Building and preserving a research object

Brian Matthews, STFC

Facilities Data Lifecycle

As in PanData-ODI – D6.1
(which has much more detail)

Proposal

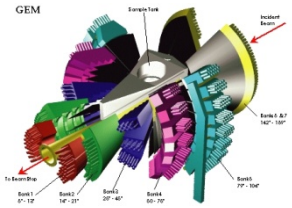
Scientist submits application for beamtime

Approval

Facility committee approves application

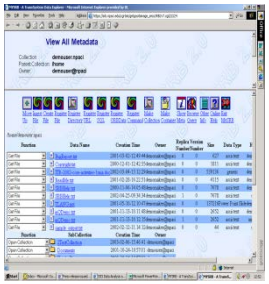
Scheduling

Facility registers, trains, and schedules scientist's visit



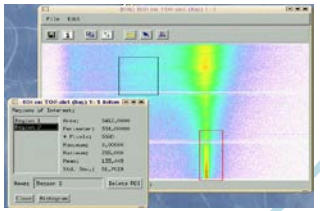
Experiment

Scientists visits, facility run's experiment



Data storage

Raw data filtered, and stored



Data analysis

Tools for processing made available

Record Publication

Subsequent publication registered with facility

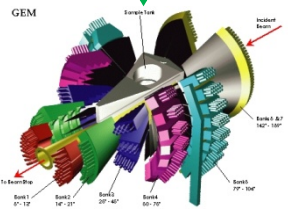
http://code.google.com/p/icatproject/

After proposal: Initialise the Research Object

:investigator

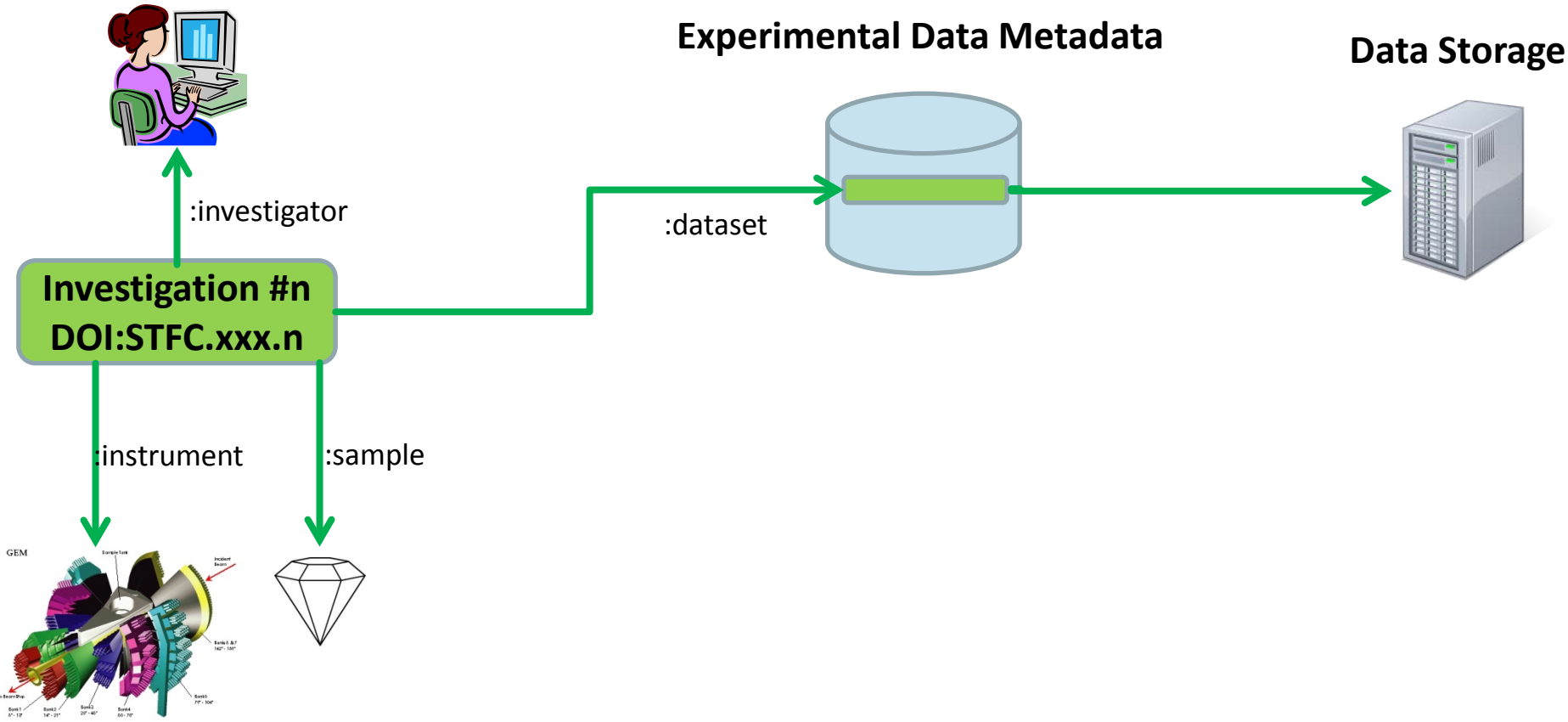
Investigation #n
DOI:STFC.xxx.n

:instrument



- Assign (but not necessarily register) a DOI for the object
- Take basic investigator and instrument information from the proposal system
- Also link to funding

After the experiment



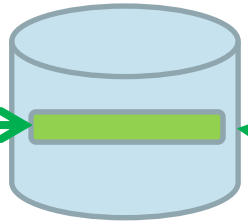
- Own metadata format (CSMD)
- More or less what ICAT currently supports
