Core semantic model for generic research activity

Vasily Bunakov

Science and Technology Facilities Council

United Kingdom

STFC

Funds and operates large scale instruments for the UK and visitor researchers in:
- physics, astronomy
- chemistry, materials
- biology, medicine

Scientific Computing develops and operates computing infrastructure:
- High Performance Computing
- Petabyte data store
- CERN LHC Tier 1 hub

also conducts applied research and does software development

Big Facilities for Small Science
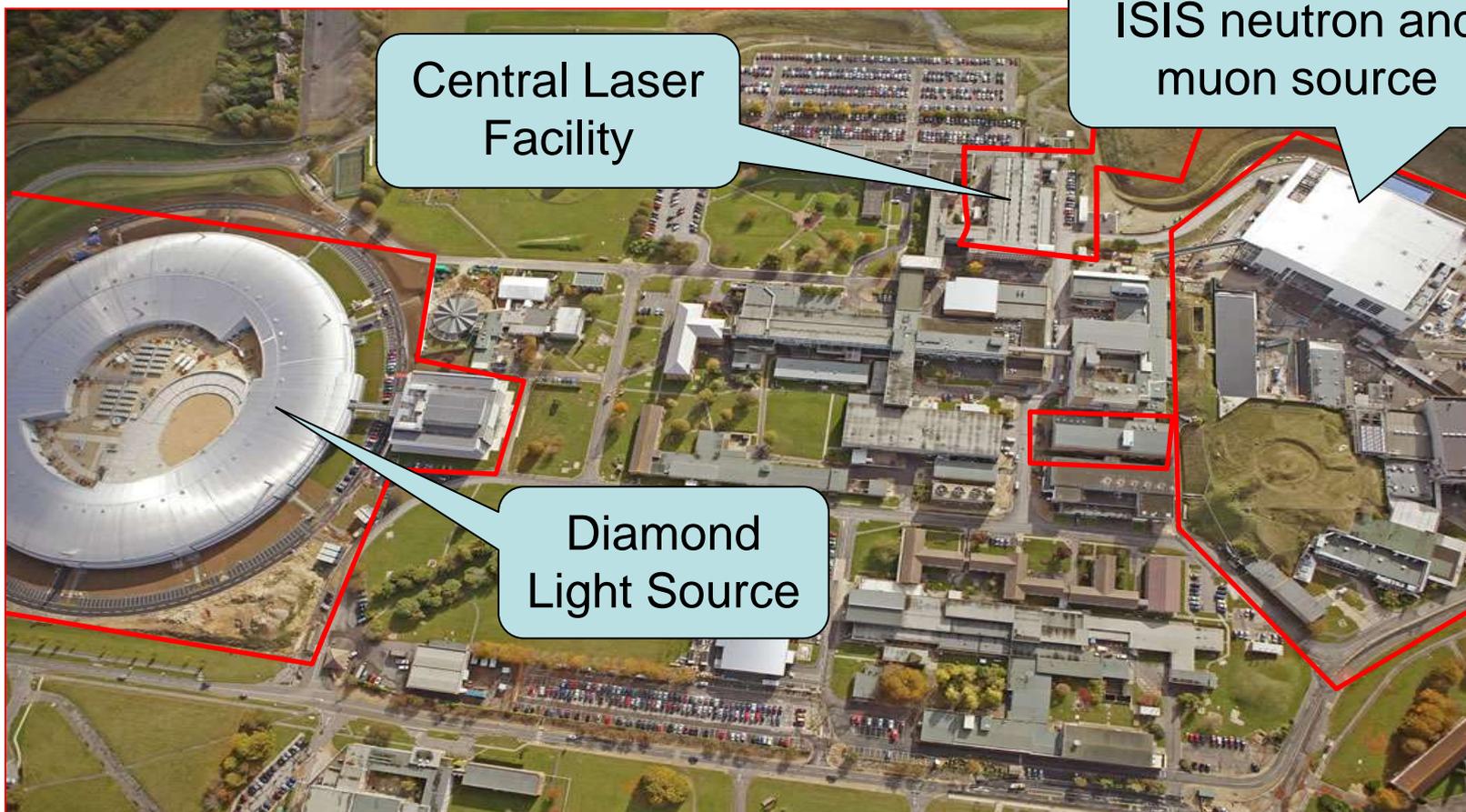
PaNdata projects

PaNdata Europe
2010 – 2011
Preparation: common policies and standards
http://pan-data.eu/pandata/?q=PaNdataEurope

PaNdata ODI
2011 – 2014
Implementation: delivering new infrastructure
http://pan-data.eu/pandata/?q=ODIWP

