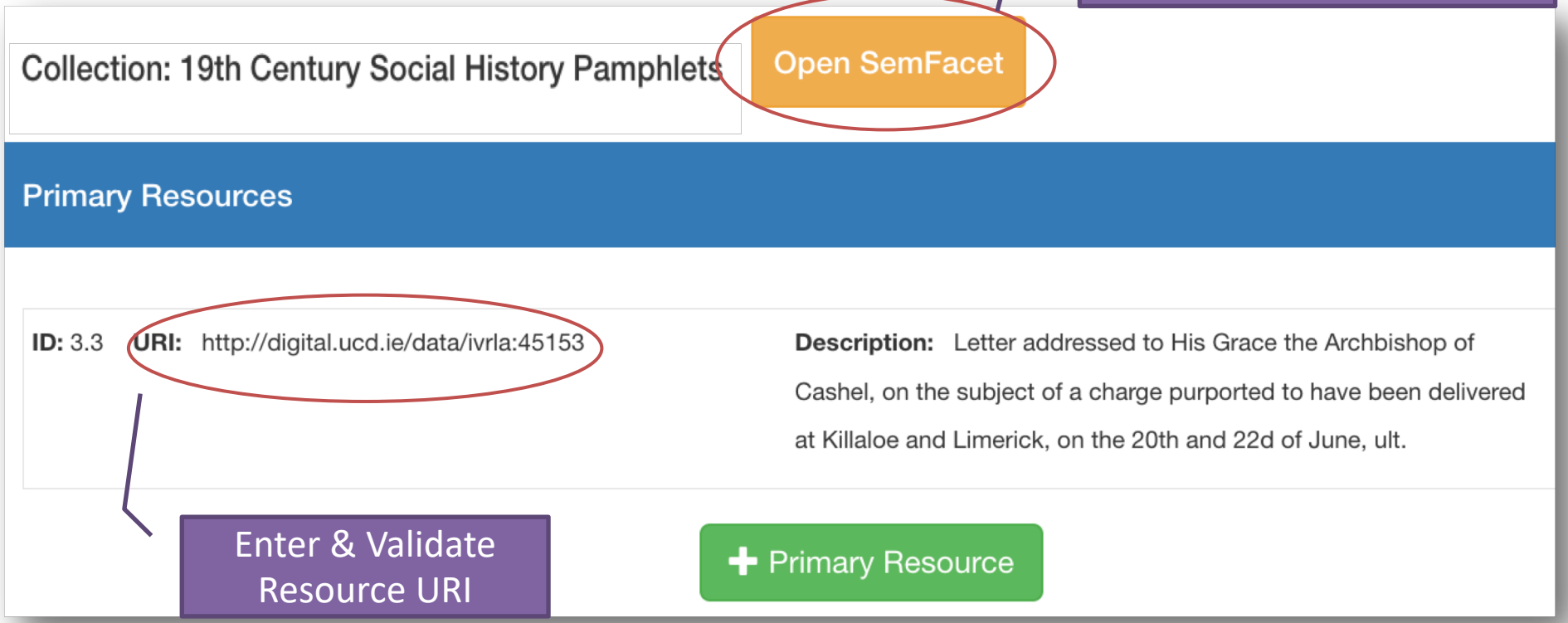


NAISC Framework

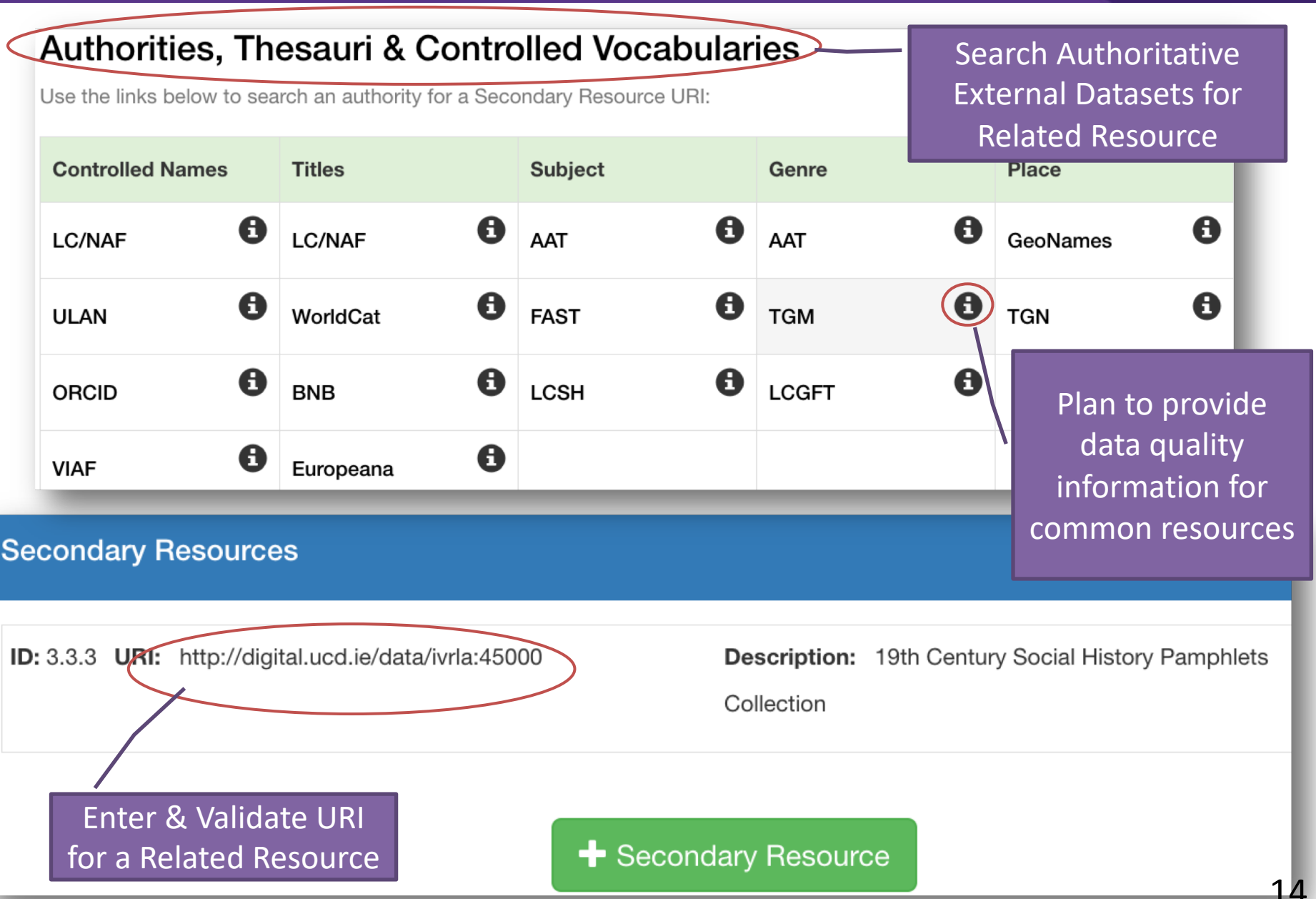


NAISC – Resource Selection

Search Internal RDF Dataset using Semantic Faceted Search Tool, SPARQL Endpoint or Web Resource



NAISC – Resource Selection

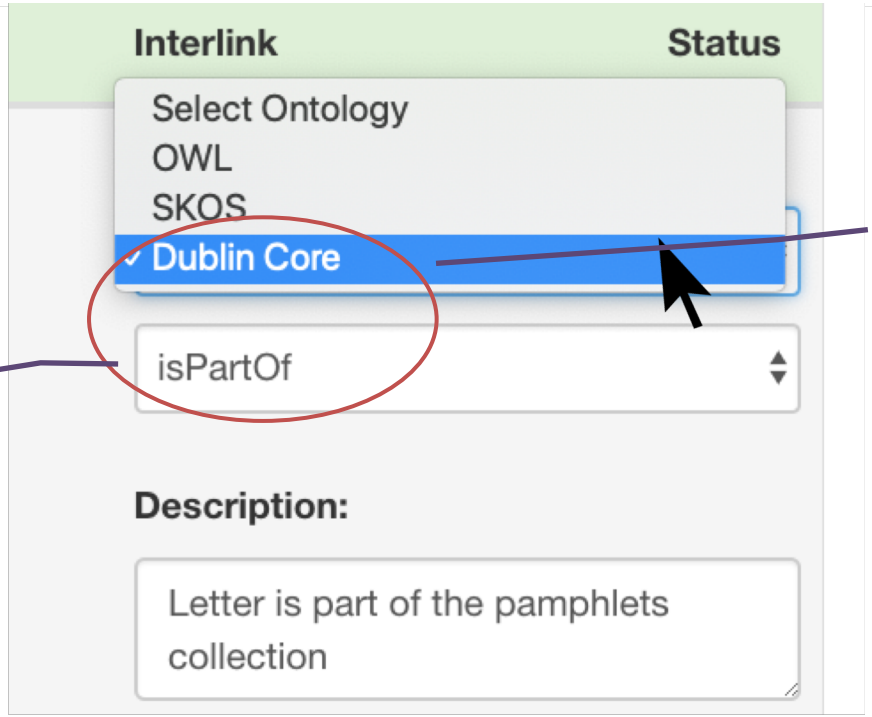


NAISC - Interlinking

Interlink Resurces

Guide

ID	Primary Resource	Secondary Resource
3.3	**Primary URI:** http://digital.ucd.ie/data/ivrla:45153 **Description:** Letter addressed to His Grace the Archbishop of Cashel, on the subject of a charge purported to have been delivered at Killaloe and Limerick, on the 20th and 22d of June, ult.	**Secondary URI:** http://digital.ucd.ie/data/ivrla:45000 **Description:** 19th Century Social History Pamphlets Collection



Plan to develop a Predicate Recommender that would suggest suitable predicates based on resource & relationship description

Select Predicate that describes the relationship between the resources

NAISC – Provenance Competency Qs

Provenance of Interlinks	
Who created the interlink?	What dataset does the interlink point to?
How was the interlink created?	Where can the dataset be accessed?
Why was the interlink created?	What resources are interlinked?
Where was the interlink created?	What is the relationship between the resources?
When was the interlink created?	Why was this predicate selected?
What dataset is the interlink part of?	When was the interlink last modified?
Who published the dataset?	Who modified the interlink?
Where can the dataset be accessed?	Why was it modified?
Provenance of Provenance	
When was the provenance data generated?	Who generated the provenance data?

