

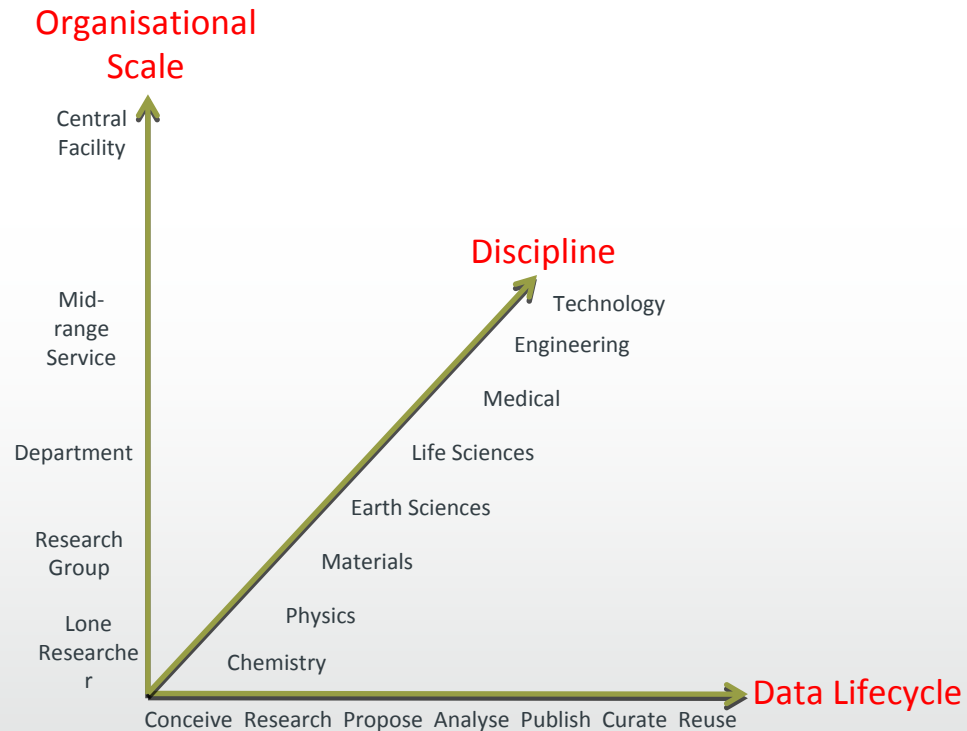
Case Study 1

Probing Data Lifecycle and Inter-
Institutional Issues.

Simon Coles (s.j.coles@soton.ac.uk)
5th May 2011

Aims

- Increase efficiency and cope with scale across institutions
- Support whole data lifecycle management so that processes become seamless
- Map across organisational infrastructures
- Develop from established starting platforms



Challenges

- Researchers capture, manage, discuss & disseminate data in relative isolation with highly fragmented data infrastructures and poorly integrated software applications
- Conventional systems of publication lead to insufficient provenance of results and irreproducible experiments
- The processes for recognition lead to a lack of inclination and incentive to share or make all the supporting information for a study publicly available
- Low awareness of data curation and preservation issues leads to data loss and reduced productivity

The Lone Researcher

- Considers research as ‘linear’ &/or ‘individual’ processes
- Considers data volumes to be manually manageable
- Relatively low-powered instrumentation
- Is less concerned with sharing
- Has little or no support infrastructure



Conceive

Research

Propose

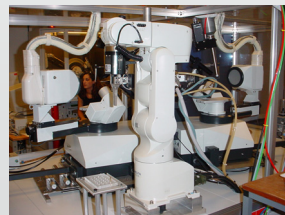
Experiment

Analyse

Publish

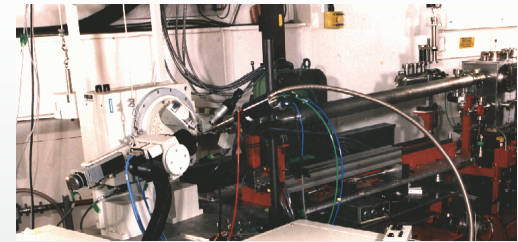
Mid-range Facility

- Service provision function / working as a team
- Technique focused
- High powered lab-based instrumentation
- Operating across institutions
- Moderate infrastructure
- Local data management solutions



Central Facilities

- Operate on behalf of multiple institutions (community)
- Have established and often bespoke & formulaic experimental processes
- Very powerful custom instrumentation
- Have large supporting infrastructure
- Users take data away (different mechanisms)
- Don't often know what happens to data