Facilities Research Lifecycle

Proposal

Approval

Scheduling

Experiment

Data storage

Data analysis

Record Publication

Scientist submits application for beamtime

Facility committee approves application

Facility registers, trains, and schedules scientist's visit

Scientists visits, facility run's experiment

Raw data filtered, and stored

Tools for processing made available

Subsequent publication registered with facility

Data catalogue software: *http://code.google.com/p/icatproject/*

CSMD: Core Scientific MetaData Model
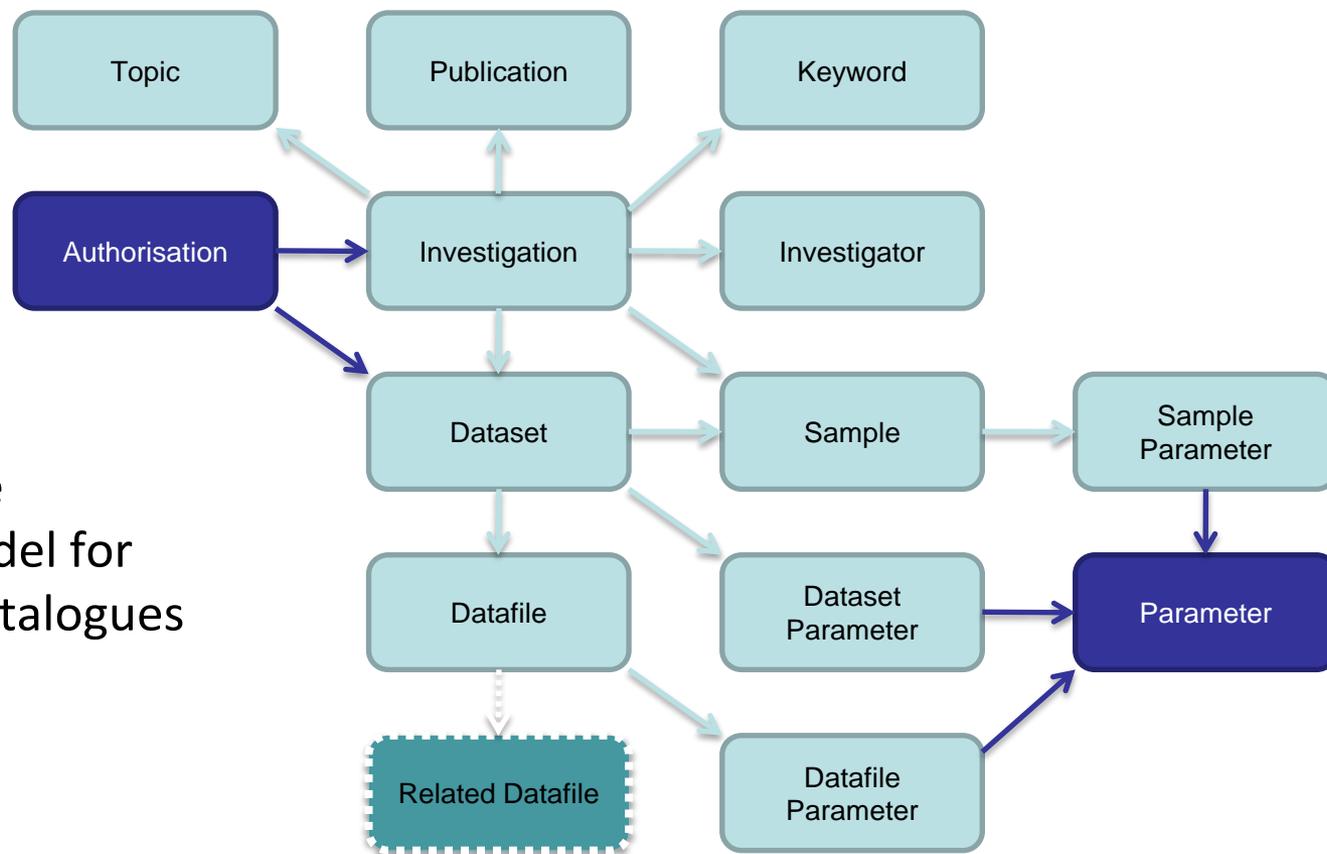
CSMD forms the information model for facilities data catalogues

