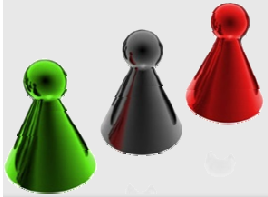
Security policies in scientific data sharing agreements

A Arenas, S Crompton, B Matthews, M Wilson, B Aziz,
Science and Technology Facilities Council

Emil Lupu, E Scalavino,
Imperial College London.

F Martinelli, M Petrocchi.
CNR Istituto di Informatica e Telematica, Pisa, Italy

SEVENTH FRAMEWORK PROGRAMME



The Issue

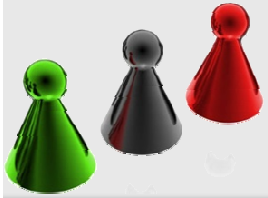
CONSEQUENCE

The on-line publication of data is encouraged by funding bodies to promote re-analysis and interdisciplinary research.

The publication of scientific data is governed by legislation and data sharing agreements between the different parties.

Agreements are drafted by senior managers and lawyers to express what can be decided in court, rather than what can be enforced computationally.

But, what can be enforced computationally … ?



Data Sharing Scenario

CONSEQUENCE

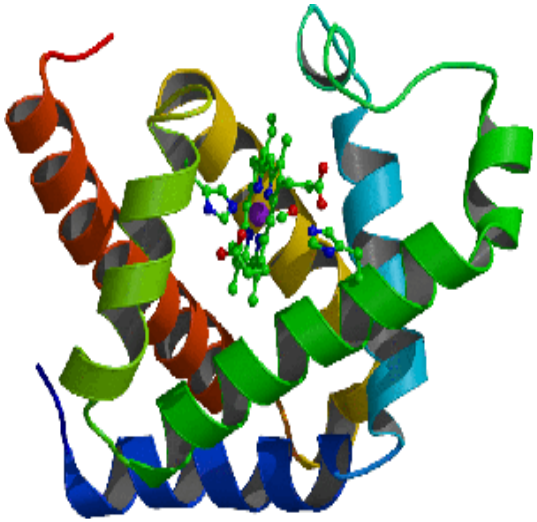
A Research Council awards a grant to a consortium to study an enzyme that is essential to the lifecycle of HIV virus.

The project will pave the way for the development of new drugs, so is co-funded by a drug company.

The consortium includes researchers from international Universities, and a commercial partner.

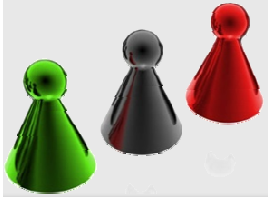
STFC Diamond synchrotron beamline is used to:

- collect X-ray diffraction patterns of an enzyme bound to the interacting host proteins
- *In vitro* experiment of docking drug compounds



STFC store the experimental data

Who has access to it and when ?

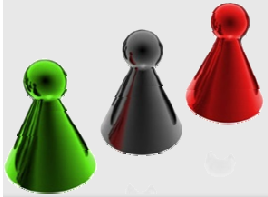


Actors

CONSEQUENCE

1. Funding bodies
2. Exploiting companies
3. STFC facilities management
4. Annually over 15000 visiting Scientists from around the world from both Academia and Industry.
5. The collaborative projects the scientists work in
6. The team at each university
7. Universities who employ scientists

Each have their own data policies



a policy use case

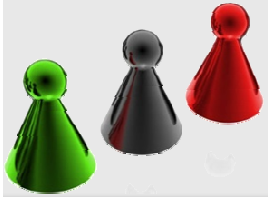
The national funding body who has paid for the research has a data policy which states that the funded researchers, staff in the funding body and their reviewers should have access to the data for 3 years, but nobody else.

The researchers work in a university who have a data access policy that all researchers in the university should retain IPR on their data and not allow others access to it for 5 years. All researchers in the university have access to the data of all other researchers in the university in order to facilitate interdisciplinary research.

The pharmaceutical company who co-sponsor the research have a policy that although others can have access to the data, they are the only ones who can use the data for commercial purposes.

One researcher on the project is submitting part of the work to her university to acquire a PhD, and does not want anybody else, even in the university, to see it for 3 years.

STFC has a policy that our staff can have access to the data produced on our facilities for administration and for use in developing the facilities.