Proposal

Approval

Schedule

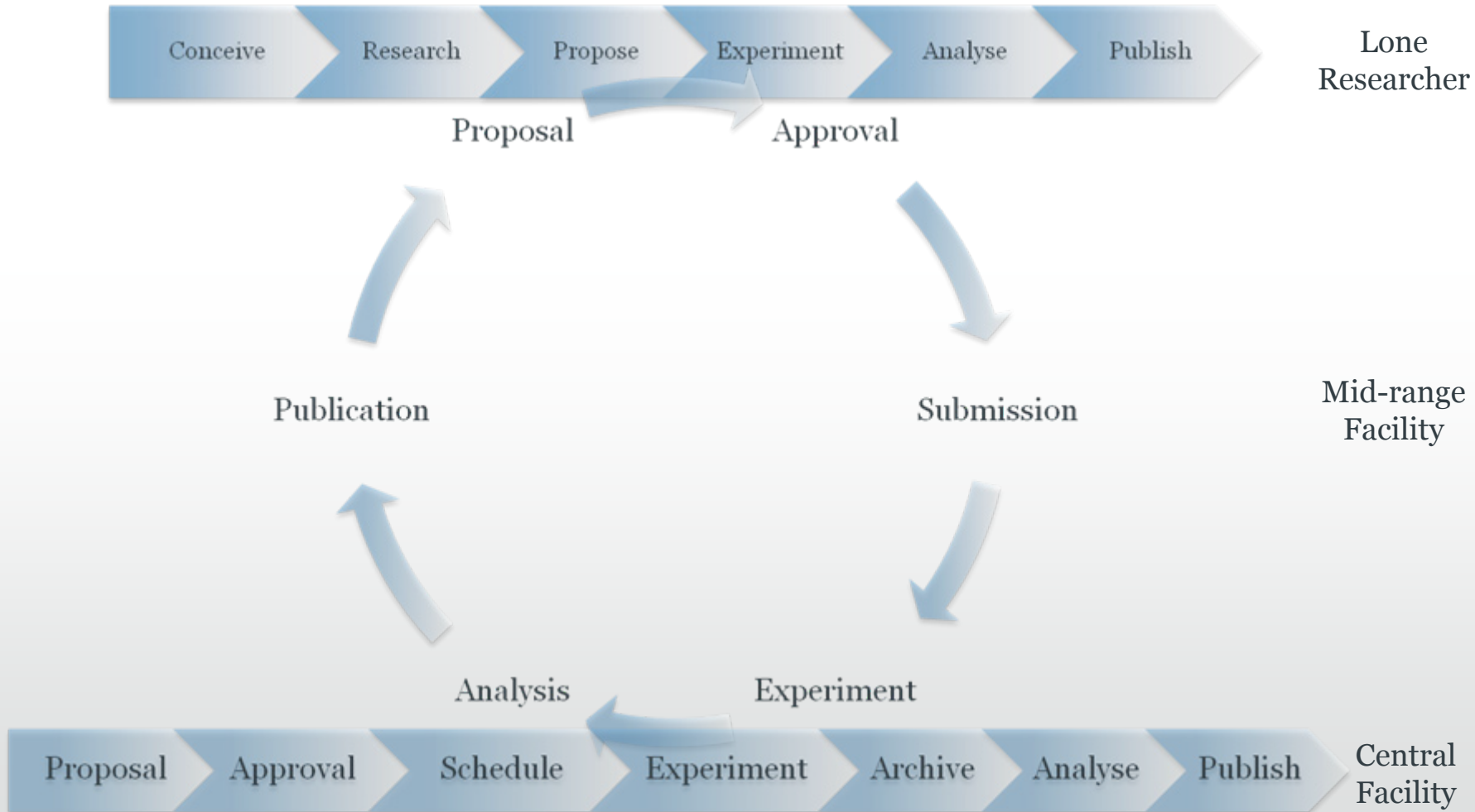
Experiment

Archive

Analyse

Publish

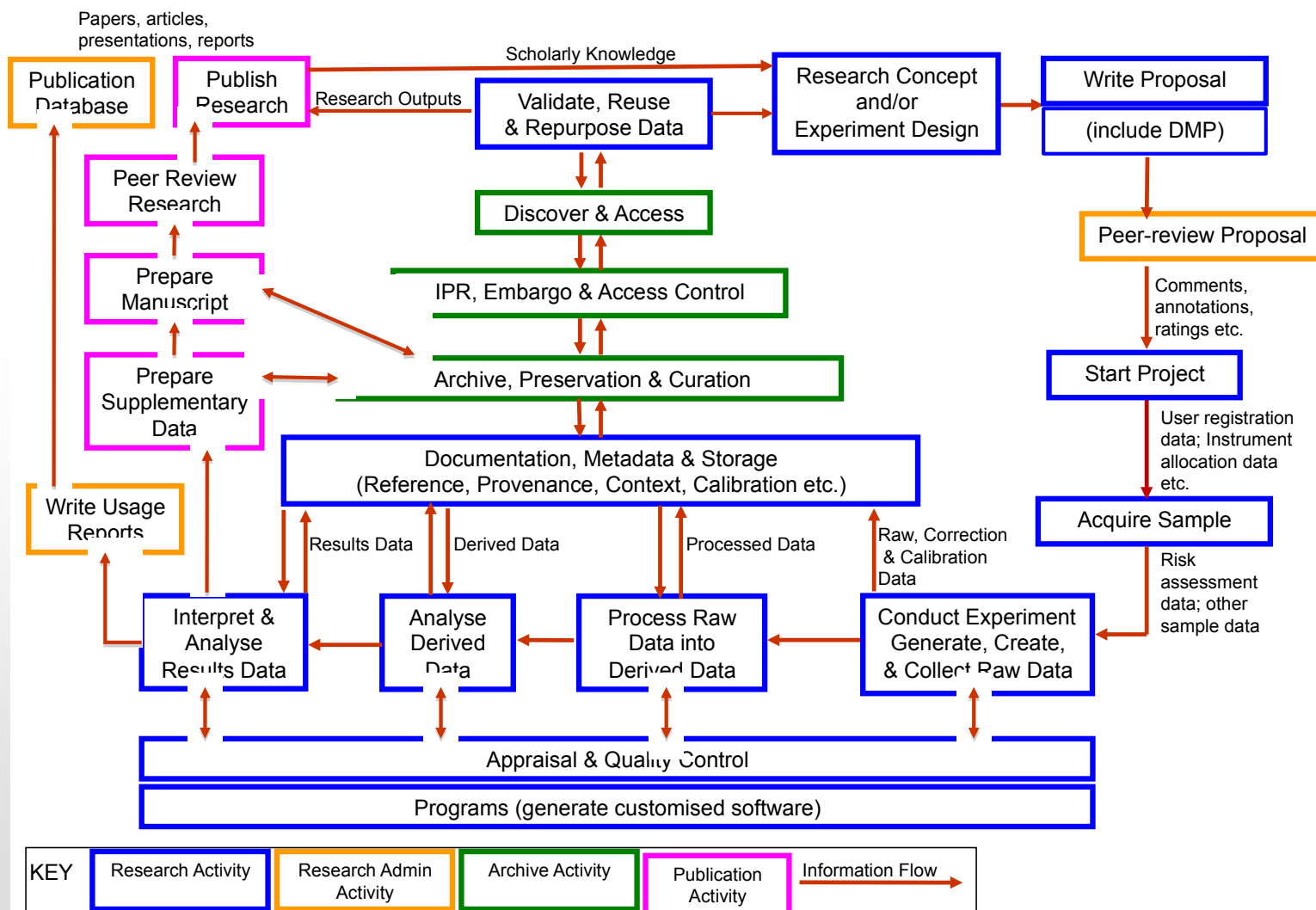
The Bigger Picture



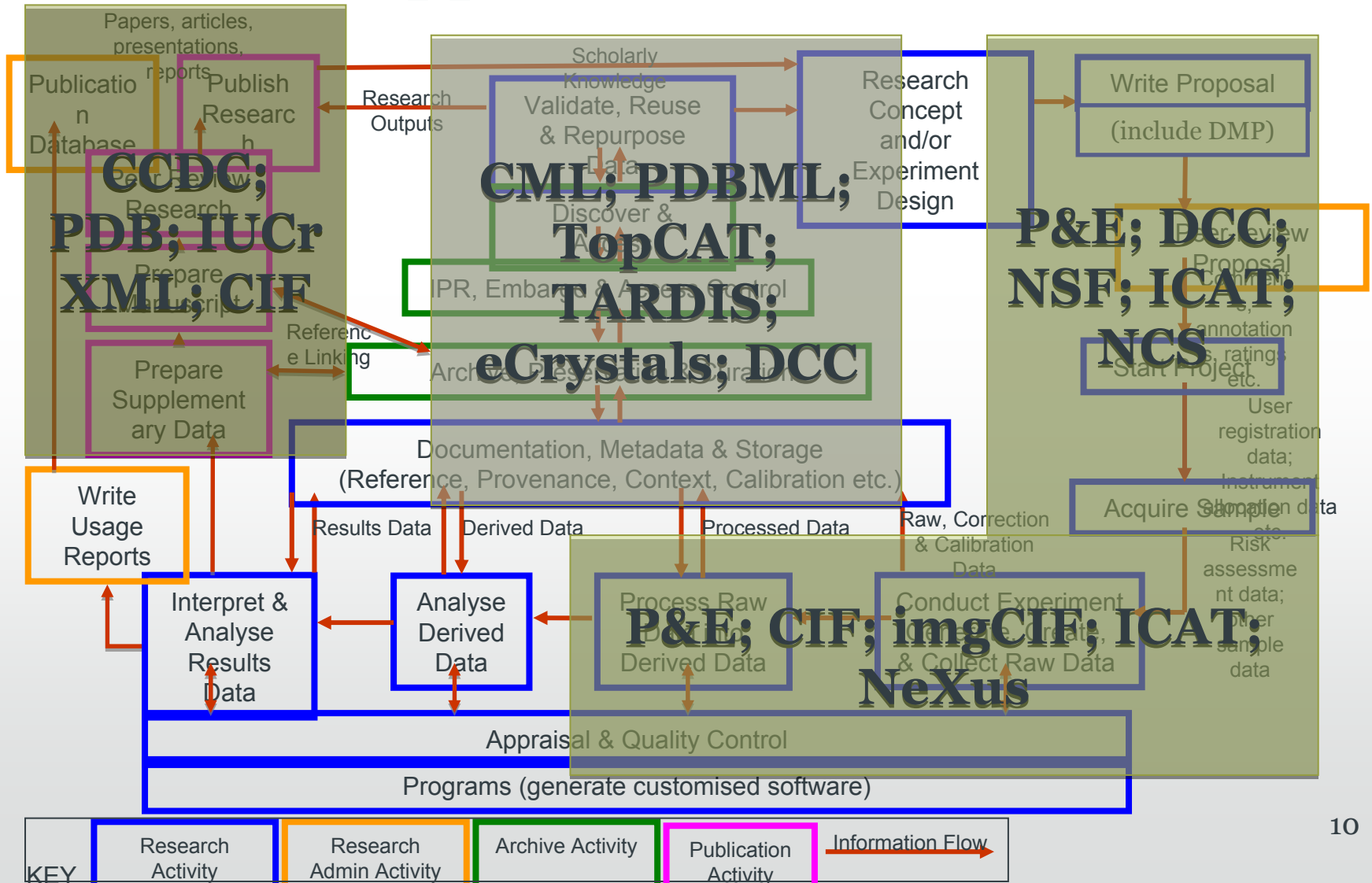
Our Approach

- Model the research processes
- Analyse established working practice
- Adopt and develop established systems
- Examine the interfaces between infrastructures
- Develop a harmonised generic model that fits all infrastructures