Topic

Publication

Keyword

Authorisation

Investigation

Investigator

Dataset

Sample

Sample Parameter

Datafile

Dataset Parameter

Parameter

Related Datafile

Datafile Parameter

We joined DataCite

Much cheaper DOIs than directly from DOI Foundation

International DOI Foundation
Member

TIB

Managing Agent

DataCite

carries

MEMBER INSTITUTE

ASSOCIATE STAKEHOLDER

works with

DATA CENTRE (N)
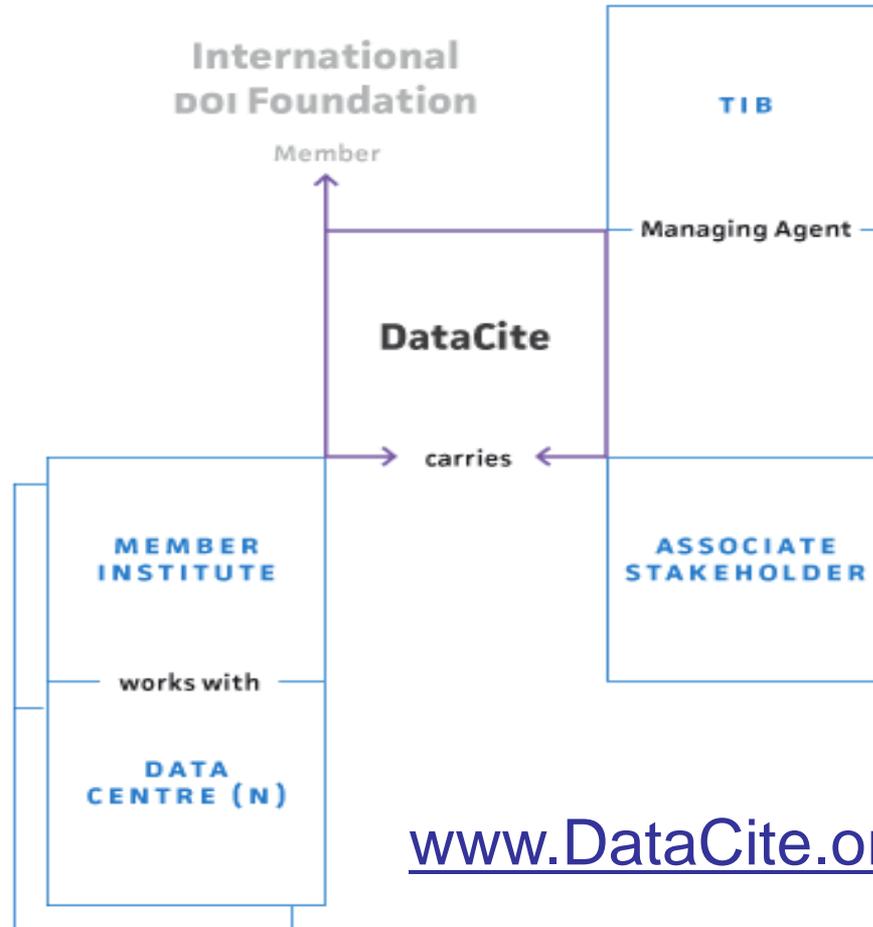
www.DataCite.org

Is it really about data?
Our DOIs landing pages are in fact for Investigations (Experiments)

Red is for "data" notion, and green is for "investigation"

We are not alone in DataCite "abuse"

TUDelft Delft University of Technology · · · **Institutional Repository**

Home · About · Disclaimer · Terms of use ·

[   ] [Search] Options

Help and search tips

Enhanced corrosion protection by microstructural control of aluminium brazing sheet

Attachments
FARID_NOROUZI_AFSHAR.THESIS.pdf (32.0 MB)   🖨 ✉ 𝕏 in t f ➕ 0

Cite or link this publication as: doi:10.4233/uuid:b7113b48-08aa-41a1-871b-bf5248fbf5db   ⟸ DataCite DOI

Author:	Norouzi Afshar, F.
Promotor:	De Wit, J.H.W. · Terryn, H.
Faculty:	Mechanical, Maritime and Materials Engineering
Department:	Materials Science & Engineering
Type:	Dissertation
Date:	2013-08-30
ISBN:	9789077172933
Keywords:	aluminium brazing sheet · accelerated corrosion test · microstructural characterization · electrochemical characterization · heat treatment

We used to think our metadata is for "data" but in fact, quite often it is for "activity", e.g. Experiment or Study

Research activity is not restricted to Experiment or Study and can be a part of a longer "value chain"

DDI record for social
science Study decomposed

Archives:
www.data-archive.ac.uk
www.gesis.org
and many more

DDI portal:
www.ddialliance.org

Research funding

Funding agency
Grant ID

Research per se

Study ID
Study title
Study description
Contributor (author)
Temporal coverage
Spatial coverage
Subject coverage
Related publications

Research distribution

Contributor (distributor)
Copyright
Access type
Access description
Access contact
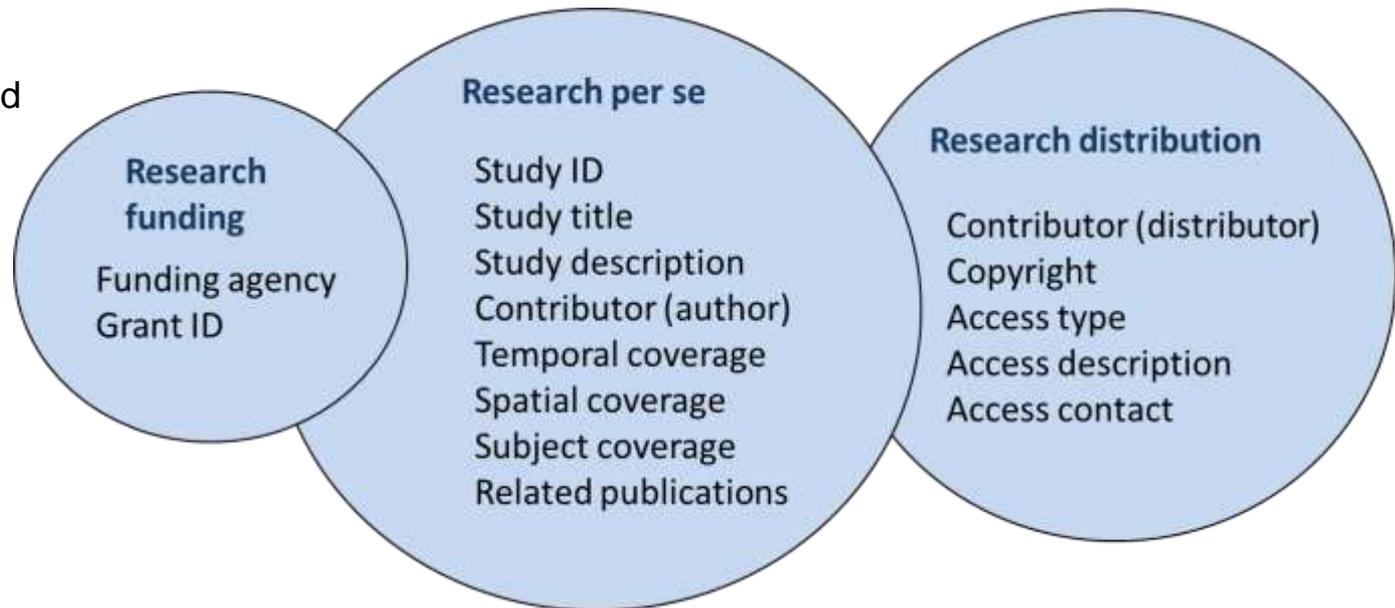
An Infrastructure for Open, Linked Governmental Data Provision
towards Research Communities and Citizens