DSA – Example Policies

Funding Agency → Grantees

- Research data with no outstanding IP must be made publicly available after project completion or on publication, whichever the earlier.
- Its reviewers have read only access to data during project lifetime.

Facility Provider → Facility Users

- All results from non-commercial experiments older than 3 years will be made public unless an extension has been granted by the Director.
- Appropriate personnel (instrument scientists, administrators) can access facility user's data or metadata for facility-related purposes (eg. maintenance, beamline development). All relevant staff are bound by standard non-disclosure restrictions.

Participating Organisation (academic) → Employees

- In-house colleagues have read access to other researchers' data to facilitate cross-discipline research.
- External researchers are barred from accessing data arising from employee's research for a period of 5 years.
- Retains IP arising from staff's research.

Participating Organisation (Commercial) →Consortium Partners

- Access to its proprietary libraries and data derived from its application are restricted to
  - the EU
  - named investigators
  - the project lifetime
- Retains share of IP on data and patent/s derived from the use of its library.
- Has exclusive licence to take targeted active compounds into drug development.

Consortium Member (Ph.D. Student) → The World

- Nobody can access his research data until after the completion of the project which is tied to the submission deadline of his thesis.

Example Policies in Data Sharing Agreements

5 DATA SHARING AND ACCESS

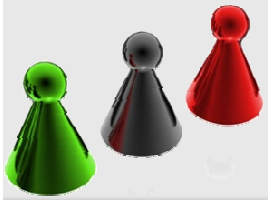
5.1 The parties agree in principle to the open and free sharing between the parties of meteorological, air, water and terrestrial data gathered from monitoring locations;

5.7 The parties agree to seek opportunities to provide the public with seamless Web access to federal and provincial environmental data and information;

5.8 Data or information acquired by one party from the other under the Agreement shall not be disseminated commercially or otherwise disclosed, transmitted or sold to any organizations other than the parties to the Agreement without the written consent of the party that owns the information. The parties thus recognize that the organization that produces a particular set of data or information item is the only one, at its option, entitled to receive payment for subsequent distribution or use;

- 5.9 Quality controlled data or information acquired by one party from the other under the Agreement may be disseminated or otherwise disclosed or transmitted without cost to any individual or organization without the need for written consent from the party that owns the information providing that (1) the data or information are not modified, (2) the jurisdictional ownership is clearly acknowledged and (3) the release does not compromise the privacy or commercial competitiveness of the data source;
- 5.10 Raw data or draft information acquired by one party from the other under the Agreement may not be disseminated or otherwise disclosed to any individual or organization other than the parties to the Agreement without the written consent of the party that owns the data or produced the information;
- 5.12 The Agreement authorizes the production, dissemination and sale of derivatives by the parties. Where appropriate (e.g., in scientific reports and articles), the party that owns the data shall be explicitly acknowledged as the owner of the data used;
- 5.15 The parties agree to acquire and use data and information in a manner that respects the scientific value of the data, legislation and intellectual property rights of the parties;

Environmental Data Sharing AGREEMENT Between Environment Canada, Pacific and Yukon Region and British Columbia Ministry of Water Land and Air Protection, April 29, 2004



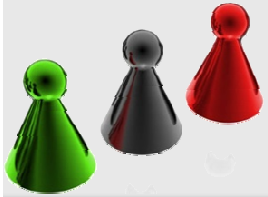
US Restrictive data policies

VI. LOCATION OF MATCHED DATA AND CUSTODIAL RESPONSIBILITY
This agreement represents and warrants further that, except as specified in an attachment or except as authorized in writing, that such data shall not be disclosed, released, revealed, showed, sold, rented, leased, loaned, or otherwise have access granted to the data covered by this agreement to any person. Access to the data covered by this agreement shall be limited to the minimum number of individuals necessary to achieve the purpose stated in this section and to those individuals on a need-to-know basis only.

VII. CONFIDENTIALITY
The User agrees to establish appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to it. See - OMB Circular No. A-130, Appendix III

VIII. DISPOSITION OF DATA
The requestor and its agents will destroy all confidential information associated with actual records as soon as the purposes of the project have been accomplished and notify the providing agency to this effect in writing.

Inter-Agency Data-Sharing Agreement, Centers for Medicaid and Medicare Services, http://www.cms.hhs.gov/states/letters/smd10228.asp



UK restrictive data policies

CONSEQUENCE

3 Nomination of Staff

State team members working within the project
To whom data shall be sent
To whom disclosures should be made
Who are responsible for Data Protection and Security
Requests from non-authorised personnel / organisations shall be declined.