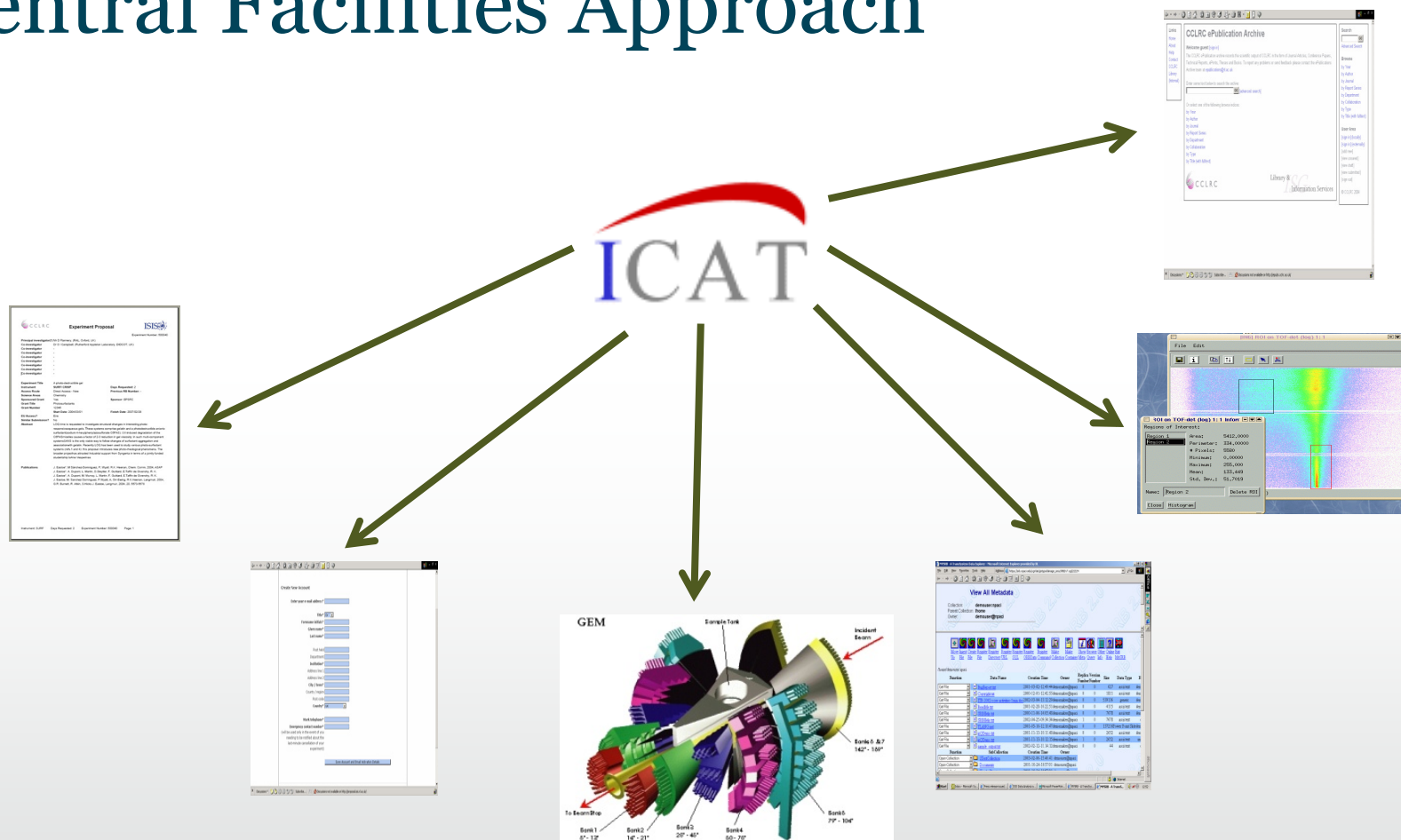
Idealised Research Activity Model



Established Approaches



Central Facilities Approach



Proposal

Approval

Schedule

Experiment

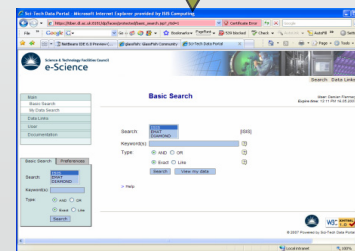
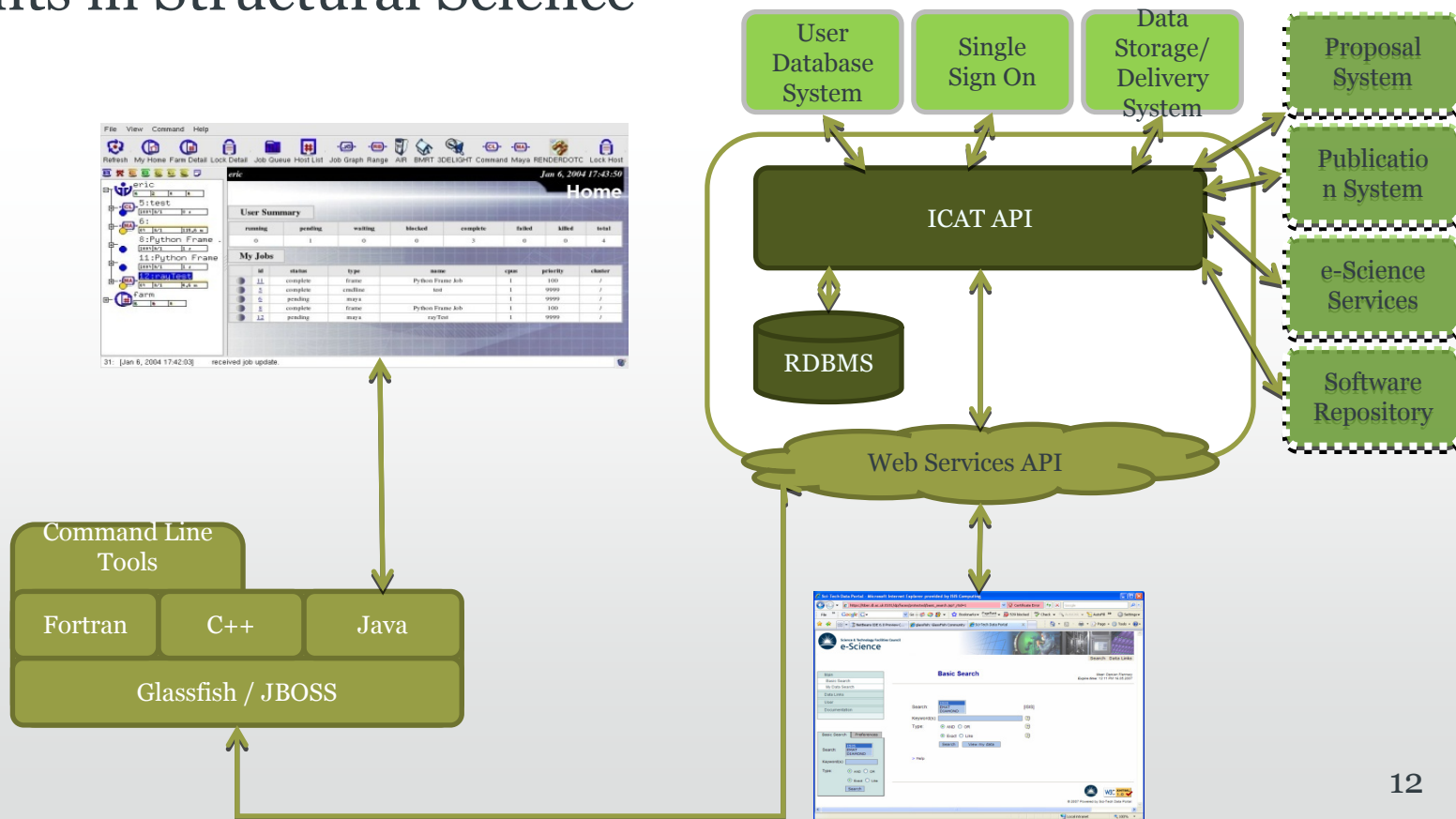
Archive