NAISC – Provenance Model

Used 3 graphs:

1. Interlink Graph: A Named Graph for a set of interlinks

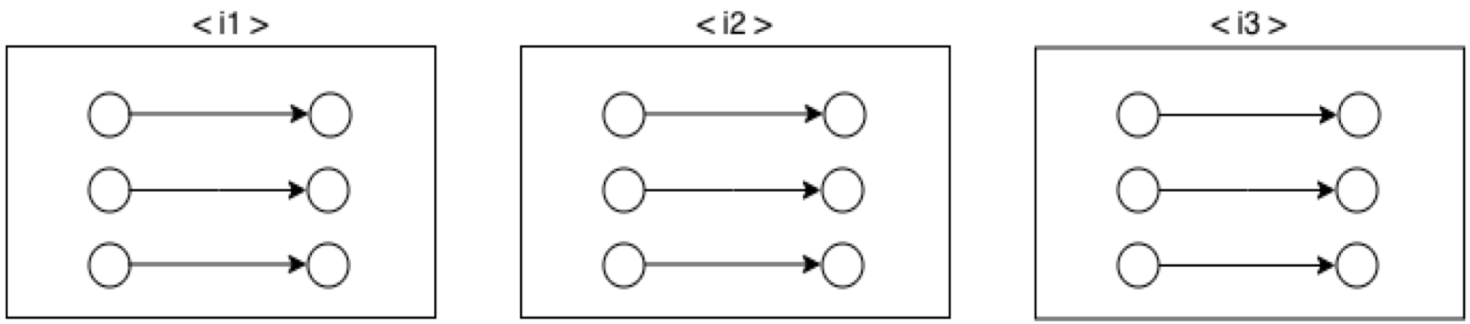
2. Provenance Graph: A prov:Bundle containing a set of provenance descriptions for a set of interlinks

3. Relationship Graph: A graph that represents the relationship between an Interlink Graph and a Provenance Graph.

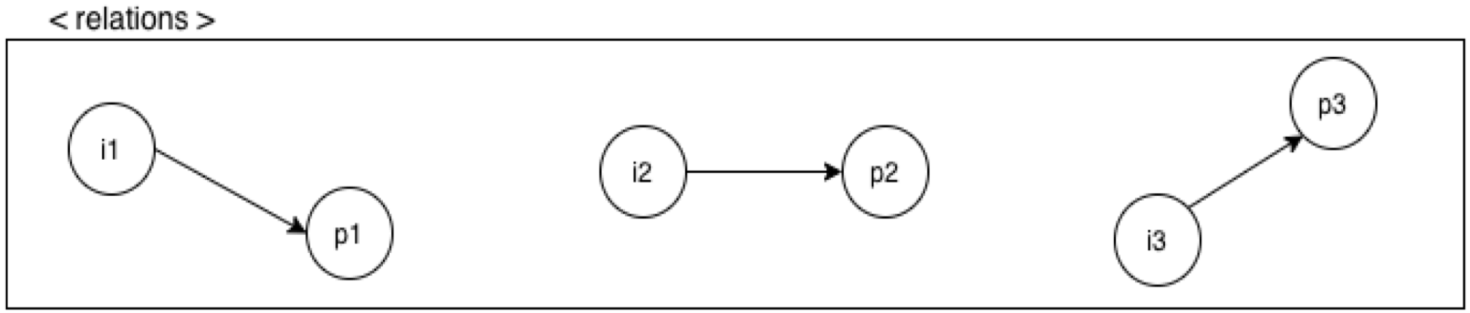
- As a Prov Bundle is an entity we can describe the provenance of the interlink provenance data contained in the bundle.

