



*National Institute for  
Health Research*

# **NIHR Information Systems**

## **Enterprise Architecture**

Document Number: NIHR 4.2EA001  
Version 1.0  
Issue Date: 01/10/2007

# Executive Summary

This document is an introduction to the NIHR Enterprise Architecture. It provides a formal mechanism for aligning business requirements and information systems so that IT systems support the goals of NIHR.

The Architecture is composed of three parts – the business architecture which describes organisational goals and processes; the information architecture which describes the information used to support these processes; and the technology architecture, which describes the underlying computing technologies which should be used to support NIHR.

## **Business Architecture**

The Architecture is intended to cover all the processes and operations associated with the NIHR. These can be described broadly as the management of NHS research expenditure in England.

Management of research takes place across many agencies and at many levels in the NHS. The Architecture is of relevance to all of these.

The Architecture does not currently apply to the collection and management of research data, i.e. the data collected as part of a research study.

## **Information architecture**

The architecture emphasises the need for strategic management of information across NIHR. Four classes of information are emphasised throughout and form the basis for the information architecture. These are: people, research projects, facilities, and finance.

These classes will form the basis for a number of services and applications. For example, the person class is the basis for and NIHR Identity Management service. The research project class forms the basis for a lifelong research record.

Data classes and their attributes are defined within a data dictionary, to be based on the NHS Data Dictionary with extensions as necessary. The WHO clinical trial data set forms the core of the attributes in the Research Project Class.

Particular emphasis is placed on unique identifiers. Digital Object Identifiers and OpenIDs are currently being considered.

Integration and interoperability will be achieved through web services. These need not be complex. For example the Programme is not certain that an Enterprise Service Bus is needed. It may be that simple web services such as RSS and REST provide much of the service layer.

Security standards underlie a range of services which ensure the integrity, confidentiality and appropriate use of information. Services include event monitoring, virus checking, encryption, firewalls, and access control. NIHR Information Systems will comply with all appropriate standards including ISO 27001. Detailed planning for standards compliance will be completed in due course.

## **Technology Architecture**

The proposed technology architecture is intended to support the business and information architectures and of note has the following features:

Internet enabled – it follows that the underlying network architecture will be the Internet, including N3. Human Computer Interfaces will be delivered mainly through web browsers, which should be at least WAI AA compliant. Some applications and services may also be delivered through mobile devices

Loose-coupling – a base assumption is that the overall system will be composed from a series of semi-independent software applications which nonetheless have the capability of appearing to the user to be a single system

Service orientation – the fundamental way in which loose-coupling will happen is through the exposure and consumption of services

Software as a Service – defined as a software application delivery model where a software vendor develops a web-native software application and hosts and operates (either independently or through a third-party) the application for use by its customers over the Internet.

Transition to the new Architecture is intended to take 3-5 years. It involves:

1. Implementing key system components, which are defined within the R&D MIS and portal. These include: personal home page (part of the NIHR portal) and lifelong research record (part of the R&D MIS)
2. Implementing a range of shared services within R&D MIS and portal, such as identity management and a contacts directory
3. Taking a strategic approach to data management for each main class, ensuring accuracy and completeness. This will require greater focus on data management than is presently the case.
4. Establishing data standards including a data dictionary for data that is to be shared across applications and APIs to enable interoperability
5. Influencing stakeholders and existing projects to work in an orchestrated way, by establishing governance for the Architecture and through partnership working with National Co-ordinating Centres, networks, regulatory agencies, and others.

## Table of contents

Executive Summary .....	2
Purpose and structure.....	5
Introduction .....	6
Overview .....	9
Developing the Enterprise Architecture.....	15
Research project class/entity .....	20
Person class/entity .....	23
Finance class/entity .....	26
Facility class/entity .....	27
Implementation .....	29
Data Dictionary .....	32

# Purpose and structure

The NIHR Enterprise Architecture provides a blueprint for the design of information systems to support the management of health research in the NHS in England.

This document is an introduction and is organised as follows:

Section 1 – the **Introduction** describes the policy and operational context for which the architecture is being developed.

Section 2 - the **Overview** provides a summary of the architecture.

Section 3 – **Developing the Enterprise Architecture** – adds detail to the architecture by developing the business, information and technical architectures of the fundamental entities of the NIHR information system.

Section 4 – the **Implementation roadmap** – sets out the approach for implementing the Enterprise Architecture

Section 5 – the **NIHR Data Dictionary** – sets out an initial data dictionary, which will develop quickly over the next few months.

# Introduction

The NHS spends around £700M per annum on research, including: the direct funding of research studies, programmes and centres; and the support costs of studies taking place in the NHS but funded by other agencies.

Health research is inseparable from health care. It enables the development of better treatments and new technologies; it provides evidence about the effectiveness of existing interventions. It also offers hope for cures where there is none, and it is a major source of industry and enterprise for the UK economy.

## Policy framework

Current strategy for supporting research in the NHS is contained in *Best Research for Best Health*<sup>1</sup>, the Government's five year strategy for the NHS contribution to Health Research. The vision for the programme is to:

- Support the Science & Innovation Framework 2004-14
- Improve health outcomes through research
- Harness the capacity of the NHS
- Tackle R&D bureaucracy
- Stop loss of industry trials from the NHS

Central to the realisation of this vision is the establishment of a new body, the National Institute for Health Research (NIHR), whose role is described below. The subsequent review of health research by HM Treasury (Cooksey Review) of December 2006<sup>2</sup> confirmed the vision of Best Research for Best Health and made a number of further recommendations:

- The creation of an Office for Strategic Coordination of Health Research (OSCHR) which will oversee all of government expenditure on medical research, including NHS and MRC budgets.
- Changes to MRC board membership and clarification of responsibilities of MRC and NIHR
- Establishment of NIHR as a real, rather than virtual, Institute from 2009
- Greater focus on translational research (ie the translation of scientific discoveries into beneficial applications)