- Adds extra details on parameters, datasets, formats etc.

Linking the Publications

:investigator

Investigation #n
DOI:STFC.xxx.n

:dataset



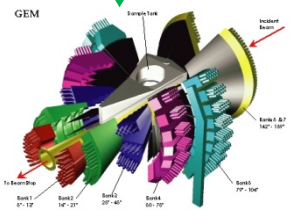
Experimental Data Repository

cito:cites

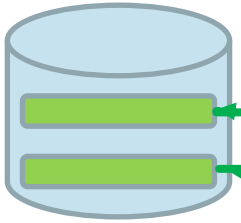
cito:cites

:instrument

:sample



Publication Repository



Publication Store

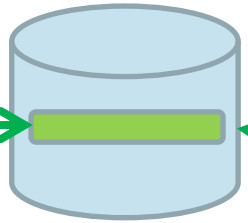
- Own metadata format (CSMD)
- cito – citation ontology (Oxford)
- Would also need to take into account pub metadata
- Publication repository could be on a different site

Linking Publication into Research Object

:investigator

Investigation #n
DOI:STFC.xxx.n

:dataset



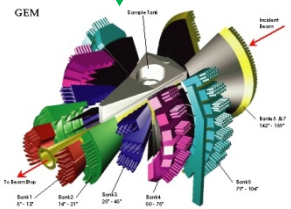
Raw Data Repository

cito:cites

cito:cites

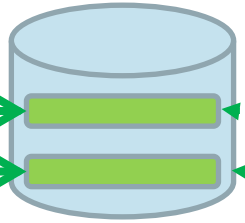
:instrument

:sample



:publication

:publication



Publication Repository

- Own metadata format (CSMD)

Adding Derived Data

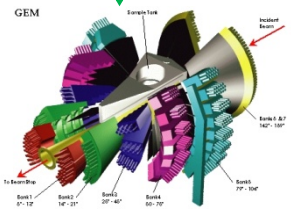
:investigator

Investigation #n
DOI:STFC.xxx.n

:dataset

:instrument

:sample



:publication

:publication

Raw Data Repository

Derived Data Repository

Publication Repository

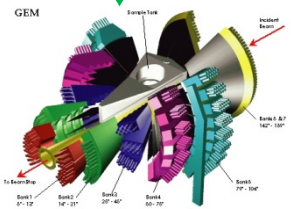
- Own metadata format (CSMD)
- Note that derived data could be on a different site