NAISC Provenance Model

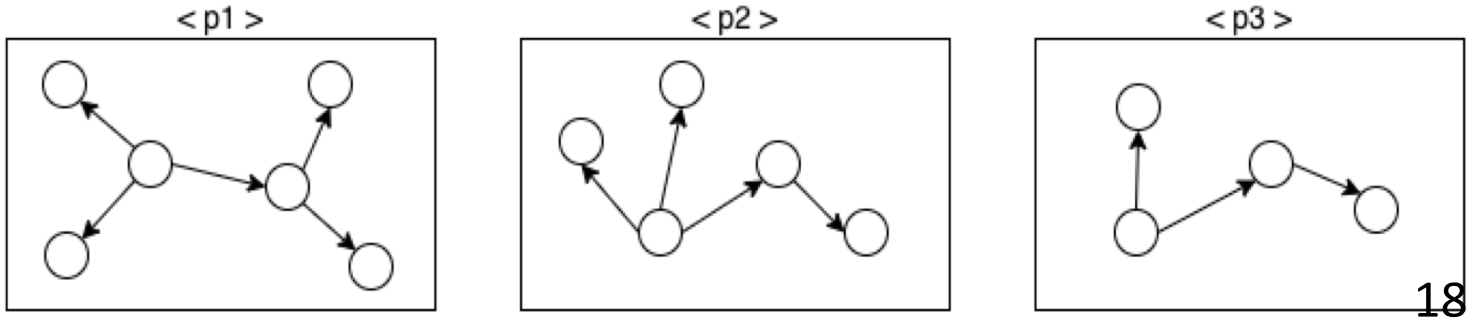
Interlink Graphs



Relationship Graph

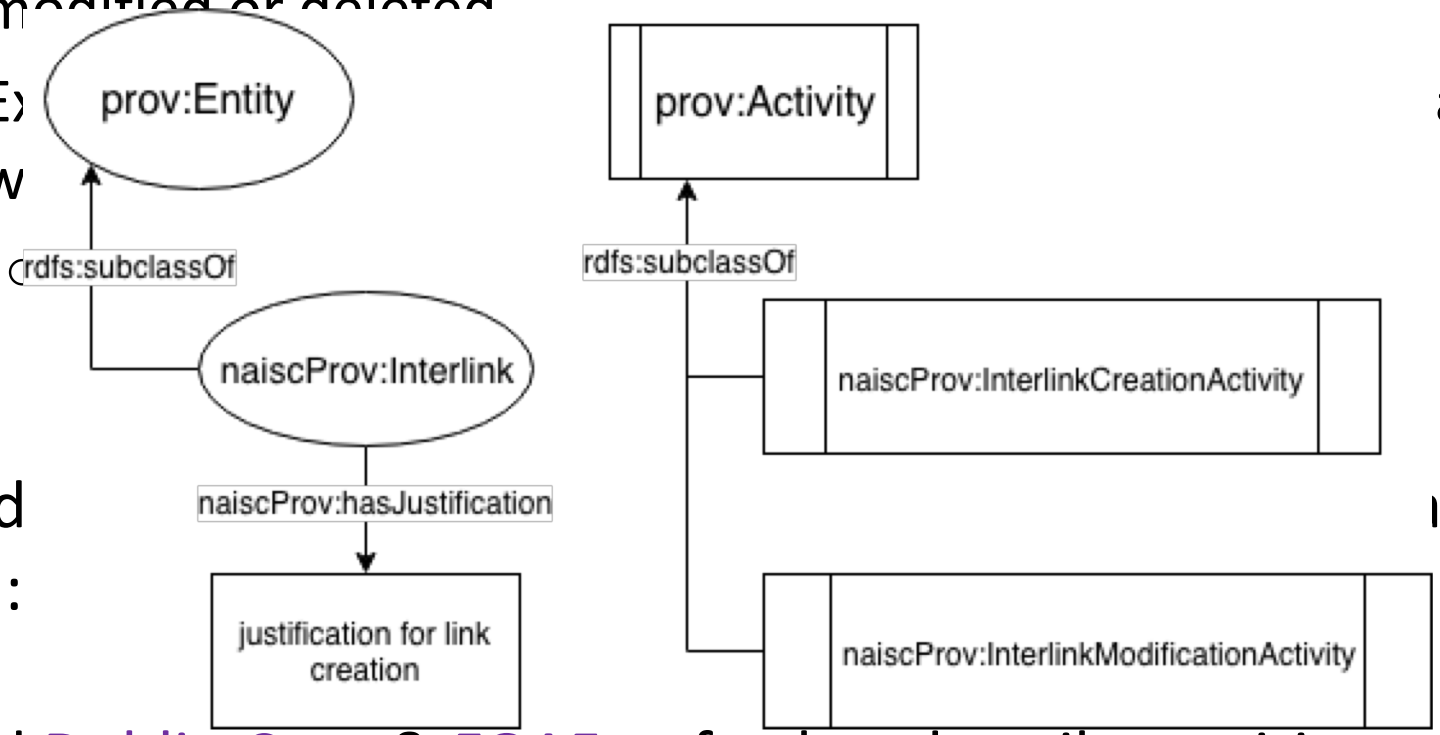


Provenance Graphs



NAISC – Provenance Ontologies

- Used the Prov Ontology
  - Describe who, where, and when interlinks were created, modified or deleted
  - Ex and w



- Used taset, void:
- Used Dublin Core & FOAF to further describe entities e.g. dct:title, dct:description, foaf:name, foaf:givenName