---

<sup>1</sup> <http://tinyurl.com/br6tz>

<sup>2</sup> [http://www.hm-treasury.gov.uk/independent\\_reviews/cooksey\\_review/cookseyreview\\_index.cfm](http://www.hm-treasury.gov.uk/independent_reviews/cooksey_review/cookseyreview_index.cfm)

## The National Institute for Health Research

*Best Research for Best Health* is being taken forward through a series of implementation plans<sup>3</sup> Of central importance to these is the creation of the National Institute for Health Research (NIHR) Originally intended as a 'virtual organisation', the Cooksey Review recommended its establishment as a real institute from 2009. As a programme, the NIHR will be responsible for several outcomes in support of Best Research for Best Health:

- Establishing the NHS as an internationally recognised centre of research excellence
- Attracting, develop and retain the best research professionals to conduct people-based research
- Commissioning research focused on improving health and social care
- Managing knowledge resources
- Acting as a sound custodian of public money

## Business requirements

From these outcomes a number of new business requirements for the NIHR can be identified. These include:

- Establish the identity of the NIHR nationally and internationally
- Engage and connect with the public
- Increase recruitment into clinical trials
- Engage and connect with researchers individually and in networks
- Simplify and speed-up ethics approval
- Ensure transparency and equity of resource allocation
- Facilitate the improvement of health outcomes through the outputs from NIHR

These are essentially about the management of research, as opposed to the conduct of research. Changes to the management of research – through reorganising, streamlining and coordinating operational processes – will help meet the business requirements of NIHR. But given the degree to which research management is a knowledge-based activity it is clear that process change must be intensified and supported by change to information systems.

The current NHS R&D information system landscape can be characterised as fragmented and piecemeal. It is not surprising that this is so. The many agencies and bodies involved in research each have their own information systems, almost none of which can share data because there has been no imperative to share data and no standards for doing so.

## NIHR Information Systems Programme

Recognising the need for change in information systems, the NIHR implementation programme has established an Information System programme. The purpose of the programme is to create, with partners where

---

<sup>3</sup> <http://tinyurl.com/zppxa>

appropriate, an integrated set of national systems for health research and research management that will:

- unify and simplify the administrative procedures associated with regulation, governance, reporting and research administration;
- enable procedures and input of data to occur once and once only;
- make information supporting regulatory approvals and permissions available via secure systems and processes to everyone who needs to act on it;
- increase the UK's attractiveness as a site for international clinical research.

To achieve these goals the IS Programme will introduce two new systems (a portal and a management information system) and define methods and standards for exchanging information between NIHR information systems.<sup>4</sup>

The NIHR Enterprise Architecture sets out these methods and standards.<sup>5</sup> It focuses on how existing and new information systems can work together to support the strategic goals and operational needs of NIHR over the next few years. The next section provides an overview of the Enterprise Architecture.

---

<sup>4</sup> This is a summary of the IS Programme For a full description of its activities see [http://www.nihr.ac.uk/files/pdfs/Implementation\\_Plan\\_4.2\\_Bureaucracy\\_Busting\\_Research\\_Information\\_Systems\\_August\\_2007.pdf](http://www.nihr.ac.uk/files/pdfs/Implementation_Plan_4.2_Bureaucracy_Busting_Research_Information_Systems_August_2007.pdf)

<sup>5</sup> [http://en.wikipedia.org/wiki/Enterprise\\_architecture](http://en.wikipedia.org/wiki/Enterprise_architecture)

# Overview

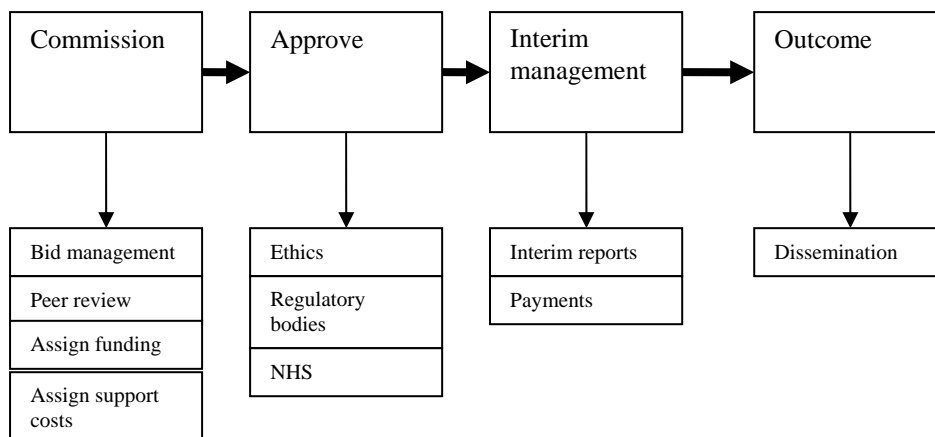
## Research processes

Research is a highly knowledge-based activity. At all stages – from conception to publication – it involves capturing, managing and exchanging information relating to:

- background research using databases, websites and journals
- writing grant applications
- collaboration with co-researchers
- processing ethics approvals
- grant application processing
- research data collection
- research project management
- statistical analysis
- writing up results
- peer review
- publishing
- translating findings into new systems and services
- research assessment exercises
- priority setting

## Research management

Research is managed by a number of agencies, which are responsible for the commissioning, approval and other management of research. The processes associated with management are outlined below:



Many information systems are involved in managing research. Of note:

- researchers say they find research management processes confusing and bureaucratic
- the various management information systems associated with these processes were not designed to work together in an orchestrated fashion, creating inefficiencies and duplication in data management

The NIHR implementation programme is addressing these concerns by rationalising the approvals process and establishing an information systems programme, described in Implementation Plan 4.2<sup>6</sup>

Research management processes takes place in many places and at many levels in the NHS, including nationally, within research networks, and within NHS Trusts. The NIHR Enterprise Architecture is intended to apply to research management at all levels in the NHS.