Linking the derived data into the Research Object

:investigator

Investigation #n
DOI:STFC.xxx.n

:instrument



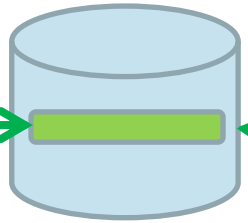
:sample

:dataset

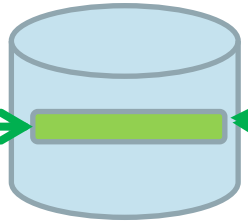
:relatedDataset

:publication

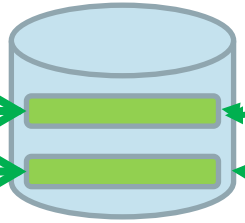
:publication



Raw Data Repository



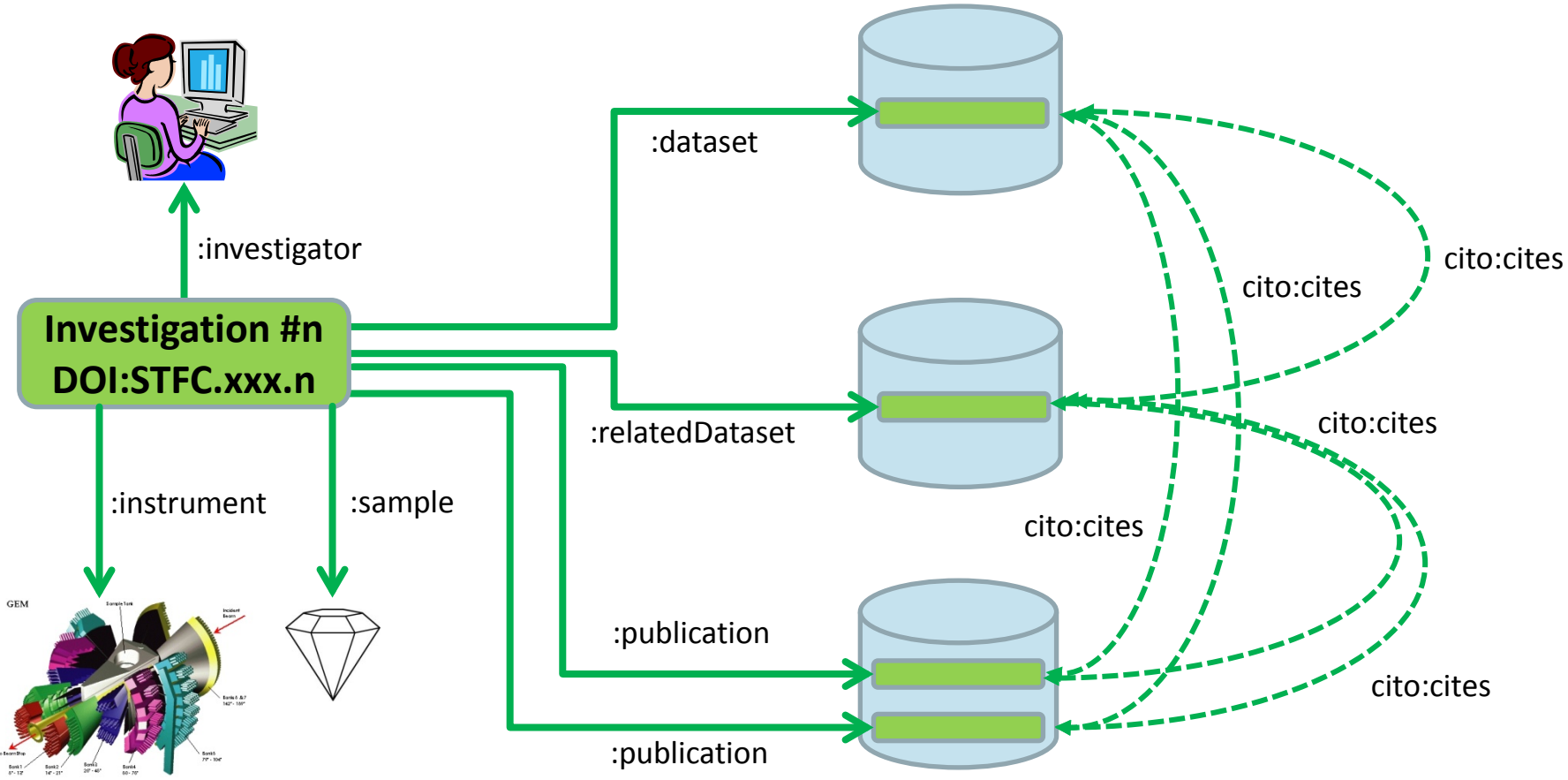
Derived Data Repository



Publication Repository

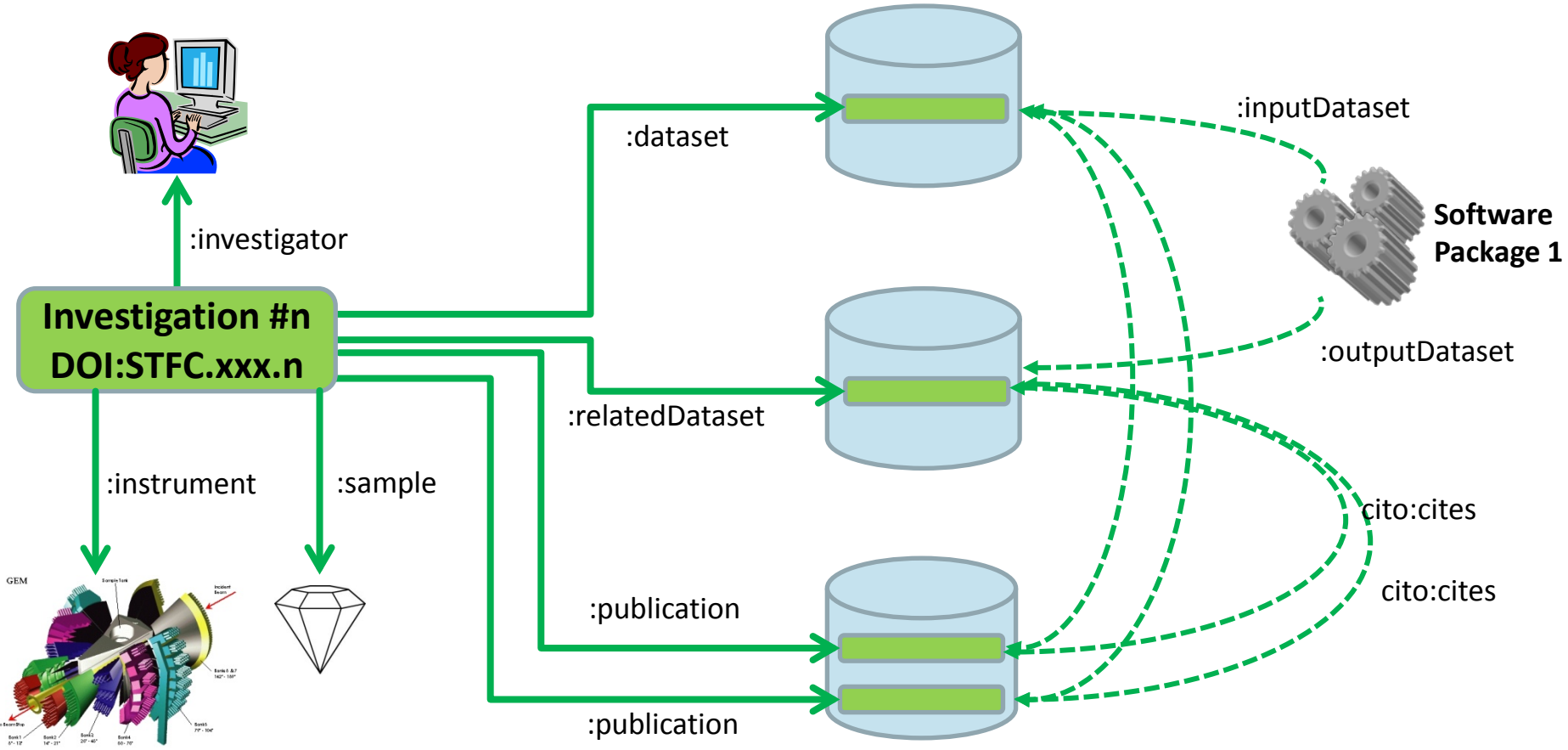
- Own metadata format (CSMD)
- Represent this in OAI-ORE ?