4 Commitment to compliance

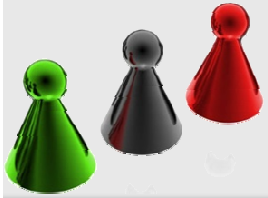
[Make reference to Policies and Procedures and State that all staff will be made aware of their responsibilities under the Data Protection Act 1998 and related legislation]

5 Security Measures

[Identify specific Security requirements to allow for the secure transfer of data]
e.g.
[Passwords]
[encryption]
[Personnel security]
[Classification guide – allocate level of security to specific sets of data]
[Data Storage – compatible with Classification guide]

6 Destruction of information

[Process]
[Retention Periods for sets of data]
[Define how data will be destroyed after Retention period]
[Make reference to Policies already in place where possible]



Analysing Data Sharing Policies

CONSEQUENCE

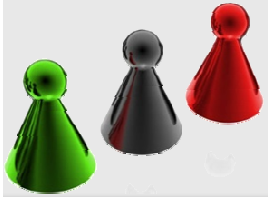
Policies need:

1) to be refined down to policy languages where they can be analysed and conflicts resolved,
2) further refined to languages which can be enforced at run time.

Informally: If B receives object o from A on c1, then B will not send o to any other principal for 1 year

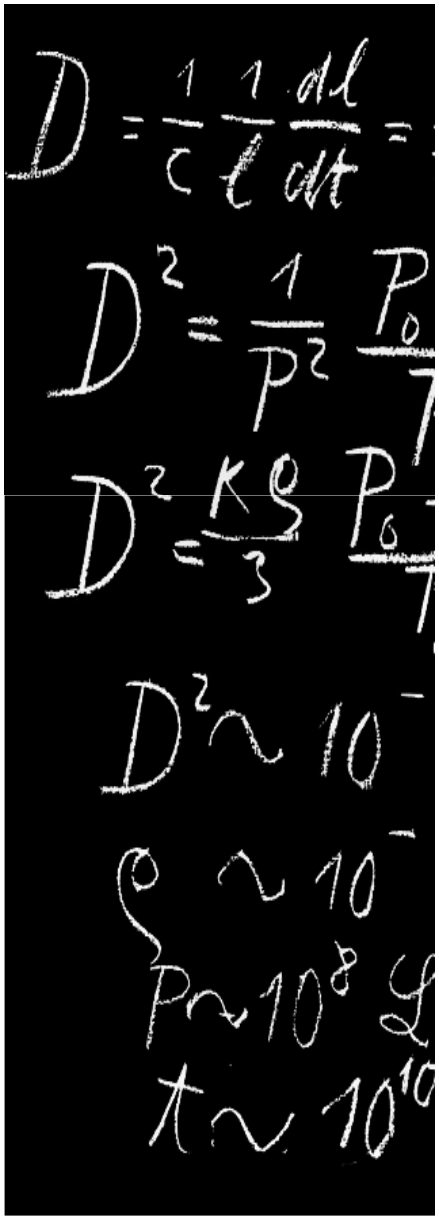
Formally:

(forall T o) if c1.rcv(o) then
(forall Channel c) if (c.dest ≠ A) then
(not c.send(o) until 1 year)
at B until 17/04/2009



Enforcing Policies

CONSEQUENCE



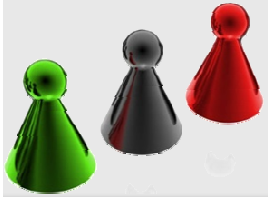
The established scientific approach is to apply access controls on the data server

- data usage is not addressed.

The established commercial approach is to apply digital rights management which constrains the use of the data in the reader or player programme

- these technologies are coarse and simplistic – play/no-play

Need to combine these two approaches, will also allow control over both data access and data usage



The Consortium

CONSEQUENCE

High demand testbeds

BAE SYSTEMS

Science & Technology Facilities Council

Industrial innovators

Microsoft Innovation Center Europe

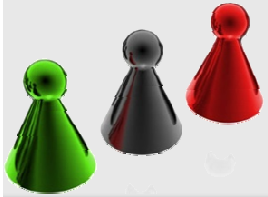
hp invent

Researchers

Imperial College London

IIT Consiglio Nazionale delle Ricerche Istituto di Informatica e Telematica

CREATE-NET



Main objectives

Define an *architecture* within a *framework*

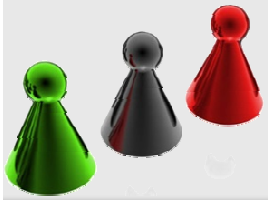
- to enable *dynamic management policies*
- based on *agreements* that
- ensure *end-to-end secure protection*
- of *data-centric information*.