### **The role and benefits of the NIHR Information Systems Programme**

The IS Programme is intended to supplement and integrate existing information systems them by introducing:

- From 2008 a common portal to provide access to all systems and new functionality
- A management information system to integrate and enhance existing systems
- An Enterprise Architecture to describe how new and existing systems can integrate and interoperate

In delivering the outputs described above, the IS Programme will provide the information systems support to allow other Programmes of work to deliver the far-reaching objectives set for NIHR. As such, there are no directly measurable benefits of the IS Programme. Rather, the benefit measure in each area of output will be whether the output does indeed provide sufficient ICT support to allow others outside this Programme to deliver the objectives.

For the Architecture, the intended benefits are to

1. Increase the capacity of NIHR information systems to integrate and interoperate
2. Ensure that information systems are responsive to business need
3. Highlight the need to manage information as a strategic resource across NIHR

<sup>6</sup> [http://www.nihr.ac.uk/about\\_implementation\\_plans.aspx](http://www.nihr.ac.uk/about_implementation_plans.aspx)

## Fundamental information classes/entities

While research processes and information flows are perceived as complex, the fundamental classes of information or entities involved in the research cycle are simple, few in number and easy to describe. They are:

**Research Projects:** research projects are the trials, studies and other scientific studies which form the core activity of the NIHR

**People:** at all stages of the research cycle people are of fundamental importance – as research investigators, managers, reviewers, and Faculty members.

**Facilities:** the places where research takes place and the organisations which sponsor and fund research.

**Finance:** the financial resource used to support research, directly or indirectly.

These classes/entities may appear self-evident. But if they are managed effectively the strategic goals and the five-year blueprint of BRfBH will be met. For example, the strategic goal:

‘Attract, develop and retain the best research professionals to conduct people-based research’

requires the NIHR to build systems and services for people who are or who will be significant NHS based researchers. These systems and services - for example the NIHR Faculty - will rely on applications and services which build trust between individuals and the NIHR.

Trust is a matter of predictable, coherent, and well targeted communications. In turn this requires an underlying database of people, and their metadata, which might run to 100 or more items about a person as systems develop. The people database needs to offer identity management services across the NIHR, so that people don't have to have separate logins for the NIHR systems they use.

The database needs to be available to all the systems in the NIHR world that use or create information about people. And it goes without saying that the information has to be accurate and as complete as it needs to be.

These services are not trivial to create and manage. But it is important to recognise that the core information requirement for at least one of the strategic goals of BRfBH is essentially a database of demographic information.

Research Project, People, Money and Facilities are the fundamental classes of information in the NIHR system. Actual systems are composed from the processes that link instances of these classes. Figure Y illustrates some of the

instances of each class of entity. It is a premise of SOA that new relationships will emerge over time so Figure 1 is not intended to be exhaustive.

**Figure 1**

NIHR: fundamental system entities (not exhaustive)				
	Research projects	People	Money	Facilities
	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Studies funded through CCF, SDO, HTA, PRP</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Studies funded by pharma and other commercial orgs and taking place in the NHS</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Ditto studies funded by charities</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Programmes of work funded by NIHR</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Other studies which receive support funding</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">People leading research projects</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">People managing research</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Members of the public interested in research or becoming a research subject</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">People involved in the governance of research</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">People who are key stakeholders or on committees</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">People who undertake peer-review of research proposals</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">NIHR funds used to meet the direct costs of research</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">NIHR funds used to support the indirect costs of research</div>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Organisations which have a stakeholder/partner interest in NIHR</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Organisations which have a governance interest in NIHR</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">All NHS organisations</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">All facilities in the NHS where research actually takes place</div>

### The NIHR Enterprise Architecture

The NIHR Enterprise Architecture (NIHR EA) provides a blueprint for the development of information systems to support the aims of NIHR. Whilst it has many similarities to the Enterprise Architecture of the National Institutes of Health,<sup>7</sup> it is not intended (at this stage anyway) as a detailed general purpose description of all the systems in NIHR. Instead, it is focussed on:

<sup>7</sup> <http://enterprisearchitecture.nih.gov>

- integration and interoperability between new and existing systems
- ensuring that the development of information systems supports the goals of NIHR

It does this by identifying four fundamental classes of information and on the basis of these describing:

- the business processes with which each is associated
- the key system components associated with each
- the data attributes of each
- the services which each supports
- the underlying technologies supporting each

Other points of note include:

Cross-linkage to other architectures. The NIHR EA sits alongside, and is compatible with, two closely related architectures:

- HM Government cross government enterprise architecture (XGEA), which is being developed as a 'meta-architecture' for UK public sector information systems<sup>8</sup>
- NIH Enterprise Architecture, which informs the development of NIH information systems<sup>9</sup>

Lifelong research record Of central importance to the architecture is the concept of a single record for each research project supported by NIHR. Together these will form part of the UKCRN portfolio and provide the basis for orchestrating a range of services to researchers and research managers

Continuous improvement and refinement The architecture will evolve in line with business requirements. At this stage the NIHR EA is not intended to be a fully detailed specification for systems development. It is focussed on methods for achieving interoperability, and for setting out a framework for bringing together business goals and information systems. The Architecture will get more detailed over time. The data dictionary will become richer as requirements for definition arise; business and technology descriptions will get more detailed through projects such as the CSOU.

Scope To fulfil its purpose the NIHR EA needs to be relevant to all business processes and all information systems in NIHR and other bodies associated with the management of research that takes place in the NHS in England.

Service Orientation The NIHR EA is Service Oriented – that is, it describes a style of integration based on independent systems providing each other with services. Service Oriented Architecture is described more fully in the next section.

---

<sup>8</sup> [http://www.cio.gov.uk/documents/cto/pdf/enterprise\\_architecture\\_uk.pdf](http://www.cio.gov.uk/documents/cto/pdf/enterprise_architecture_uk.pdf)

<sup>9</sup> <http://enterprisearchitecture.nih.gov/>

Implementation roadmap The NIHR EA sets out a vision for the re-configuration of systems (sometimes referred to as the target architecture) but is not itself the vehicle for building new systems and transforming existing ones. The EA is therefore intended to have the closest relationship with projects which develop information systems.