With all the cross linking



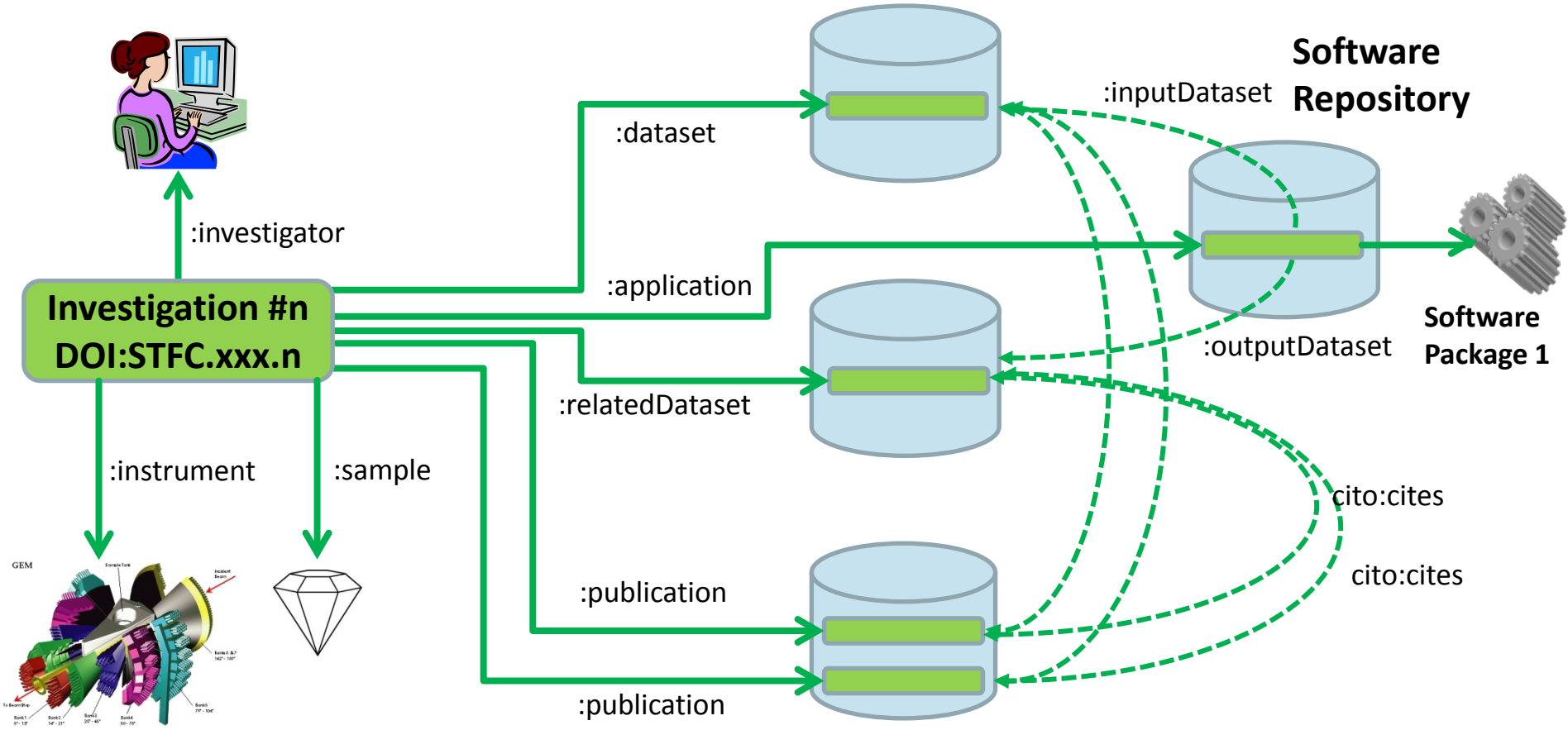
- Own metadata format (CSMD)
- OAI-ORE
- Cito

Provenance of related dataset via a software application



- Own metadata format (CSMD)
- OAI-ORE
- Relate to W3C Prov ontology

Linking the software application into the research object



- Own metadata format (CSMD)
- OAI-ORE
- W3C Prov ontology
- Assume that the software is in a repository

Adding Preservation Information – Rep Info for various items

Raw data format description (e.g. NeXus)

Parameter description (e.g. NXDL, Con Vocab)

:investigator

Investigation #n
DOI:STFC.xxx.n

:dataset

:application

:relatedDataset

:instrument

:sample

:publication

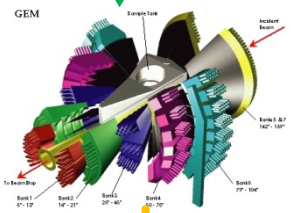
:publication

Software classification

Software description

Analysed data format description

Publication format description



Sample description

Instrument description (website)

- Would probably be more
- Work into a RepInfo Repository
- Would also have a RepInfo Network
- The Research Object itself would be piece of RepInfo