NAISC – Publication & Visualisation

Graph Interlinks

RDF Output

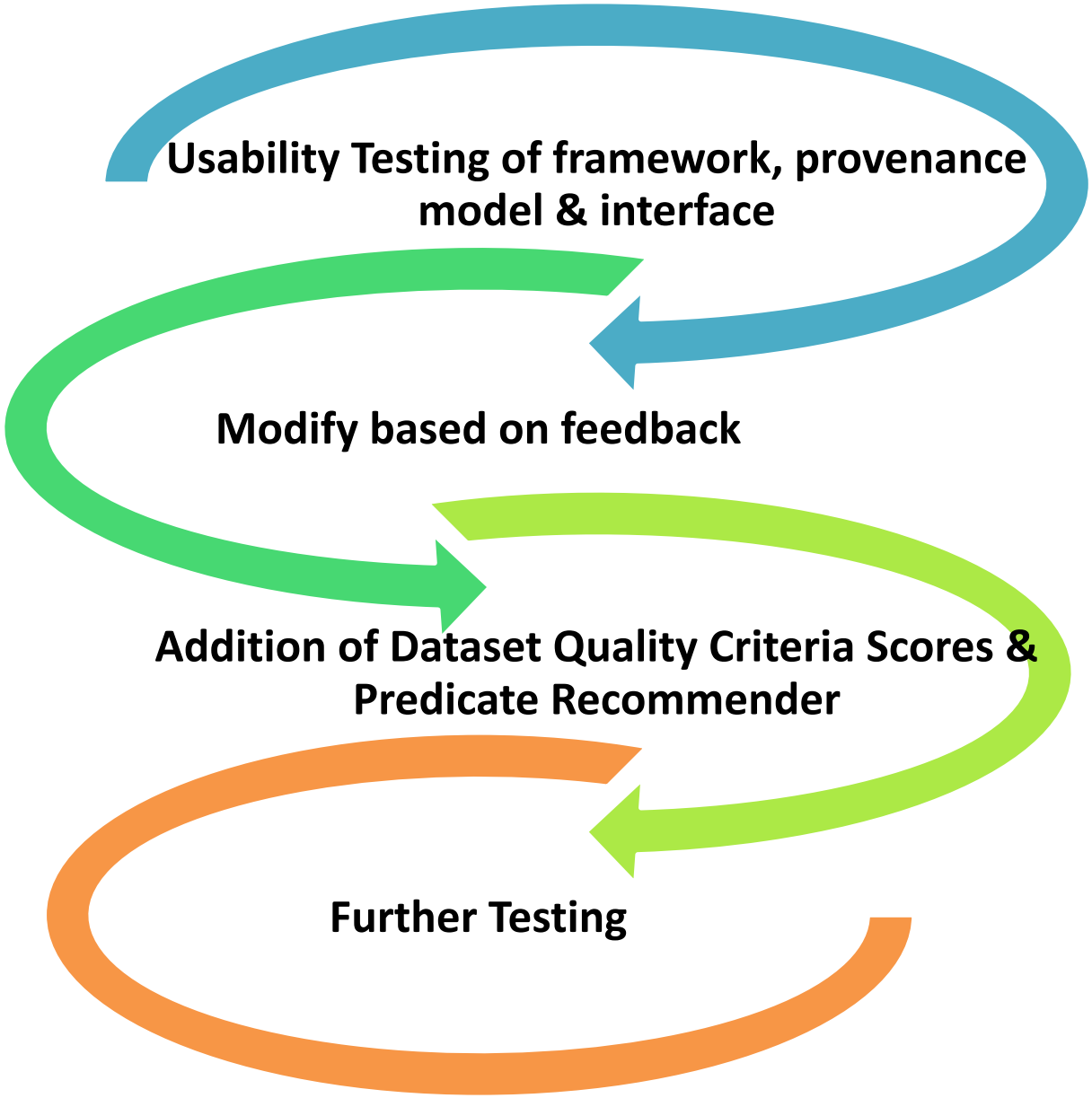
<http://digi
<htt

htt

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

  <#TriplesMap1>
  rr:logicalTable [
    rr:sqlQuery """select p.primaryurl, s.secondaryurl, concat(pr.ontologyURI,
pr.predicate)  predicate from dataset d
    join primaryresource p on d.id = p.dataset_id
    join link l on l.primaryresource_id = p.id
    join secondaryresource s on s.link_id = l.id
    join predicate pr on pr.link_id = l.id
    where d.id = {DATASET_ID} and pr.linkstatus != 'deleted'
    """;
  ];

   rr:subjectMap [
      rr:column "primaryurl";
   ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:column "predicate";
      rr:termType rr:IRI;
    ];