The transition to the target architecture is gradual, and consists of:

1. Implementing key system components. These include: personal home page (part of the NIHR portal) and lifelong research record (part of the R&D MIS)
2. Implementing a range of shared services such as identity management and a contacts directory
3. Taking a strategic approach to data management for each main class, ensuring accuracy and completeness. This will require greater focus on data management than is presently the case.
4. Establishing data standards including a data dictionary for data that is to be shared across applications and APIs to enable interoperability
5. Influencing stakeholders and existing projects to work in an orchestrated way, by establishing governance for the Architecture and through partnership working.

# Developing the Enterprise Architecture

## Composition and use

In keeping with well established approaches<sup>10</sup>, the NIHR Enterprise Architecture is composed of three sub-architectures:

- The business architecture describes the ‘business’ of NIHR, emphasising strategic goals, operational processes and the information flows that support these
- The information architecture provides detail about the data in the system, the applications which manage it, and the methods for exchanging data between applications, including data dictionary and services
- The technology architecture describes the technologies used - computers, networks, and fundamental software choices.

Each sub-architecture is applied to the fundamental entities (projects, people, facilities and finance). In practice the architecture will consist of a series of documents of varying detail, coverage and point of view.<sup>11</sup>

The present document is a general introduction to the NIHR Enterprise Architecture, including first versions of the data dictionary and implementation road-map.

As the Architecture develops it will enable:

- Expression of the NIHR information systems and their evolution
- Communication among system stakeholders
- Evaluation and comparison of existing architectures in a consistent manner
- Planning, managing, and executing the activities of system development
- Expression of the persistent characteristics and supporting principles of a system to guide acceptable change
- Verification of a system implementation’s compliance with an architectural description

---

<sup>10</sup> Standard Architecture frameworks include TOGAF <http://www.opengroup.org/>, the Federal Enterprise Architecture Reference Model <http://www.whitehouse.gov/omb/egov/a-2-EAModelsNEW2.html>, and the Zachman model <http://www.zifa.com/>

<sup>11</sup> The NIHR EA conforms to the principles in the IEEE Standard 1471-2000 ‘Recommended practice for architectural descriptions of software intensive systems’

## An integrated approach

As outlined in the previous section, the purpose of the NIHR Enterprise Architecture is to guide the development of information systems to support the aims of the NIHR.

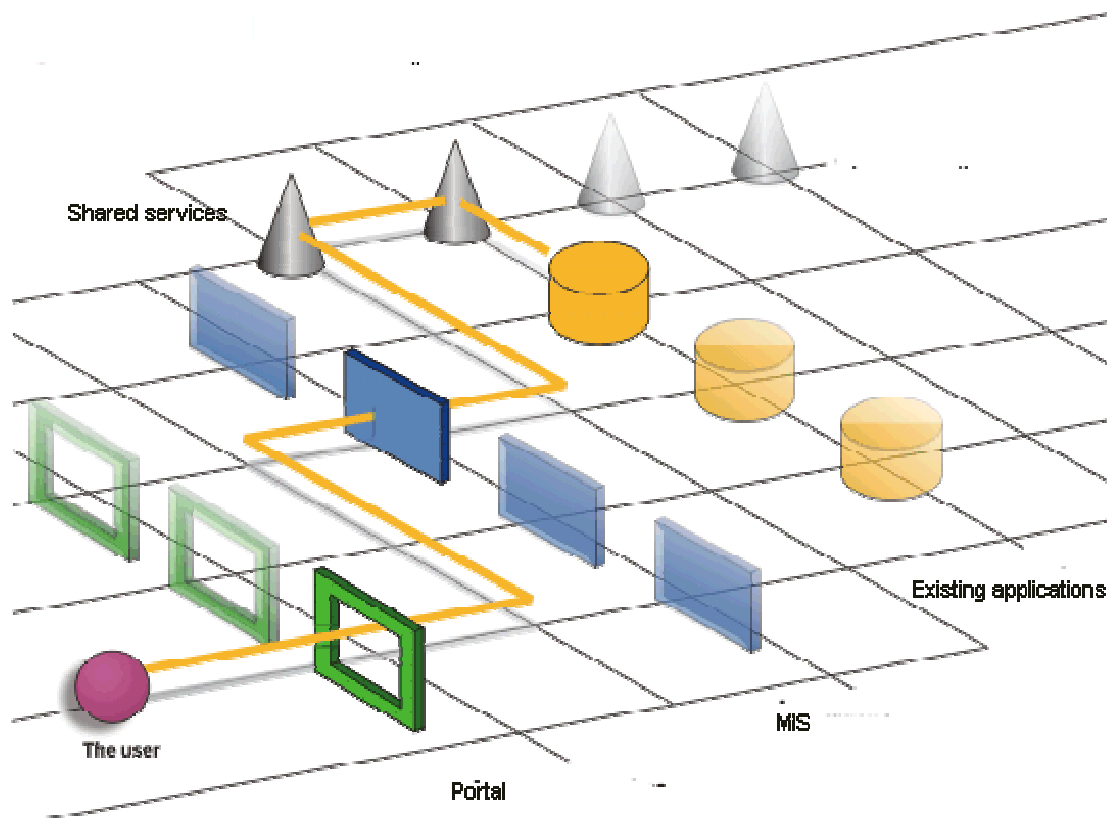
It does this by setting out methods for orchestrating information systems to meet the aims of NIHR. These methods are intended to facilitate the integration and inter-operation of information systems and are based on the following:

1. New systems components - for example a portal and a management information system as described in the NIHR programme implementation plan 4.2
2. Data standards to enable interoperability between existing systems
3. Shared services, for example an identity management service.

The overall effect of these methods is illustrated in Figure 2, which shows in outline a typical user session once the Enterprise Architecture has been implemented. Figure 2 is consistent with, and complements the systems diagram in Appendix 1.