engage

SEVENTH FRAMEWORK
PROGRAMME

Project: www.engage-project.eu   Platform: www.engagedata.eu

ENGAGE vision: promotion of Open Data to Linked Open Data through collaborative data curation

Project: www.engage-project.eu   Platform: www.engagedata.eu
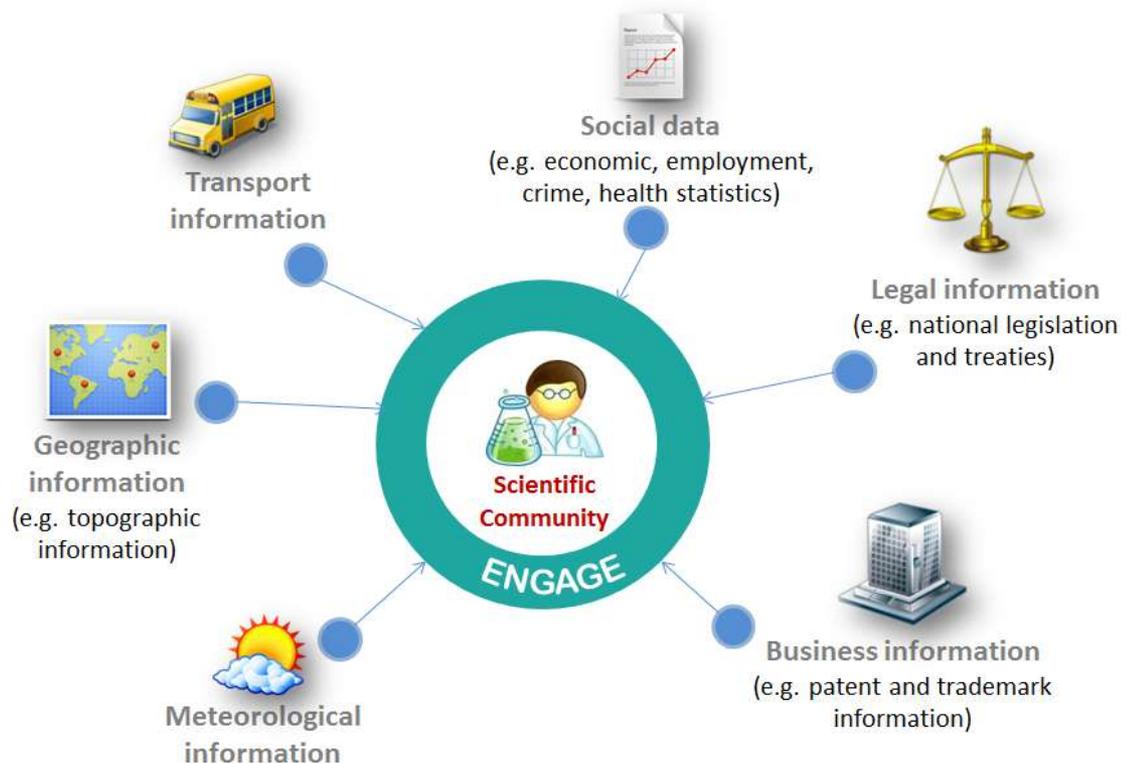
To make research data linkable, we need to <u>reasonably</u> model research activity

- Keep the model generic enough

- Keep it simple for better adoption and "opportunistic" application

- Aim it not at humans only but at machines / software agents, too

Do we have reasonable research activity models?