Analyse

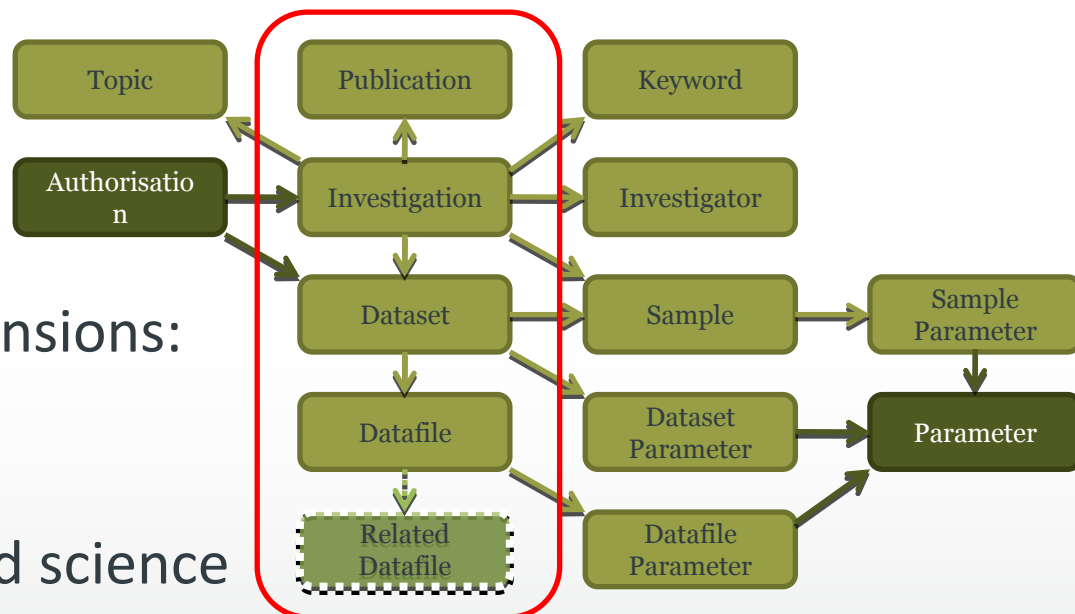
Publish

ICAT

The Core Scientific Metadata model forms the information model for ICAT & is designed to describe facilities-based experiments in Structural Science



CSMD as a Starting Point



- Forms the basis for extensions:
 - To derived data
 - To laboratory based science
 - To secondary analysis data
 - To preservation information
 - To publication data

Planning and Enactment

Rationalisation of objects by way of the
explicit statement of intent and action
(Mark Borkum).

“In theory, there is no difference between theory and practice. But, in practice, there is.” Unknown (possibly Yogi Berra)

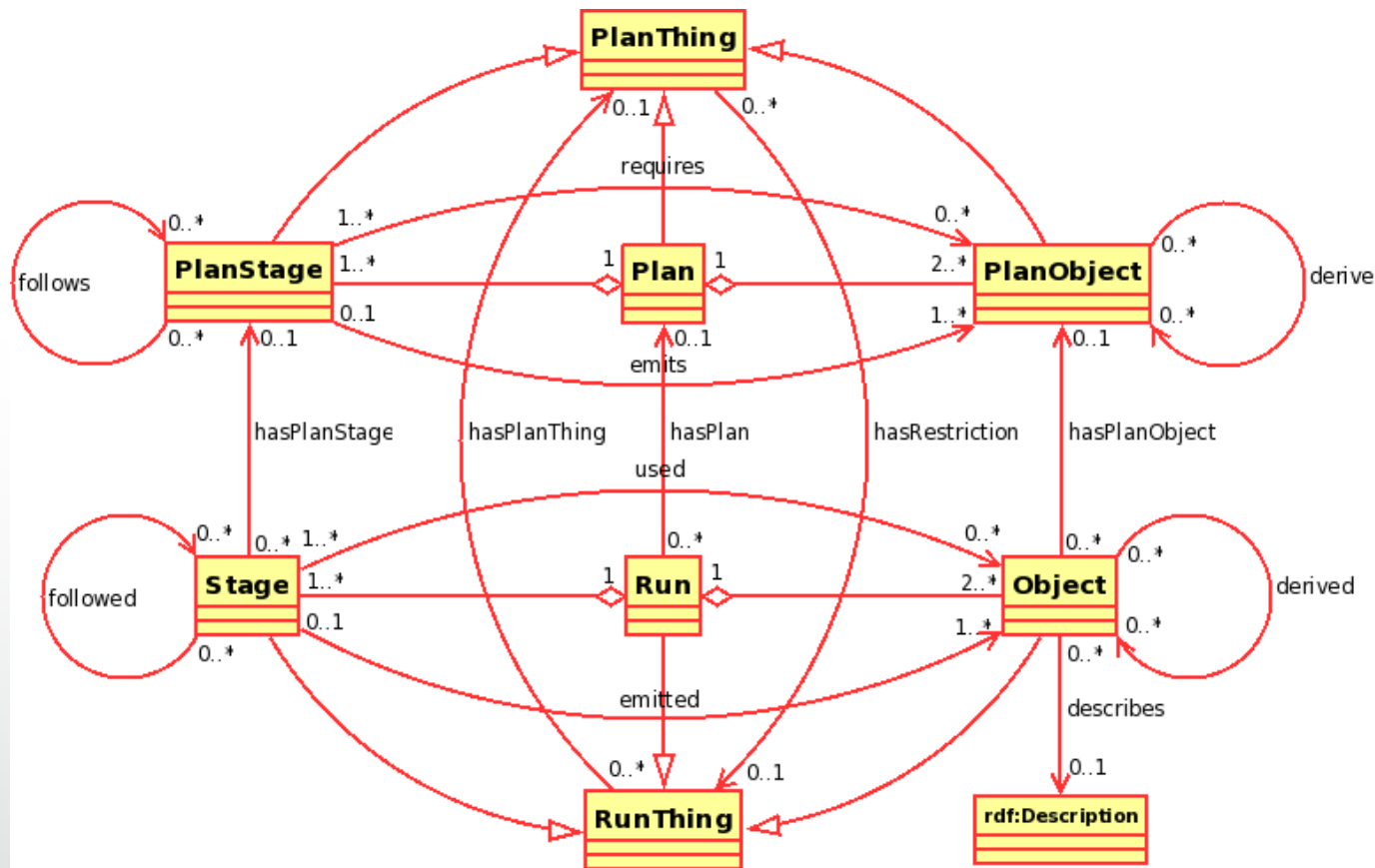
Introduction

- Q: What defines a Semantic Web application?
- A: 3x concepts:
 1. Identity
 - *Who*
 2. Vocabulary
 - *What*
 3. Provenance
 - *Where, When, Why and How*