Figure 2 shows a user logging-on via the portal. This gives them access to a personalised version of the MIS, based on a profile delivered by an identity management service and composed from the underlying NIHR (and possibly other) applications.

**Figure 2**



The composition of personalised portal and its patterns of use will vary from user to user. For example, if the user is a Faculty member the personalised version of the MIS might include a list of current studies in the portfolio, a CV, and access to grant application systems. If the user were a research manager in the CSOU or CLRN the personalised version of the MIS might include a comprehensive contacts directory, topic and location specific lists of projects, and maps of facilities in their area(s).

### **The fundamental role of services**

Many of the detailed processes within NIHR have yet to be specified and are the responsibility of work streams out-with the IS programme.

The long term architectural goal is to support developing and changing business processes by creating an environment which can respond flexibly and quickly to emerging business needs. Services are central to this goal; the following example shows how real world processes can be supported by services:

Scenario: a research manager in a network needs to hold a brief conference with several colleagues to discuss the adoption of a cross-topic research proposal.

Actions taken: She logs in to her personal workspace through the NIHR portal and sees that 3 out of 5 of the colleagues she wants to talk to are currently also logged in to the portal. She pings these colleagues and asks for a brief conference in 5 minutes. All 3 indicate their availability. In the next few minutes she adds the research proposal into a shared workspace along with a proposed solution on the whiteboard. Her colleagues come into the workspace and the decision is made to open a VOIP voice conference. The proposal is discussed and it is decided to pass the proposal to one TCRN for review. As a closing action the manager sets up a notification service so that when the TCRN review outcome is added to the research record she and selected colleagues are notified.

Services: This set of management activities is supported by a number of services built around the person and project entities, including:

- Person services –indication of presence; scheduling; workspace; VOIP; whiteboard
- Research record services – available as a digital object; can provide RSS feeds

### **Further examples of services**

Services are in effect the uses and demands that can be placed on the system and form the basis for the composition and orchestration of services across NIHR. Illustrative examples of services are:

- “show me all the research on diabetes supported by NIHR”

- “send an e-mail to all members of the XYZ committee”
- “identify people who are willing to undertake peer-review on proposals in the area of heart failure”
- “create a financial schedule for project X”
- “create a batch of payments for all projects with a payment due”
- “invite all Faculty members to meeting ABC”
- “submit my project to NRES for review”
- “allocate a Research Passport to Professor X”

## **Service Oriented Architecture**

The Architecture employs a style known as Service Oriented Architecture (SOA), which offers a distinct strategy for connecting systems and the organisations/agencies which host them. OASIS<sup>12</sup> defines SOA as:

*A paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations.*<sup>13</sup>

## **Characteristics and benefits of SOA**

SOA has the following characteristics and benefits of relevance to NIHR:

### **Business perspective**

It recognises the boundaries and interdependencies between the partners (existing systems) in the system

It encourages collaboration between partners

It also supports flexibility and change in business processes

Supports controlled interoperability with agencies outside the overall system

### **Information perspective**

Identifies the key information components/entities in the overall system

It promotes a focus on data quality in these key entities

It is flexible in its approach to the ways in which data is exchanged between partners in the system, including the exchange of ‘documents’ in XML format

### **Technology perspective**

It uses Internet standards and therefore makes integration and interoperability easy

It is geared towards the delivery of software interfaces and functions through web browsers, providing a familiar user experience

It makes maximum use of existing investments by layering services on top of them

---

<sup>12</sup> the Organization for the Advancement of Structured Information Standards

<sup>13</sup> [http://www.oasis-open.org/committees/tc\\_cat.php?cat=soa](http://www.oasis-open.org/committees/tc_cat.php?cat=soa)

It can also realise cost savings through a utility model – ie using computing power in the same way as electricity and other utilities, ie only using computing power and software as needed

### **The value of interoperability**

The NIHR is not proposing to build a new system to supersede existing systems, but will ensure that information flows effectively between existing systems. Interoperability – the capacity for individual systems to exchange information and to use the information that has been exchanged - is therefore a key concern for the Architecture, and is supported in a number of ways:

1. Identification of fundamental system entities – the key entities or classes of information in the NIHR that are common to all business processes and across all systems and applications
2. A data dictionary which defines the meaning (semantics) and structure (syntax) of data elements used across NIHR. The data dictionary is based on national and international standards and is intended to enable exchange of data between applications in a meaningful way.
3. Defining and promoting a service-oriented style of integration. This will enable applications to co-ordinate their activities without needing to merge or combine.
4. Provision of key service components, including a personalisable portal, and (as part of the MIS) a lifelong research record and an identity management service.

The next sections develop the architecture of each of the fundamental system entities, identifying for each:

- the business processes with which each is associated
- the key system components associated with each
- the data attributes of each
- the services which each supports
- the underlying technologies supporting each

# Research project class/entity

Efficient management of research projects is the over-riding requirement placed on NIHR. The research project entity is correspondingly the single most important class of information in the Architecture.

The following sections consider the business, information and technology architectures of the research project entity. Of note, the architecture proposes a **lifelong research record** – a unitary record of each study supported by NIHR, encompassing the portfolio record but with added ability to provide and consume services. The record and associated services provide the basis for supporting the reorganisation of research management processes in Work stream 4.2, including the CSOU, CLRN and other systems.

## Business Architecture

The essential business processes and workflows of a research project are:

1. Identification of needs and priorities
2. Idea/proposal generation
3. Funding and resource allocation (direct and support costs)
4. Regulatory Approval
5. Acceptance by the NHS
6. Contract award and contract management
7. Conduct including patient recruitment, data gathering and analysis
8. Reporting and disseminating to decision makers

Many agencies are involved in this workflow and it is widely accepted that the workflow is too convoluted. Key aspects of the workflow as-is have been documented by Soteriou and Bourne<sup>14</sup> and will not be replicated here.