DARIAH Scholarly Research Activity
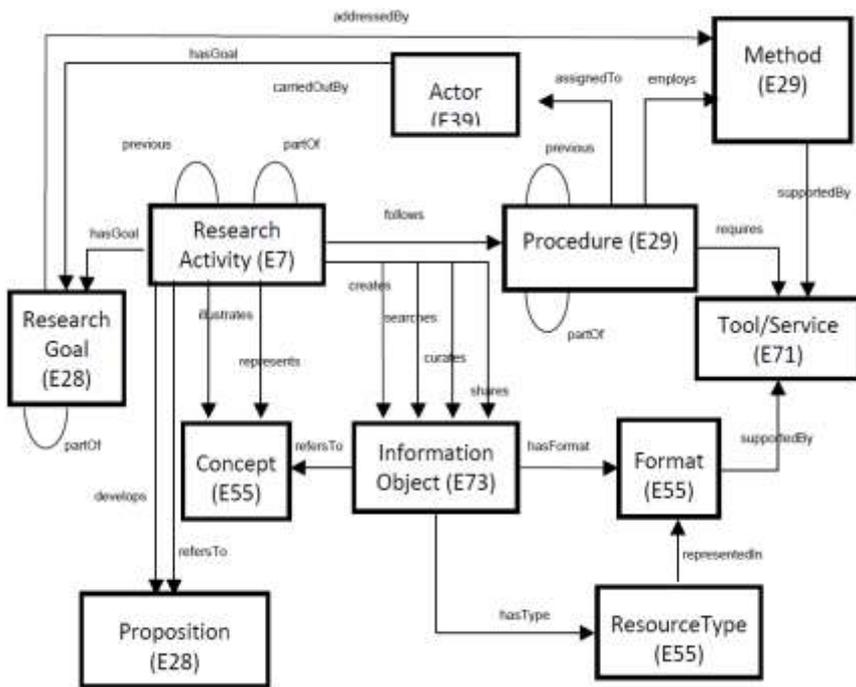
I2S2 Scientific Research Activity Lifecycle
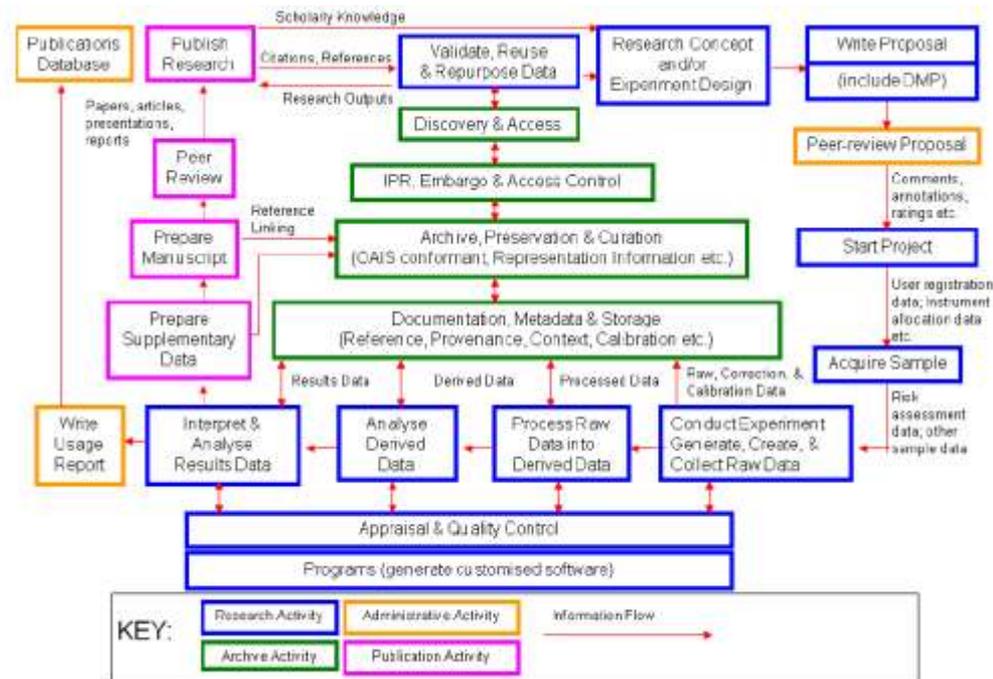
Concerns about existing research activity models

- Domain-specific
- Elements <u>seem</u> well defined but are open to different interpretations
- Are not "Linked Data ready"
- Overdone to be easily adopted and consistently used