     rr:objectMap [
      rr:column "secondaryurl";
      rr:termType rr:IRI;
     ];
  ];
  .
```

/ivrla:45000

Future Directions



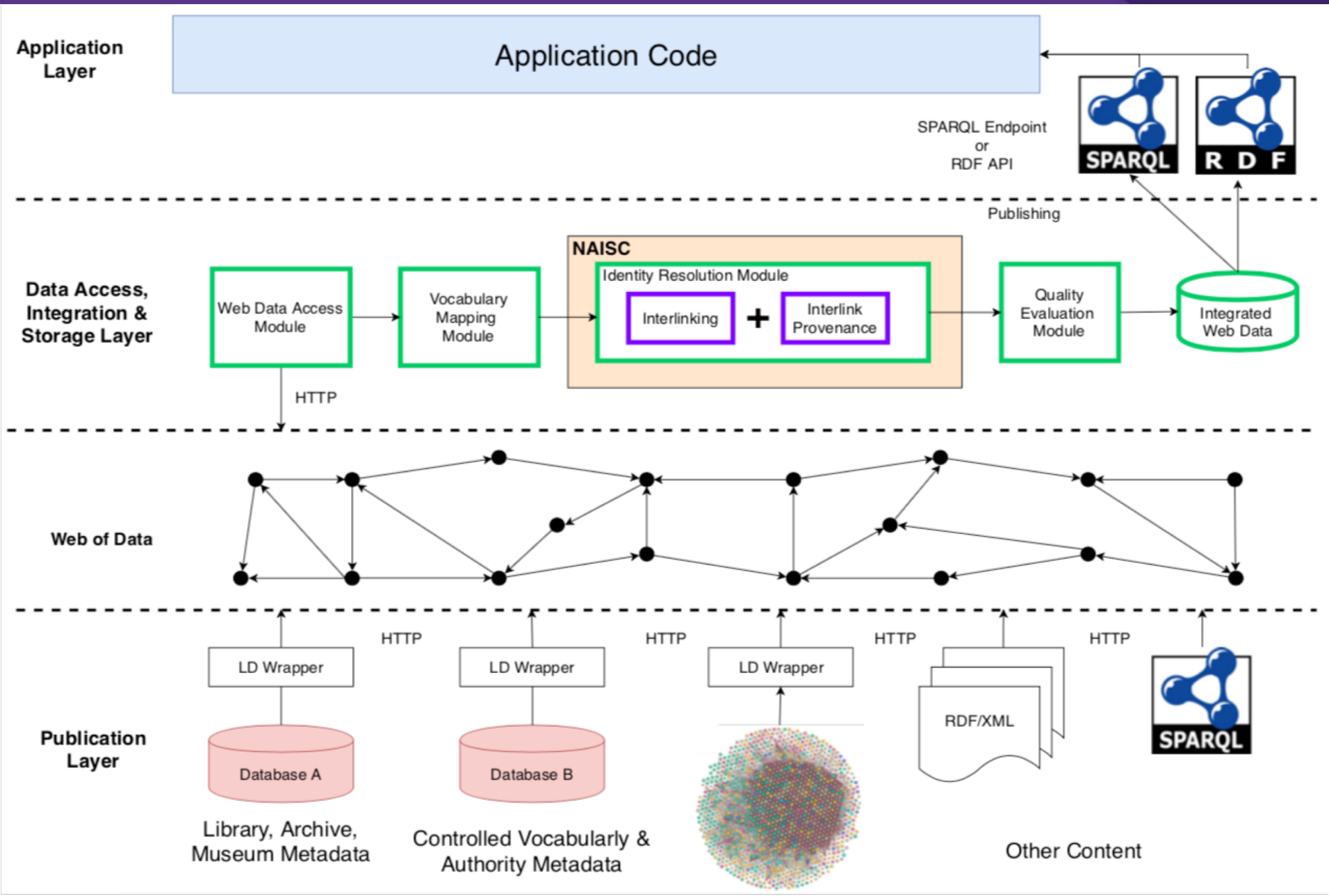
Thank you!

Any Questions?

References

- Ali, I., & Warraich, N. F. (2018). Linked data initiatives in libraries and information centres: a systematic review. 36(5), 925-937. doi:10.1108/EL-04-2018-0075
- Deliot, C., Wilson, N., Costabello, L., & Vandenbussche, P. Y. (2016). The British National Bibliography: Who uses our Linked Data?
- Hastings, R. (2015). Linked Data in Libraries: Status and Future Direction. Computers in Libraries, 35(9), 12-16.
- LaPolla, F. (2013). Perceptions of Librarians Regarding Semantic Web and Linked Data Technologies. Journal of Library Metadata 13, (2-3), 114–140.
- McKenna, L., Debruyne, C., & O'Sullivan, D. (2018, May). Understanding the Position of Information Professionals with regards to Linked Data: A Survey of Libraries, Archives and Museums. In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (pp. 7-16). ACM.
- Smith-Yoshimura, K. (2016). Analysis of an International Linked Data Survey for Implementers. D-Lib Magazine 22, 7/8.
- Smith-Yoshimura, K. S. (2018). Analysis of 2018 international linked data survey for implementers. code{4}lib, (42).
- Vander Sande, M., Verborgh, R., Hochstenbach, P., & Van de Sompel, H. (2018). Toward sustainable publishing and querying of distributed Linked Data archives. Journal of Documentation, 74(1), 195-222.
- doi:doi:10.1108/JD-03-2017-0040
- Wang, Y., & Yang, S. Q. (2018). Linked Data Technologies and What Libraries Have Accomplished So Far. International Journal of Librarianship, 3(1). doi:10.23974/ijol.2018.vol3.1.62