## Information Architecture

### Key service component – the lifelong research record

A lifelong research record would have the following characteristics:

- Manifested as a web site, part public part private
- Proposed branding identity is 'NIHR study space'
- Owned by the study PI, who would grant access to the private part of the site as they choose
- Digital Object Identifier (DOI) as unique identifier; also hold ISRCTN and other identifiers
- WHO 20 item minimum dataset, plus other items as necessary
- Indexed using UKCRC classification
- Hosted on UKCRN portfolio database
- Can hold various documents relating to the study – letters, reports etc

---

<sup>14</sup> reference needed

- Can interact with NIHR and other funding and regulatory systems
- Can link to clinical research management systems (eg PsyGRID, TRIM)
- Can hold outcomes and link to UKPubMedCentral

### **Data attributes**

Analysis shows that NIHR information systems hold 214 data items about a research study.<sup>15</sup> While many of these items are needed by local systems, the Architecture focuses on those items that are common to all information systems, or those that need to be shared by systems or made available externally.

Accordingly, the NIHR data dictionary for research studies will begin with the WHO clinical trials data set. The first set of additional data items (ie data items that need to be exposed to other systems) will come from the following applications:

- Portfolio Database
- CSOU pilot

In addition the Architecture proposes to:

- Pilot the use of Digital Object Identifiers (DOIs) as the unique identifier for a research study. The CSOU pilot has registered the prefix 10.3333 for the NIHR and will experiment with the use of DOIs
- Use the UKCRC scheme for subject classification

### **Applications and Services**

The design goal of the architecture is to establish the lifelong research record as the definitive source of information about a piece of research. At present this information is distributed among several systems, which are described in the current systems inventory, available from the Programme Office.

In meeting the design goal it is not the intention to implement a single system or replace existing systems. Instead the proposal is to bring the key information about a research study into a single record and to have this record inter-operate with funding, regulatory and management systems through web services.

It may be possible over time to bring more and more functionality into one system, and to reduce redundancy in existing management systems but we are not making these proposals now, nor does a lifelong record require significant change to existing systems.

The lifelong research record is the core of the NIHR Management Information System. It can be built as part of the portfolio management database. It can

---

<sup>15</sup> Study Data Mapping v 1.0 , 30<sup>th</sup> November 2006. Author Hazim Timimi.

start to be built over the next few months as part of the IS Programme work with CSOU, CLRN and other networks. Further development can take place as soon as resources permit engagement with NRes, CCF, SDO, and HTA.

To emphasise its functional nature the lifelong record could be instantiated and branded as Study Space – a managed space in which research studies could be ‘worked-on’ by those who need access to them. The early iterations of Study Space would provide the following minimal application level functionality:

For PIs (and their nominees) and Research Networks:

1. Create and edit a record
2. Add a document relating to a study

For CSOU staff:

1. Edit a document to indicate that approvals were in place
2. Record CSOU approval
3. Other functions as determined by the CSOU URS

In addition we recommend engagement with the forthcoming NCCHTA information system to explore the use of Study Space in relation to their new research management application.

### **Other services**

At this stage we are proposing one new service for Study Space – an RSS feed service that allows:

- a PI to list the studies they own
- a research manager to monitor newly added records

### **Future services**

A priority area for future service development concerns the exchange of data between the lifelong research record and ethics and regulatory systems. The goal will be to enable researchers to enter data once and to have this data exported or inspected by regulatory agencies.

### **Technical Architecture**

The principal technical requirement is that research project data is held in a relational database and that this database is capable of processing web services. The existing UKCRN Portfolio Management System database meets this requirement.

# Person class/entity

Many business processes across NIHR involve the use of information about people. At the moment such information is highly silo-ed and held in a variety of formats and with little common data structure.

The target architecture includes a store of information about people, including publicly accessible data and secure data. This store would offer services to every process needing information about people, and would also offer a number of end user services such as a contacts directory.

Much of the analysis relating to people forms part of Identity Management, an important topic for NIHR information systems and which will be a particular focus for development in the next year.

## Business Architecture

Fundamental business processes include:

1. Stakeholder Relationship management (ranging from simple contacts lists, through committee management, through to relationship management)
2. Faculty Management
3. Research passport
4. Services to individual researchers
5. Increasing recruitment to clinical trials
6. Collaboration and networking
7. Peer reviewer management

## Information Architecture

### Key system component

Central to the management of people within NIHR systems is a personal web page, created as a personalisation of the NIHR portal.

This web page would have a number of roles:

- It would be an individuals' 'home' within NIHR
- It would form the target for delivering personalised information from internal and external systems
- It would be configurable by the user to reflect their interests
- It would be the launch pad for applications that the user was entitled to use (eg CSOU application)
- It would be part of the identity management infrastructure
- It would provide and consume web services. For example it could play a role in support of services such as Research Passport

## **Data attributes**

The initial data set for person information will begin with the data needed by Faculty, which has been identified by DH<sup>16</sup> as:

- name,
- date of birth,
- job title,
- employer,
- email address: (initially .nhs or .ac.uk only for Faculty members)
- gender
- disability (yes/no) – if yes are you registered disabled?
- Ethnic group

These data items have been included in the NIHR data dictionary, with definitions taken from the NHS data dictionary.

## **Services/applications**

There are very many services that could be developed using the person entity including: a personal research portal; social web services in the style of Facebook and Connotea; collaboration tools such as videoconferencing, VOIP and shared workspaces; weblogs, CVs and other content generation; single sign-on to NIHR and other systems; Research passport; etc.

We propose to begin with a few simple steps. The first is to support the NIHR Faculty by establishing a simple membership database, using data supplied by CCF; SDO; NCCHTA; NCCRCD; and other sources as required by DH. At the same time we will register CSOU staff so that they can trial single-sign-on.

Initially, the database will offer four services:

- a simple personal page (ie a proto personal research portal) on the NIHR portal
- a directory look-up service, available to Faculty members and senior NIHR staff
- a login service to CSOU staff
- RSS feed of central changes to non-confidential data back to supplying agent.