Possible response:
offering a (simple) generic research activity model suitable for adoption by different stakeholders
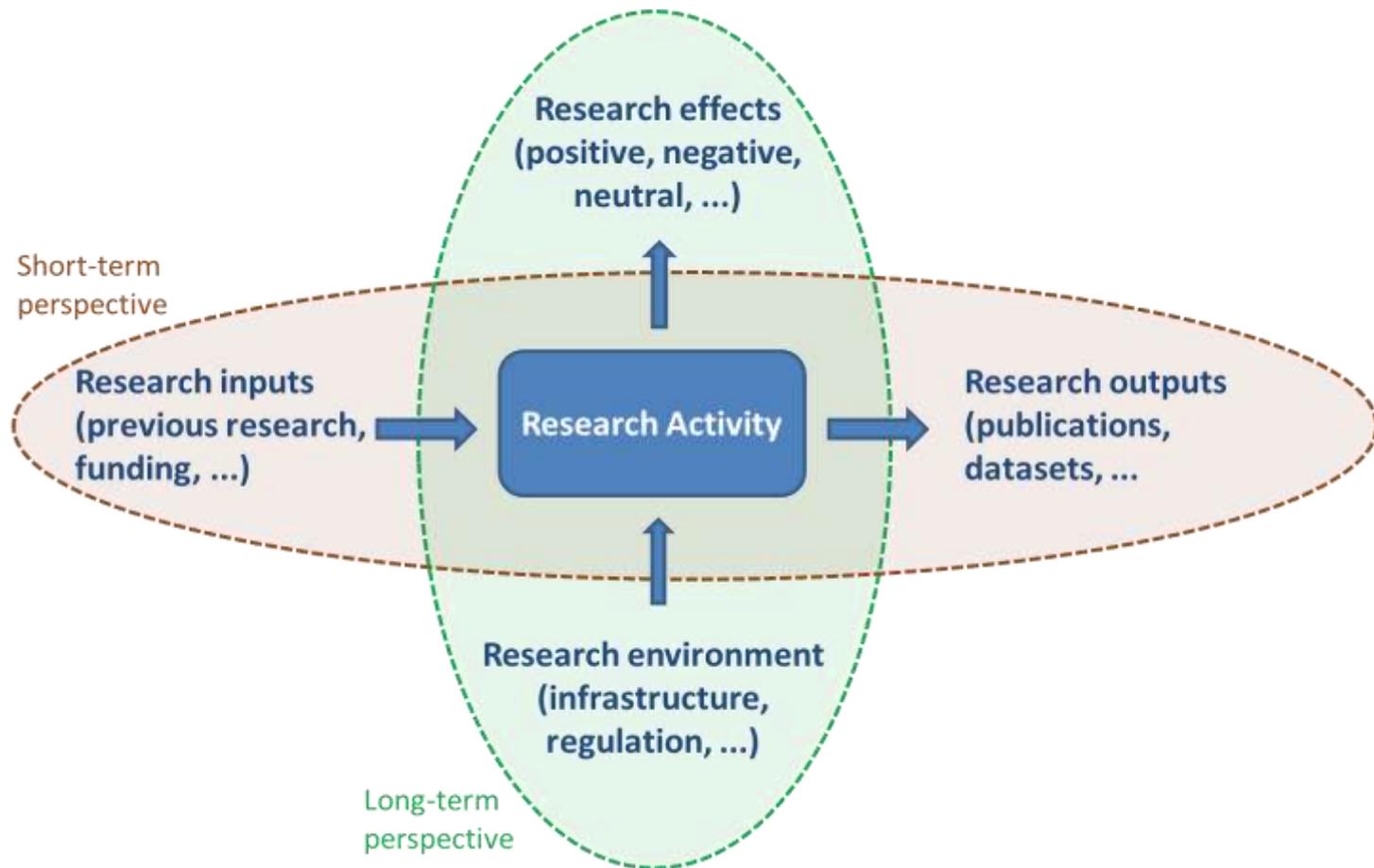
Research activity cell
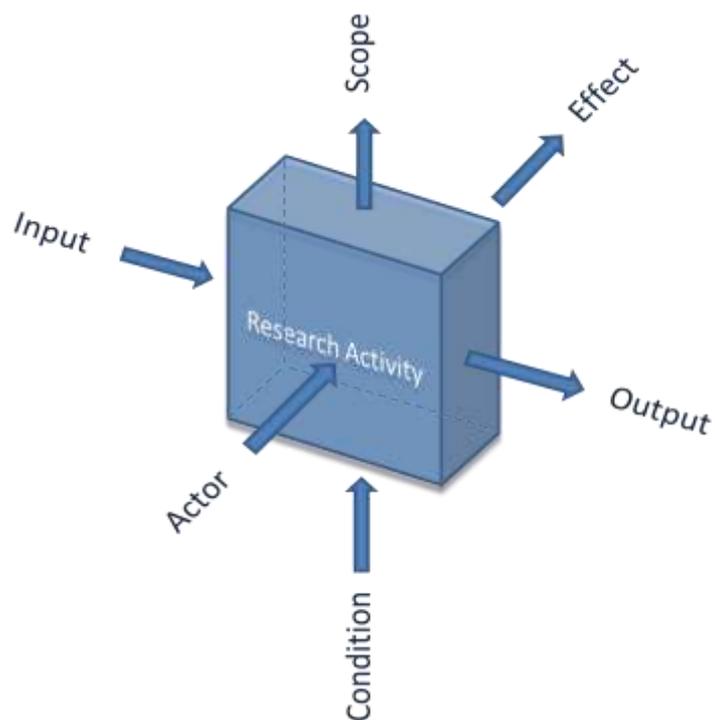
Aspect	Description	Examples	
		Research per se	Research data analysis
Input	Something that is taken in or operated on by Activity	Previous research	Raw data
Output	Something that is intentionally produced by Activity	Raw data	Derived (analyzed) data
Scope	Something that Activity is aimed at or deals with	Sample properties	One or more experiments
Condition	Something that affects or supports Activity, or gives it a specific context	Scientific instrument	IT environment
Actor	Something or somebody who participates in Activity	Investigator	Data analyst
Effect	Something that is a consequence of Activity	Environment pollution	New software module

**What we (different stakeholders of the research lifecycle) actually want to monitor and exploit is "research value chains",
to ensure the golden-eggs-laying goose of research is productive =
brings enough eggs for everyone involved.**