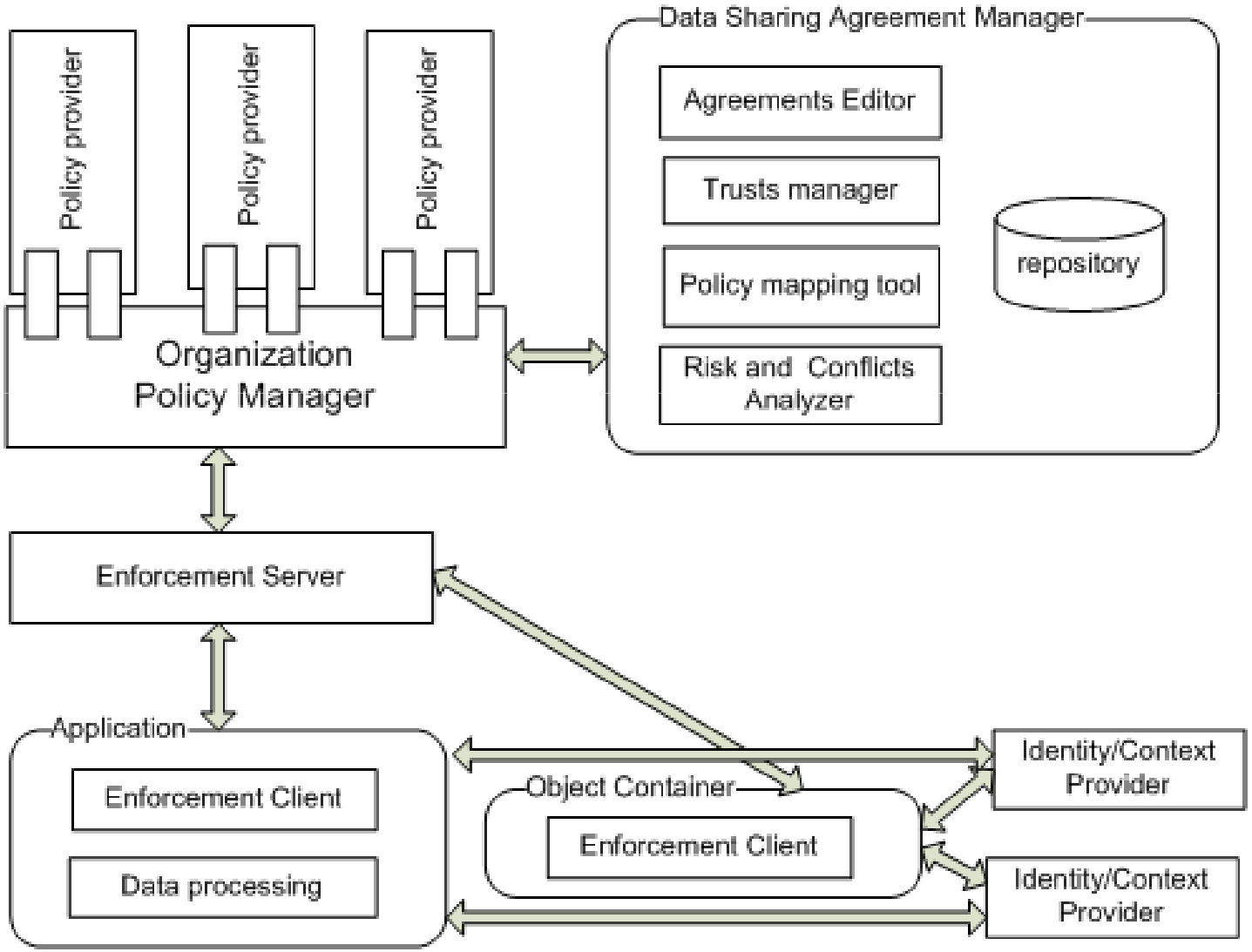
Implement the architecture in *software*.