## **Technical architecture**

As discussed, the key technical requirement is to establish a definitive database of people associated with NIHR, including Faculty members, other grant holders, members of networks, stakeholders, and NIHR staff.

As of July 2007 there is some complexity in the configuration of databases and directories that hold information about people and manage access to

---

<sup>16</sup> NIHR Faculty. Final contact, data and app form details, June 2007.

systems. Further work is in train to investigate the following applications to support identity management:

- NIHR portal
- UKCRN CMS
- OpenID
- MS Cardspace

# **Finance class/entity**

To date little work has been done to define the architecture of financial information management for NIHR as a whole.

It is recognised that with the new roles for CLRNs in 2008 it will be important to manage financial information across NIHR in an effective and efficient way.

It is proposed that a future EA work package focuses on business, information and technical aspects of financial information.

# Facility class/entity

Within the NIHR Architecture, the Facility class of data concerns any organisation of relevance to the NIHR. It includes to any specific places within an organisation at which research takes place or which provides resources to support research.

Given the broad range of business architecture associated with facilities, the scope of potential inclusions is similarly broad. They include:

- Funding bodies (including research charities)
- NIHR networks
- NHS organisations
- Specific facilities within NHS bodies
- HE/FE bodies
- Regulatory agencies
- Commercial organisations
- etc

## Business Architecture

Understanding the location and distribution of research resources is arguably of secondary importance to the overall management of NHS research. However, for some processes, location is of great significance. These include:

- Managing and supporting research infrastructure in the NHS – the CLRNs will in future be responsible for ensuring that finances are directed in a carefully targeted way to provide the support budgets for NIHR supported research.
- Patient recruitment to clinical studies. This requires a detailed knowledge of the places research is happening to map recruitment and accruals

In addition, good quality location based information is important to efficient operational processes such as:

- Communications and stakeholder management
- Distribution of business

## Information Architecture

### Key system components

The key system component is a database of facilities.

### Data attributes

Data attributes for facilities are comprehensively enumerated in the NHS data dictionary and HM Government. In addition:

- Unique identifiers Hazim Timimi has analysed the current approach to providing unique identifiers for organisations. There are several part-coverage schemes (eg the NHS organisation codes) but no overall scheme. The data dictionary suggests a pragmatic approach to unique identifiers.
- NHS organisational codes The NHS Organisational Codes Service supplies a database of organisational codes. This dataset includes 'boundary' information, showing which organisations are part of Strategic Health Authorities

It is proposed that the data items in the UKCRN organisation database (see below) form the basis for the organisation dataset, subject to further analysis.

### **Services/applications**

At this stage a simple directory look-up service is proposed. Further services can be developed as required.

### **Technical Architecture**

The UKCRN Portfolio Management System includes a number of Organisation tables. Initial analysis of these suggests they would be suitable for holding Facility data.

# Implementation

## General approach

The blueprint contained here is for an information system that is somewhat different from the existing information systems in NIHR. Implementation of a new architecture therefore requires change to the configuration of existing systems. Many of these were conceived pre-NIHR, and are (more or less) meeting the needs of the individual agencies they reside in, making the implementation of an overall NIHR approach to managing information more challenging.

One option would be to commission a new all encompassing NIHR information system. But this would be very high risk, and has been rejected in favour of a systems integration approach.

How can systems be integrated (and hence modified) whilst continuing to deliver the operational services they were intended for? The general approach to implementation proposed here is to introduce some systems elements which enable a more integrated approach (for example portal, identity management and data dictionary); and to work on new projects as they arise to implement a more integrated approach. For example the CSOU application is starting to implement the lifelong research record; work with the Faculty is seeking to establish the basis for identity management.

## Changes needed

The changes needed to take the NIHR information system to its target architecture include:

1. Implementing key system components. These include: personal home page and lifelong research record
2. Implementing a range of shared services such as identity management and a contacts directory
3. Taking a strategic approach to data management for each main class, ensuring accuracy and completeness. This will require greater focus on data management across NIHR than is presently the case.
4. Establishing data standards including a data dictionary for data that is to be shared across applications and APIs to enable interoperability
5. Influencing stakeholders and existing projects to work in an orchestrated way, by establishing governance for the Architecture and through partnership working.

## Change methods

The proposed methods for delivering these changes are:

Item		Action
System component	Personal home page	Developed as part of NIHR portal project and