- W3C (2013). PROVO-O: The PROV Ontology. Retrieved 19/11/18 from https://www.w3.org/TR/prov-o/

NAISC - Linked Data Application Framework



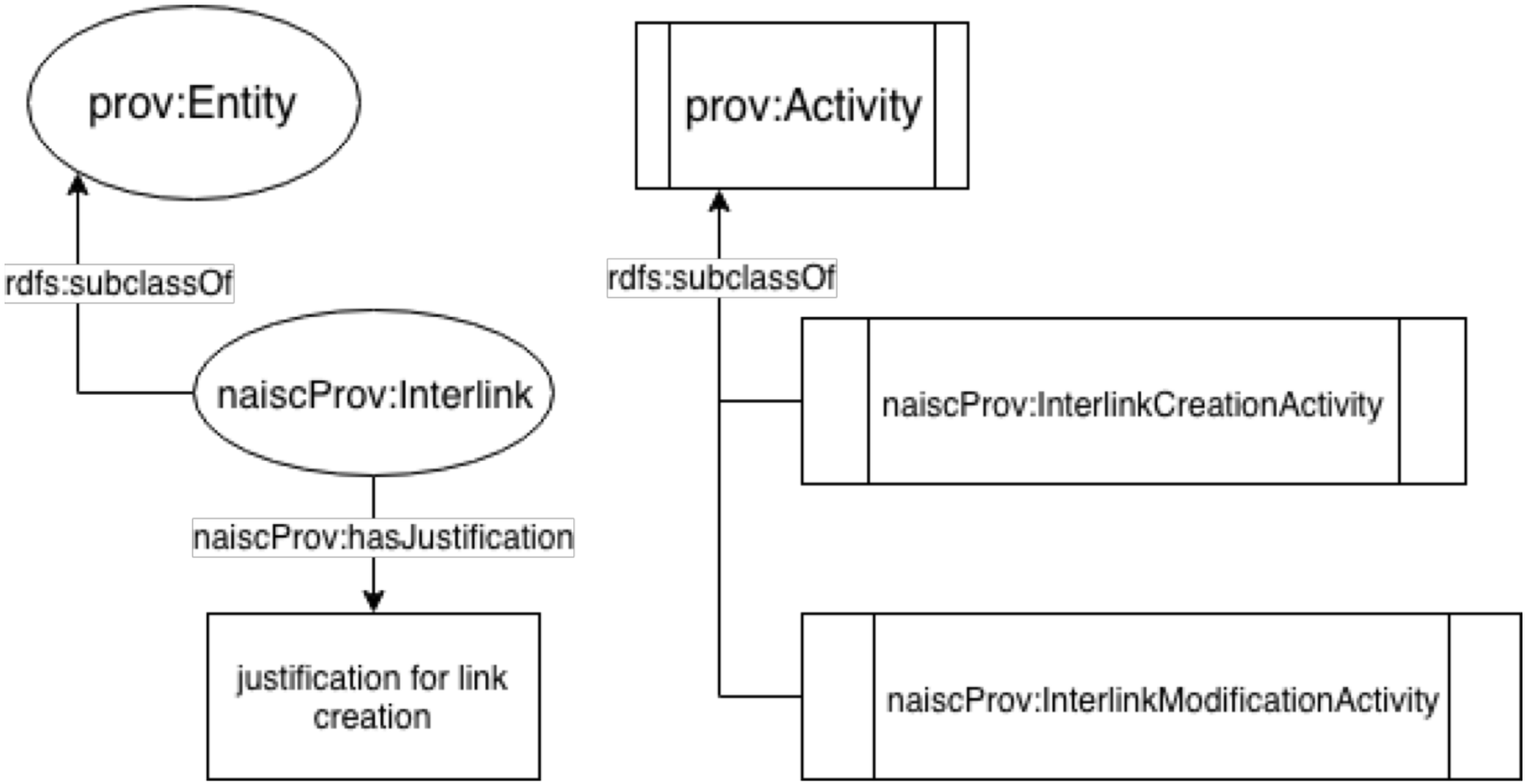
NAISC – Provenance Ontologies

- Used the Prov Ontology
  - Describe who, where, and when interlinks were created, modified or deleted
  - Extended ontology - NaiscProv – to describe what, how, and why interlinks created
    - Added interlink specific sub-classes & properties e.g. naiscProv:Interlink, nasicProv:hasJustification

- Used Void Ontology for dataset description e.g. void:Dataset, void:sparqlEndpoint, void:dataDump

- Used Dublin Core & FOAF to further describe entities e.g. dct:title, dct:description, foaf:name, foaf:givenName

NAISC – Provenance Ontologies



NAISC – R2RML Mapping

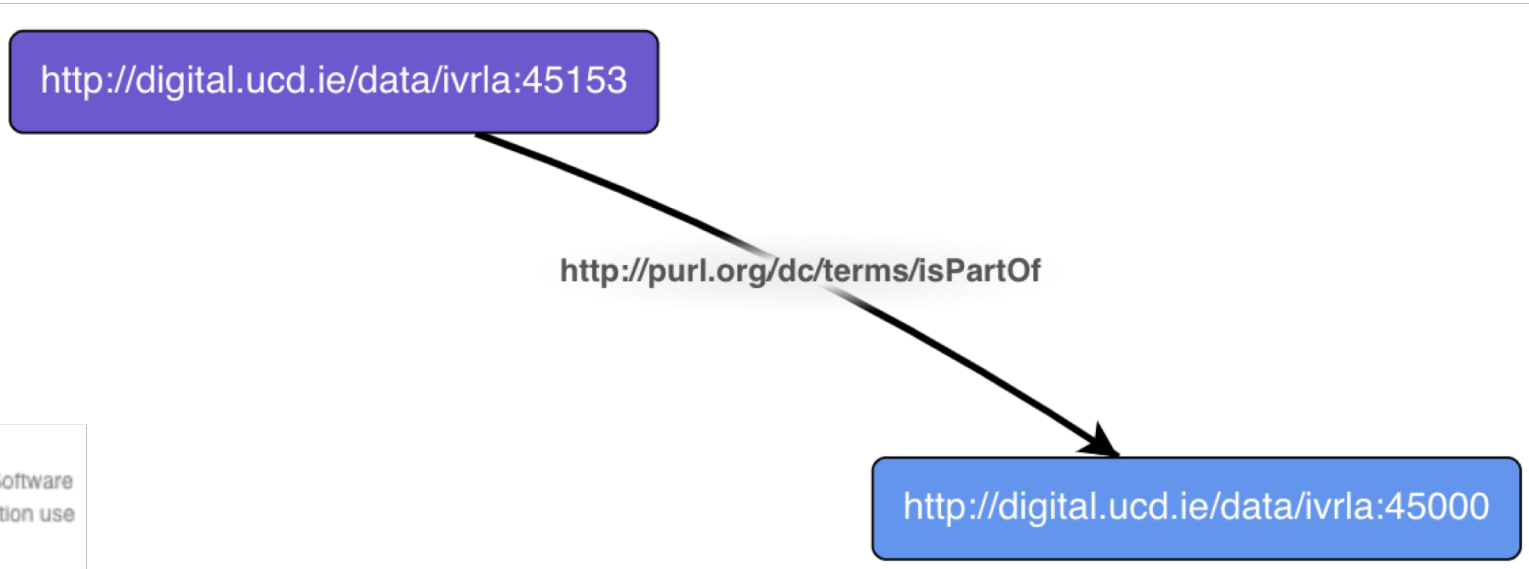

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

  <#TriplesMap1>
  rr:logicalTable [
    rr:sqlQuery """select p.primaryurl, s.secondaryurl, concat(pr.ontologyURI,
pr.predicate)  predicate from dataset d
    join primaryresource p on d.id = p.dataset_id
    join link l on l.primaryresource_id = p.id
    join secondaryresource s on s.link_id = l.id
    join predicate pr on pr.link_id = l.id
    where d.id = {DATASET_ID} and pr.linkstatus != 'deleted'
    """;
  ];

   rr:subjectMap [
      rr:column "primaryurl";
   ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:column "predicate";
      rr:termType rr:IRI;
    ];

     rr:objectMap [
      rr:column "secondaryurl";
      rr:termType rr:IRI;
     ];
  ];
  .
```