Evaluate the *technical* and *business benefits* of the implementation and framework via *two test beds*:

- *Sensitive scientific data*
- *Crisis management data*

Policy Management & Enforcement Architecture

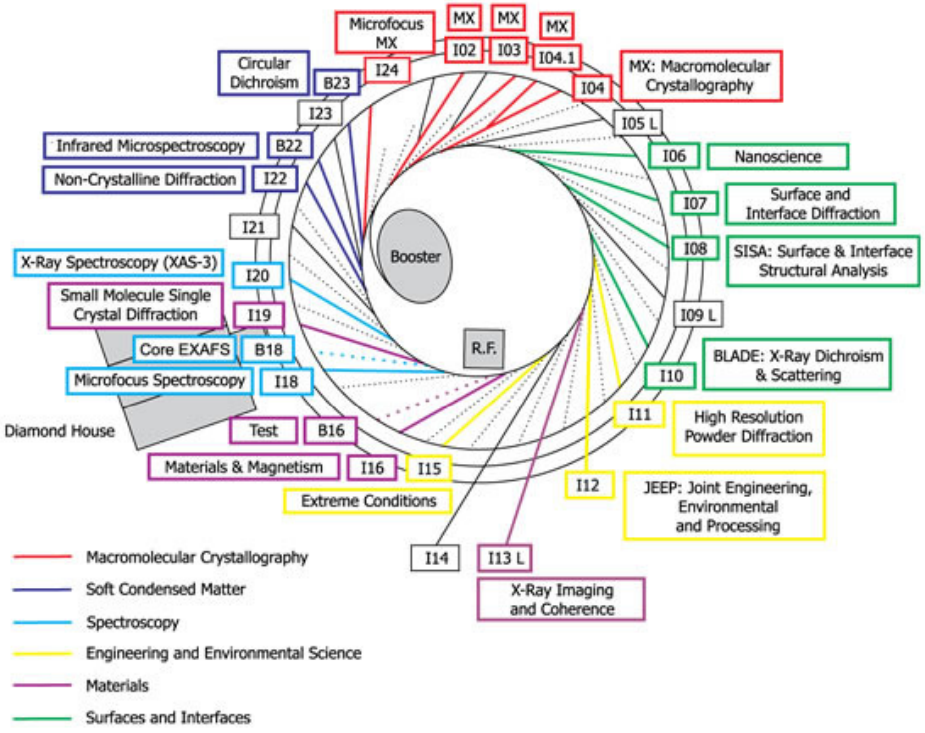




Science & Technology Facilities Council

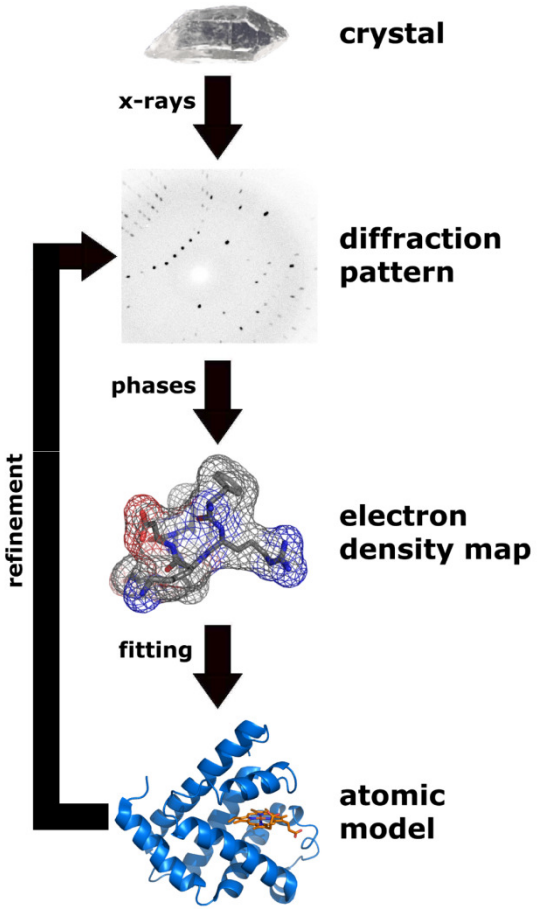
e-Science

Diamond Light Source



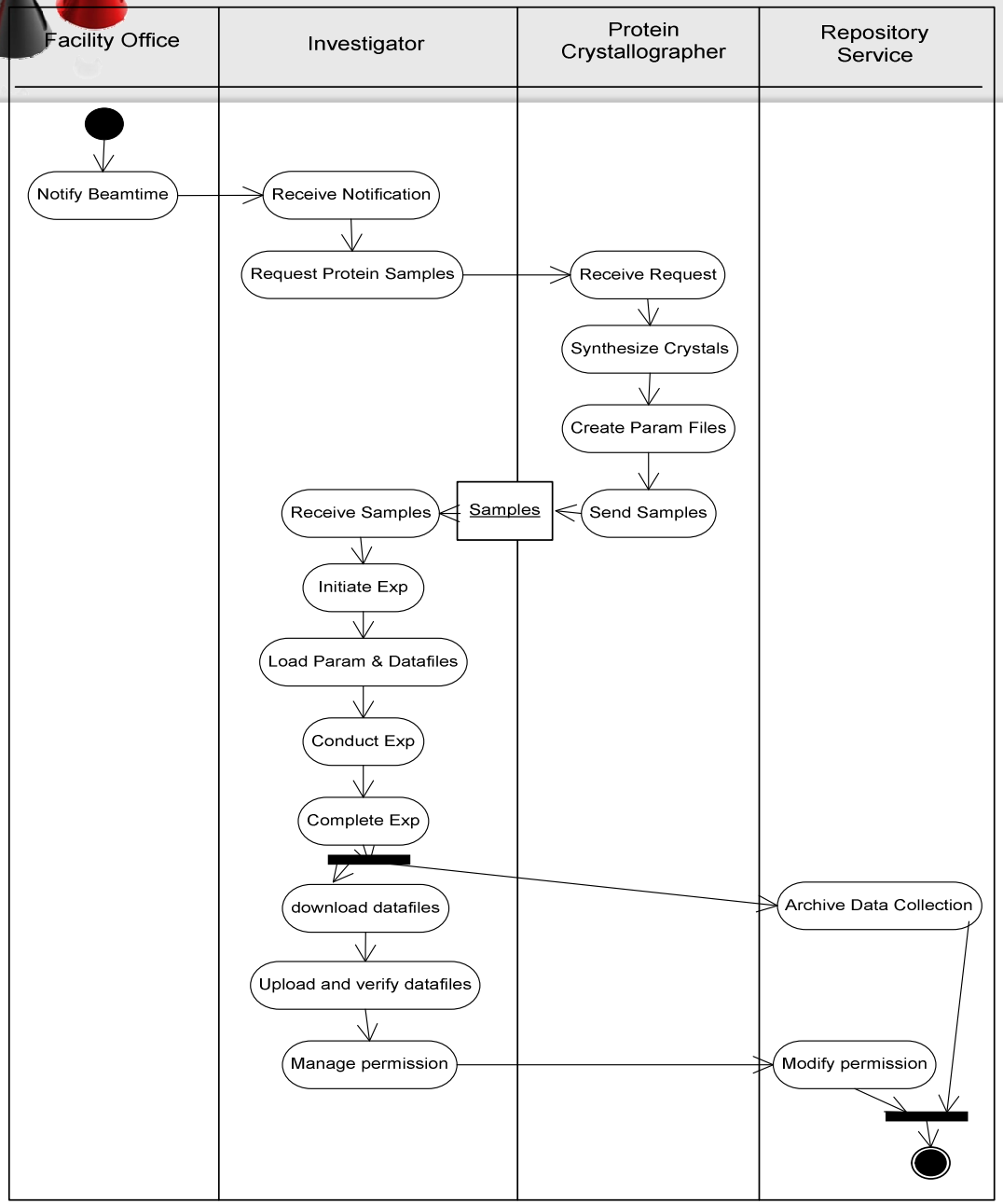
- £175 million facility
- Electrons are accelerated in a synchrotron ring
- Electrons are passed through magnets
- Light and x rays are piped to experimental stations

DLS experiments



- X-rays are 1 billion times brighter than hospital x-rays
- X rays are focussed on materials
- Detectors collect diffracted x-rays
- Computers re-construct the materials structure

Mini-scenario 1 : Beamline Experiment Workflow



Mini-scenario 2 – Concurrent Research Activities