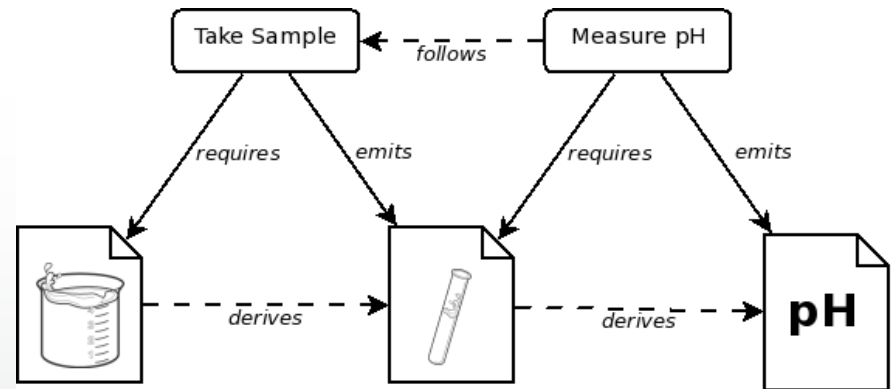
http://memory-alpha.org/wiki/File:Sickbay_laboratory.jpg

Planning and Enactment Ontology – UML



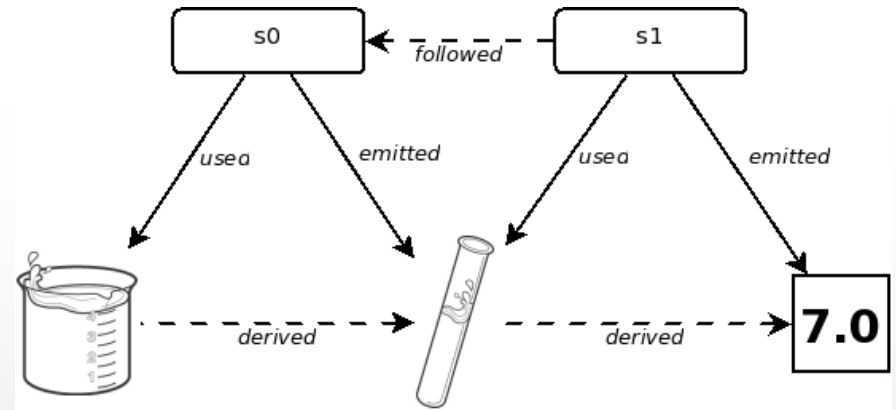
Planning

- Prospective provenance
- Describes a method that *will be enacted* (in the future)
- Three entity types:
 - Plan
 - Plan Stage
 - Plan Object



Enactment

- Retrospective provenance
- Describes a method that *was enacted*
- Three entity types:
 - Run
 - Stage
 - Object