NAISC – Publication & Visualisation

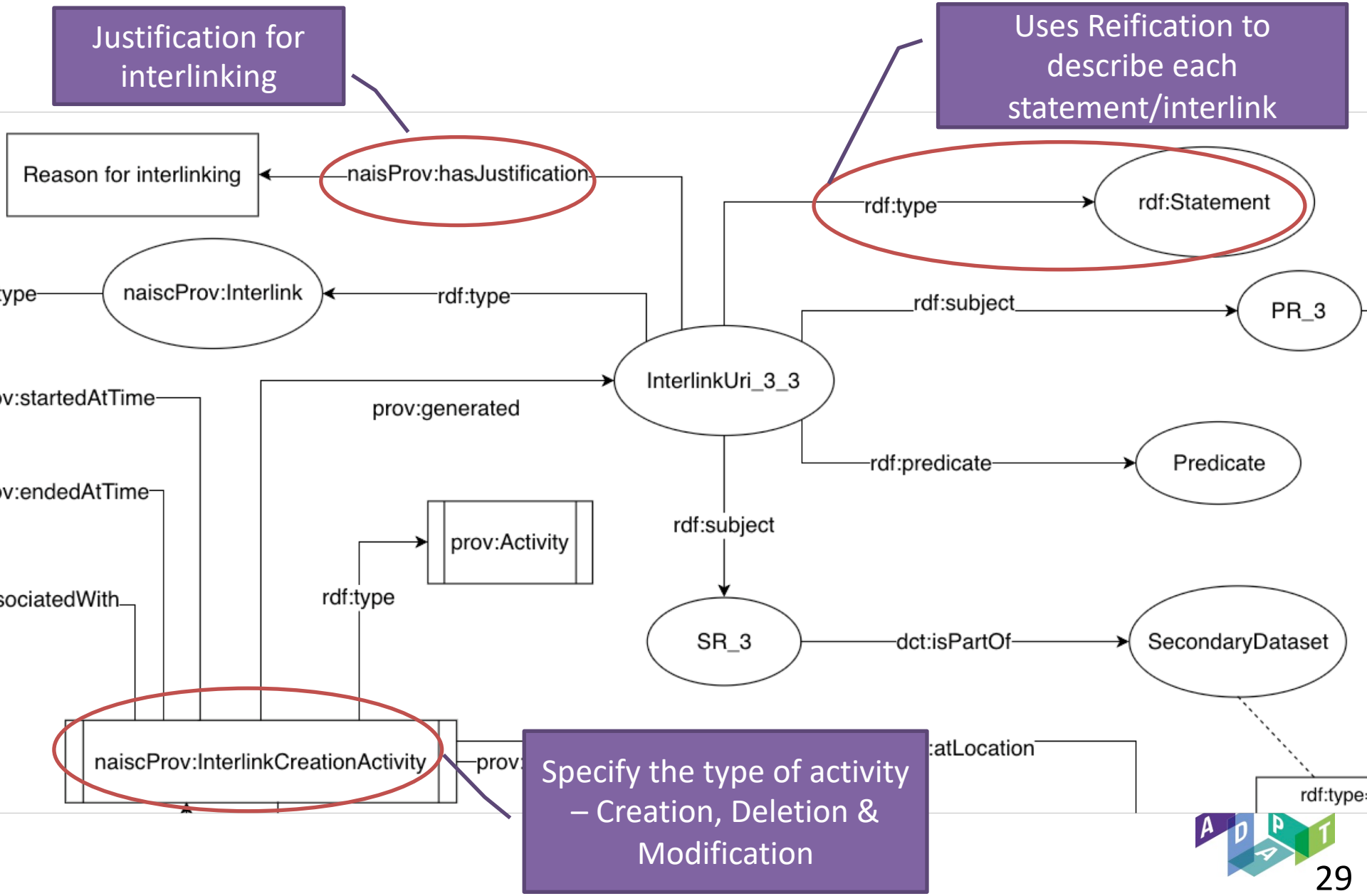
Graph | Interlinks

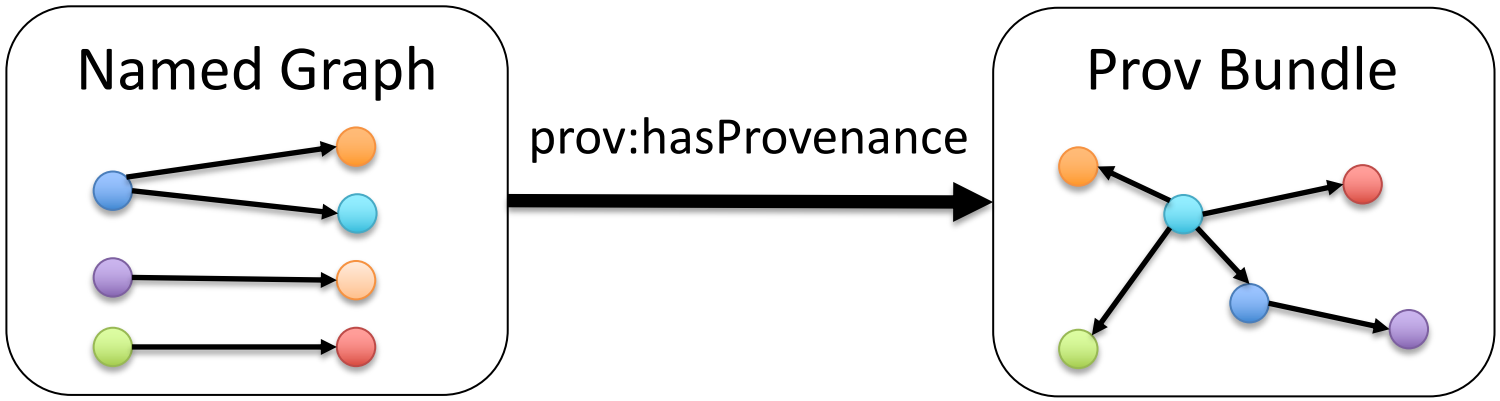
RDF Output

```
<http://digital.ucd.ie/data/ivrla:45153>
        <http://purl.org/dc/terms/isPartOf>
                <http://digital.ucd.ie/data/ivrla:45000> .
```



Naisc – Provenance Model




Named Graph
prov:hasProvenance
Prov Bundle