Research activity cells combined in "grid" should result in better research navigation and research contextualization for everyone involved
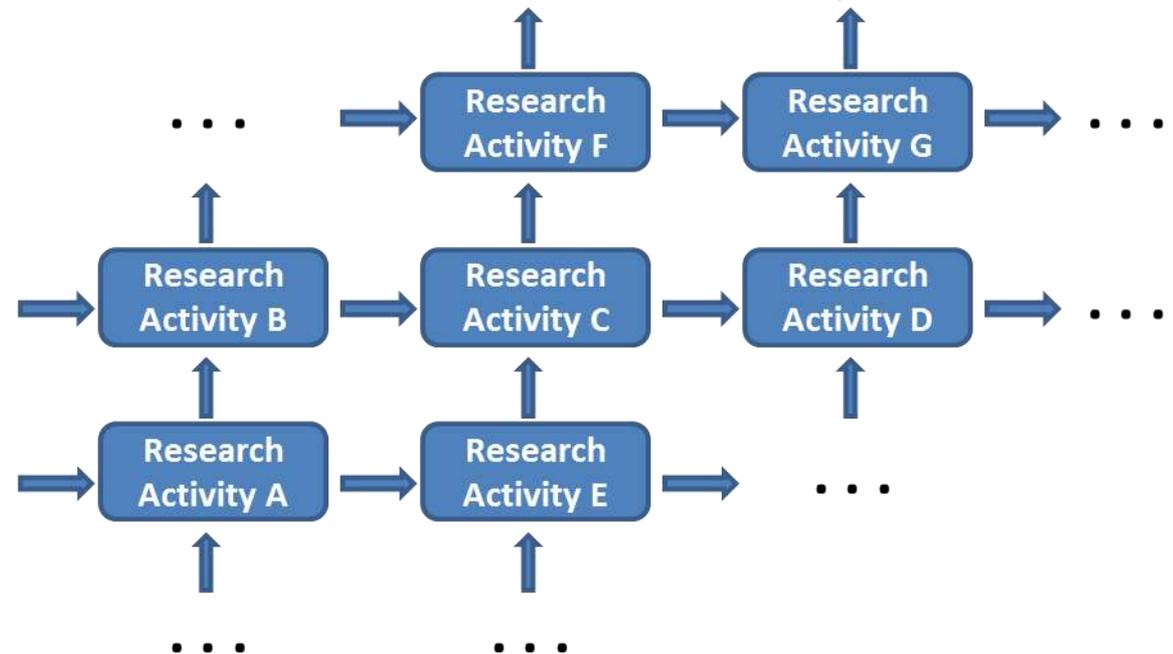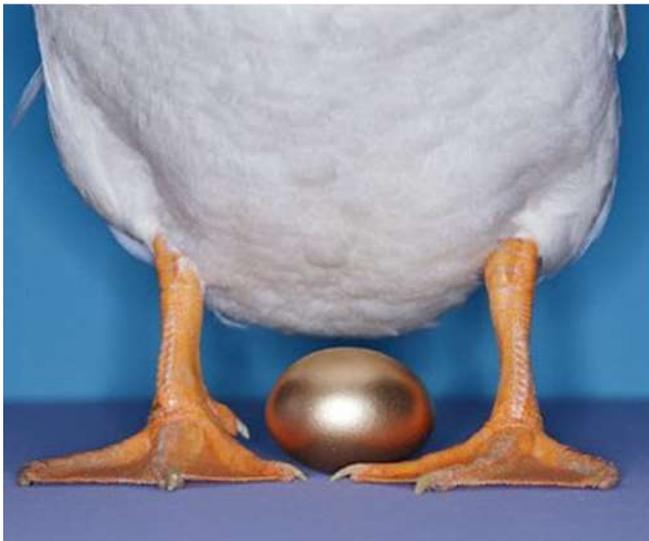
RDFS Plus representation (see in paper) and model extensions