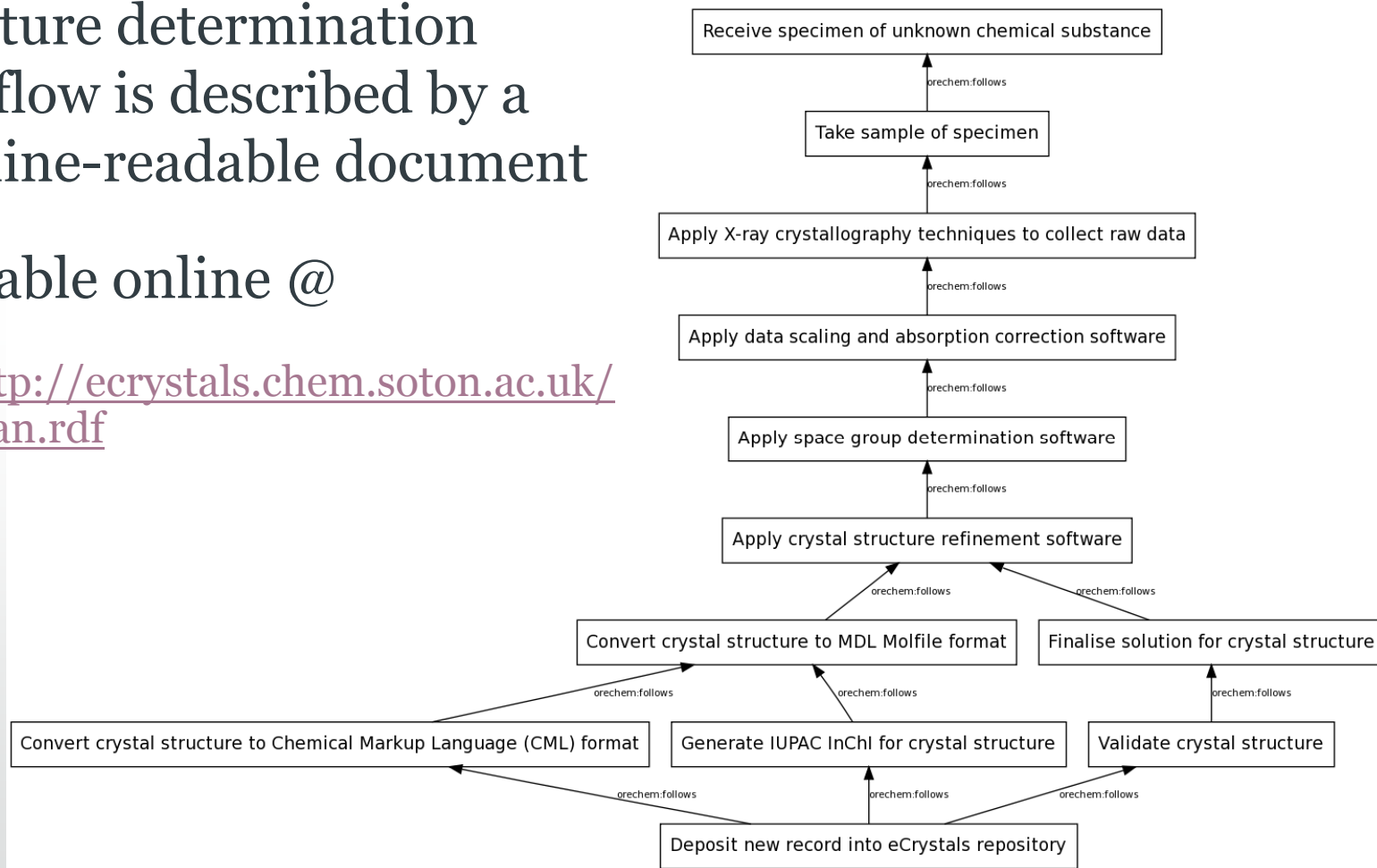


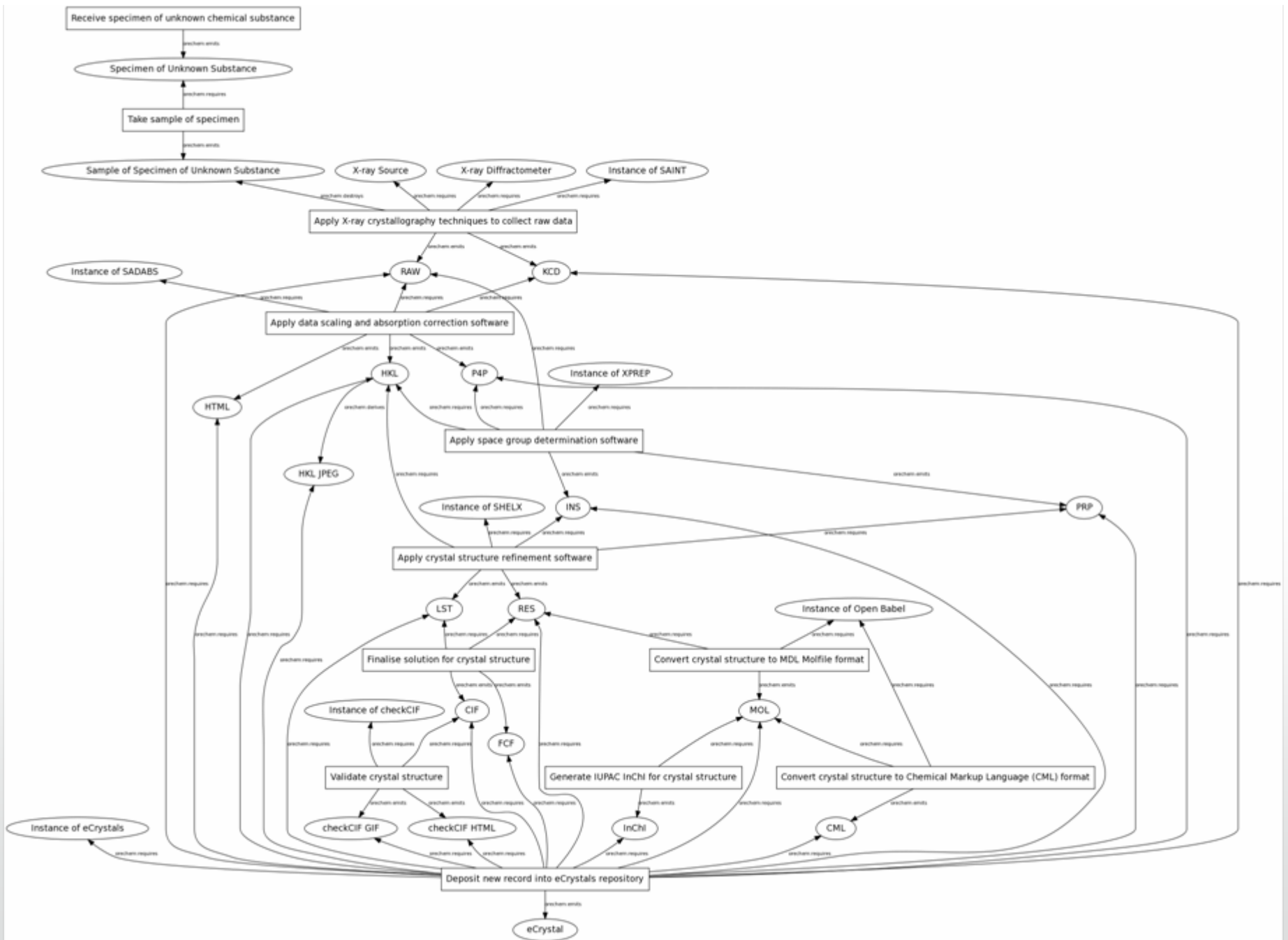
Satisfaction

- Given an enactment that has a plan, iff for all entities in the plan, there exists a corresponding entity in the enactment, then the enactment has “satisfied” the plan.
 - Satisfying the plan is equivalent to “being able to finish” the enactment.
 - This is distinct from “finishing” the enactment, e.g., the agent may repeat an action.
 - Each iteration creates a separate chain of events.

Plan for eCrystals Record

- Structure determination workflow is described by a machine-readable document
- Available online @
 - <http://ecrystals.chem.soton.ac.uk/plan.rdf>