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix rm: <http://example.org/stuff/ResearchModel#>.
# For Conditions
rm:Regulation  rdfs:subClassOf  rm:Condition .
rm:DataManagementPolicy  rdfs:subClassOf  rm:Regulation .
# For Output
rm:Publication  rdfs:subClassOf  rm:Output .
rm:Dataset  rdfs:subClassOf  rm:Output .
# For Scope
rm:ExperimentalTechnique  rdfs:subClassOf  rm:Scope .
rm:SubjectCoverage  rdfs:subClassOf  rm:Scope .
# For properties
rm:activity_location  rdfs:subPropertyOf  rm:hasScope .
rm:activity_subject  rdfs:subPropertyOf  rm:hasScope .
```

SPARQL queries in support of use cases

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix rm: <http://example.org/stuff/ResearchModel#>.

\# How much research output, and how much of each type is out there:

SELECT ?output_type (COUNT(?output) as ?total)

WHERE {?output_type rdfs:subClassOf rm:Output .

   ?output a ?output_type .

  }

GROUP BY ?output_type

\# Discover the chains of interrelated activities:

SELECT ?previous_activity ?current_activity

WHERE {?previous_activity rm:hasOutput ?output .

           ?output am:inputFor ?current_activity .}

Possible application: research provenance
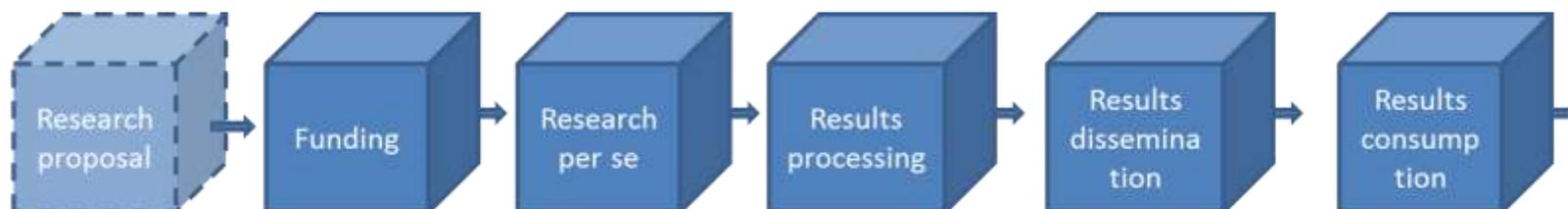
RA model class	Funding	Research per se	Results processing	Results dissemination	Results consumption
Input	Research proposal	Award (grant)	Dataset	DDI record	DDI record or its manifestation
Output	Award (grant)	Dataset	DDI record	Web service	Feedback
Actor	Researcher candidate	Contributor (author)	Data archive	Dissemination service	Web service user
Effect	Researcher's department budget	Whatever is claimed in proposal	Economical effect of processing	Economical effect of distribution	Impact on further research
Condition	Funding body rules & regs	Microdata regulation	Data processing guidelines	Data access regulation	Research purpose statement
Scope	Certain branch of science	Certain geolocations	National research	International research	Certain HASSET keywords

Collaborative curation of research data in "cloud of clouds"
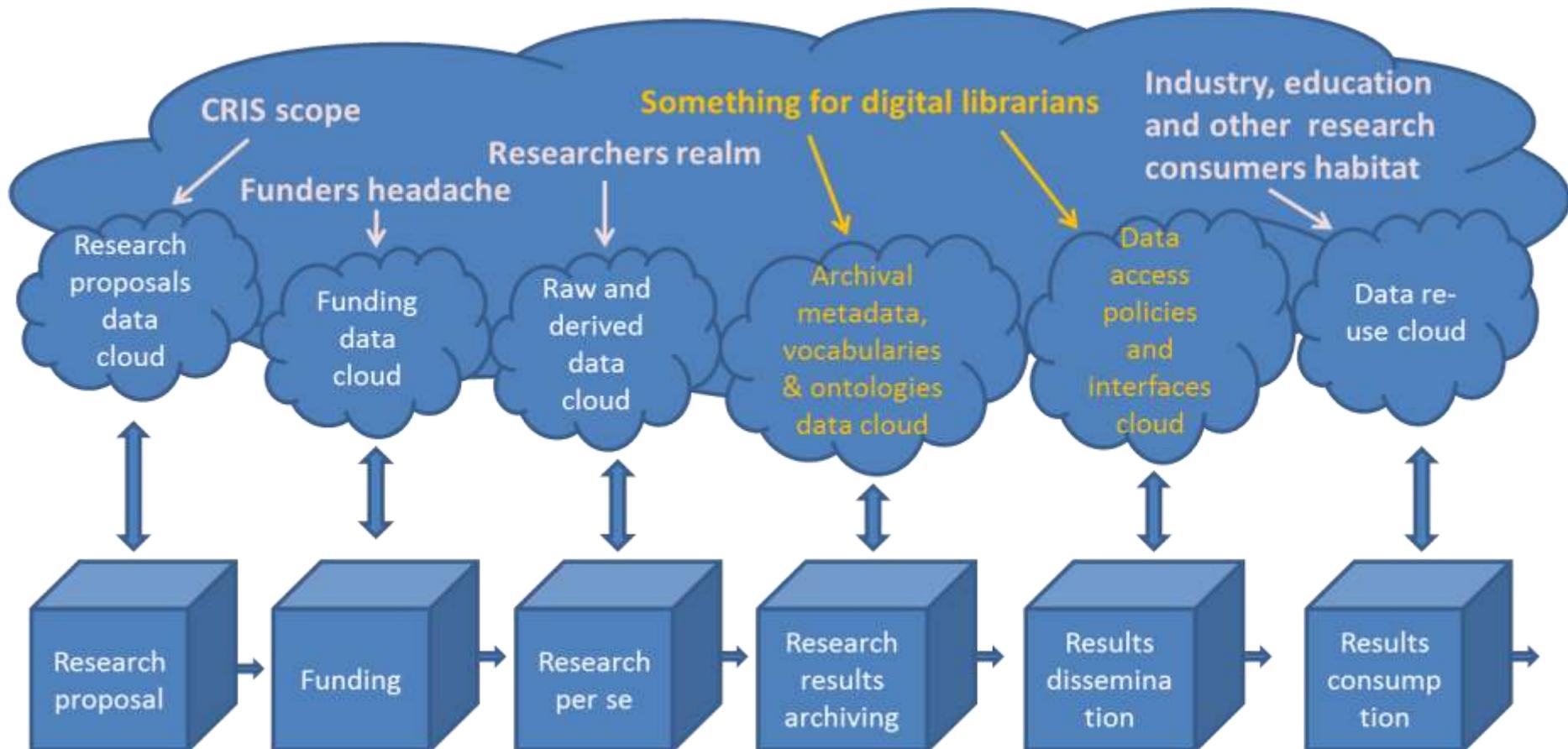
The model selling points ☺

- Small
- Extendable
- Allows widely adopted RDFS Plus manifestation
- (Right) balance between simplicity and expressivity
- (Right) balance between modeller's freedom and results interpretability

Use cases for applying the model

- Research provenance, navigation and contextualization

- Semantic analysis and annotation of domain-specific metadata (DDI, CSMD, …)

- Distributed discovery, curation, and re-use of the research information

- Long-term digital preservation

Thank you!