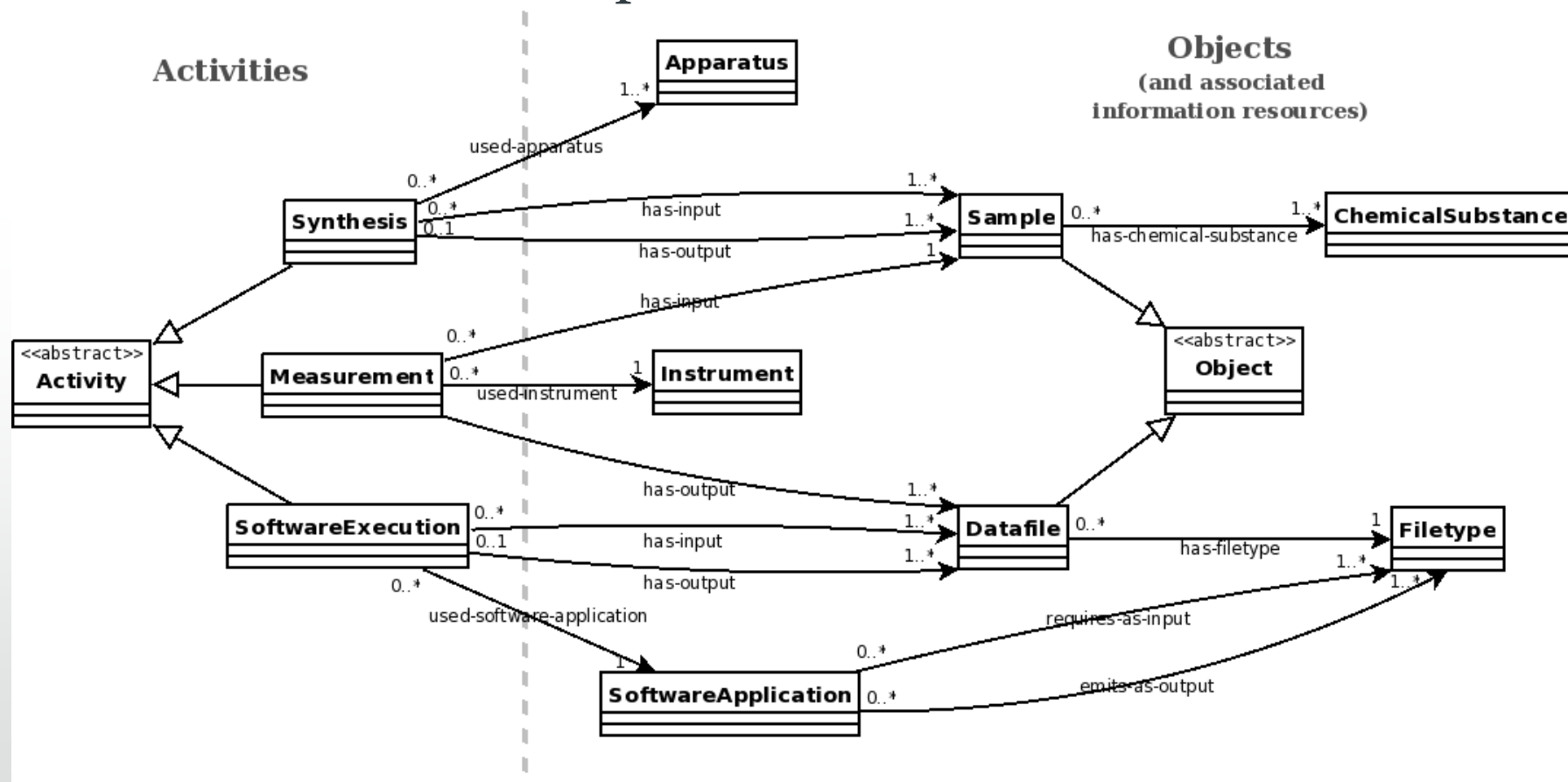


To Conclude

Bringing the models together

I2S2 Information Model (Core)

- Entities in the CSMD planned and enacted



NCS Implementation



ncs UK National Crystallography Service Management Portal

Current User: Andrew Milsted (a.j.milsted@soton.ac.uk) [Log out](#)

Home

[New Sample](#)

Search

2011NCS0014
User Code: andrew

Mr Andrew Milsted
University of Southampton

Dr Andrew Milsted
University of Southampton

Lab Status

Samples

Sample Code	Your Reference	Type	Priority	Status
2011NCS0015	Test1	Full Structure	Medium	Submitted
2011NCS0021	Hello2	Full Structure	Low	Logged In

Messages

ncs UK National Crystallography Service Management Portal

Current User: Andrew Milsted (a.j.milsted@soton.ac.uk) [Log out](#)

User: Mr Andrew Milsted

Basic Information

Name: Mr Andrew Milsted
Email: a.j.milsted@soton.ac.uk
Type: Admin
Department: National Crystallography Service
Institution: University of Southampton

Messages

Email Settings

List

Users List

Addresses

Mr Andrew Milsted
National Crystallography Service
University of Southampton
SO17 1BJ

Allocations and Samples

Allocation Details

Period: Sept 2010 (01/09/10 - 01/04/11)

No Allocation Awarded

Samples

Sample Code	Your Reference	Type	Priority	Status
2011NCS0015	Test1	Full Structure	Medium	Submitted
2011NCS0021	Hello2	Full Structure	Low	Logged In

ncs Website © NCS 2010

NCS Implementation



ncs | UK National Crystallography Service Management Portal

Current User: Andrew Misted (A.Misted@ncs.ac.uk)

Edit Sample: 2011NCS0021

Sample Sensitivity

Sensitive To: Air
 Water
 Light
 Solvent Loss
 Other

Sample Safety

Brief description of sample: Amide
In a vial, under nitrogen

Name of Solvent: No Solvent
 Acetonitrile
 Dichloromethane (DCM)
 Diethyl ether
 Dimethyl Sulfoxide
 Dimethylformamide
 Ethanol
 Hexane
 Methanol
 Trichloroethane (Trichloroform)

Is Sample Radioactive? Yes
 No
Your sample is radioactive, please complete the form BUT DO NOT send us the sample until we have made contact with you.

Is Sample Biologically Active? No
 Yes

Head Phrases:

- H205: Oxidising substance
- H226: Flammable liquid and vapour
- H228: Flammable gas
- H252: Highly flammable liquid and vapour
- H253: Highly flammable gas and vapour
- H260: Contact with water releases flammable gas
- H272: May cause an explosion if exposed to fire or heat
- H280: Contains pressurised gas, may explode when heated
- H302: Harmful if swallowed
- H314: Causes severe skin burns and eye damage
- H315: Causes skin irritation
- H317: Causes allergic skin reactions
- H330: Fatal if inhaled
- H331: Toxic if inhaled
- H332: Irritating to the respiratory tract
- H334: May cause respiratory irritation
- H335: May cause respiratory distress
- H336: May cause drowsiness or dizziness
- H337: Causes serious eye irritation
- H350: May cause genetic defects
- H360DF: May damage the foetus through the placenta
- H360FD: May damage the foetus through the placenta
- H370: May cause cancer
- H373: May cause cancer or reproductive harm
- H400: May be harmful to aquatic life
- H410: May cause long-term adverse effects on aquatic life
- H411: Toxic to aquatic life with long lasting effects
- H412: Harmful to aquatic life with long lasting effects
- H413: May cause long term harmful effects to aquatic life
- H500: Hazardous to the ozone layer

Additional Information: With ethanol

Expected Structure:

Expected Formula: C₇H₉N₂O

Experiment Details

Non-Standard Experiment: No

Details: Details for experiment

Sensitivities and Safety

Sensitive to: Air, Solvent Loss

Brief Description: Amide

Name of Solvent: Methanol, Dichloromethane (DCM)

Hazards: H225: Highly flammable liquid and vapour, H240: Heating may cause an explosion, H301: Toxic if swallowed, H311: Toxic in contact with skin, H331: Toxic if inhaled, H351: Suspected of causing cancer, H370: Causes damage to organs

Radioactive: Yes
Biologically Active: No
Other Hazards:

Sample Location: [Print Label](#)

nCS Website © NCS 2010

Thanks!

- Project partners
 - UKOLN
 - STFC
 - Cambridge
 - DCC
 - Neil Beagrie
- Developers: Andrew Milsted & Mark Borkum
- oreChem project team & Microsoft
- JISC