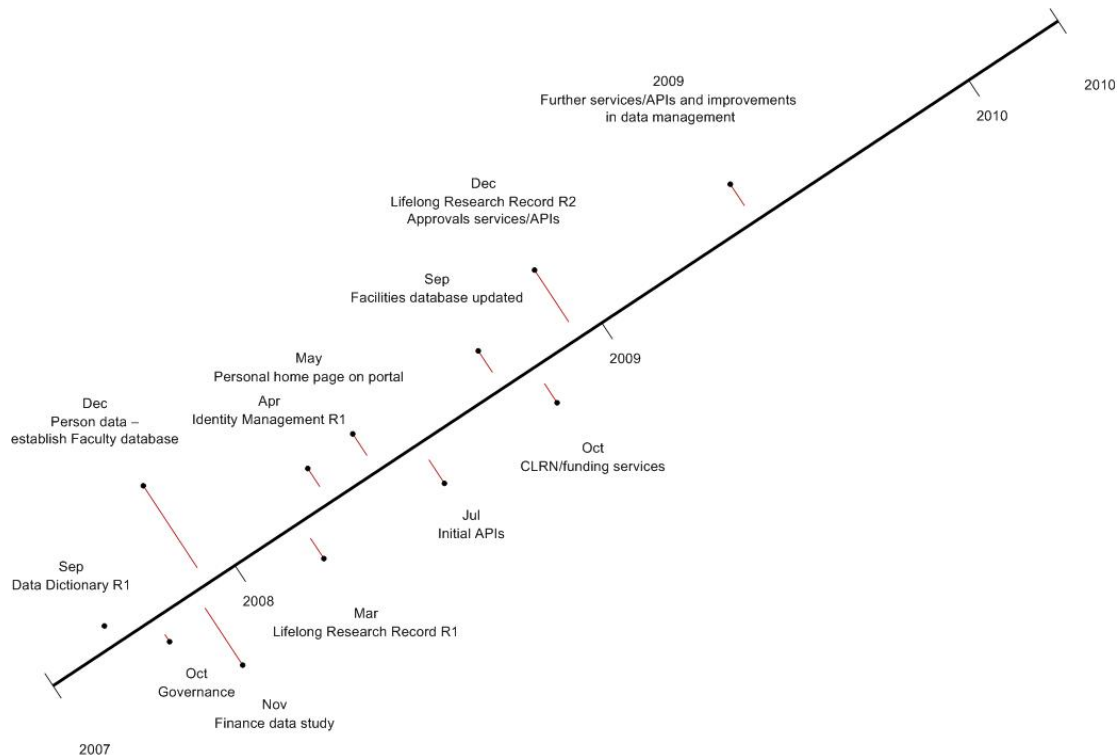
		refined according to user requirements
System component	Lifelong Research Record	Developed by extending the UKCRN portfolio database Implemented in stages, beginning with CSOU project; then integration with ethics systems.
Shared service	Identity Management service	Establish infrastructure to support single sign on as part of portfolio database development
Shared service	Directory service	Implement as a portal service as soon as possible
Strategic data management	People data	Developed in stages by developing the UKCRN contacts management system. Begin with data about Faculty members
Strategic data management	Research project data	Developed in stages based on UKCRN portfolio database. Begin with CSOU application.
Strategic data management	Facilities data	Create a definitive database of all facilities, organisations and networks associated with NIHR, based on the UKCRN organisations tables in the portfolio management system
Strategic data management	Facilities data	Create a definitive database of all facilities, organisations and networks associated with NIHR, based on the UKCRN organisations tables in the portfolio management system
Data standards	Data Dictionary	Establish a data dictionary, based on the NHS data dictionary and others
Data standards	APIs	Develop initial APIs for main classes of data and further ones in response to need
Influencing	Governance	Establish a sub-group of

stakeholders		the NIHR IS Programme Board
Influencing stakeholders	Partnership working	Work in an open way; publish documents and drafts on NIHR architecture web site. Incorporate architecture into new systems.

## Roadmap

The transition from the current multiplicity of architectures to a single enterprise architecture can be largely completed in 3-5 years. Most of the changes described above are not one-off events, but continuous processes, which can begin now, or as soon as dependencies and resources allow.

A timeline is shown below. It is not intended as a detail plan, but shows 'way-markers' towards achievement of the target architecture.



# Data Dictionary

The NIHR Data Dictionary is composed as follows

1. It is divided into four data classes
2. Each class contains a number of attributes
3. Each attributes has an identifier, a semantic definition, a syntactic structure, and any other information of use to a system designer, about an element of information in the NIHR system. For example 'First Name' is the identifier of an attribute about a person's name
4. As far as possible semantics and syntax will be derived from existing data dictionaries, such as the NHS Data Dictionary<sup>17</sup> and the UK Government Data Standards Catalogue<sup>18</sup>
5. Attributes are provided only for data that is likely to be shared between systems. Data that remains internal to a single system will not be included in the Data Dictionary though it may be held in an application data dictionary
6. The Dictionary will evolve so that it continues to meet requirements
7. The Dictionary is related to but not the same as database table schema.

## Data Class: Person

Definition: Any person included in any NIHR business process.

Includes: staff, researchers, Faculty members, stakeholders, including members of the public interested in research..

Excludes: research study subjects

The primary source for data structure about people is the UK Government Data Standards Catalogue Person Name element<sup>19</sup>

## Attributes

Identifier	Definition	Field syntax	Notes
NIHRID	A unique identifier for a person	NIHRxxxxxx	Allow for up to 999,999 unique IDs
Person Title	Standard form of address used to precede a person's name.	Max 35 characters	Eg Dr, Professor, Mr, Ms
Person Given Name	The forename or given name of a	Max 35 characters	

<sup>17</sup> <http://www.connectingforhealth.nhs.uk/systemsandservices/data/datamodeldictionary/datamodeldictionary>

<sup>18</sup> <http://www.govtalk.gov.uk/gdsc/html/default.htm>

<sup>19</sup> <http://www.govtalk.gov.uk/gdsc/html/noframes/PersonName-1-1-Release.htm>

	person.		
Person initials	Used to record a person's initials.	Max 35 characters with a space between each initial.	This can be used to record a person's full set of initials or used in conjunction with and in addition to a Person Given Name. The general practice of central departments is to record the first Given Name in full and then use initials for other Given names.
Person Family Name	That part of a person's name which is used to describe family, clan, tribal group, or marital association.	Max 35 characters	
Person Name Suffix	A textual suffix that may be added to the end of a person's name	Max 35 characters	for example, OBE, MBE, BSc, JP, GM.
UK Postal Address	The address of a postal delivery point. This is usually a building and usually comprises a name/number, street, town and county/state/province and sometimes country.	Max. 35 characters per address line. Minimum 2 lines, no maximum although a de facto standard exists for 5 lines for presentation purposes.	Use these fields to record a person's postal address if they have no organisation/employer, or if their address is different to their organisation/employer
Postcode	The code allocated by the Post Office to identify a group of postal delivery points.	Max 8 characters. Use a space between first and second part.	
Person Birth Date	Date of birth	10 Characters in the format CCYY-MM-DD	
Job Title	Description of persons main job/role		
Employer	The body which provides the main		Store the organisation code. Screens can

	contract of employment		display organisation
e-mail address	The string of characters that identifies an addressee's post box on the Internet.	255 characters	
Person URL	The URL associated with a person	Use format http://<host>	This would normally be the NIHR home page of a person
UK telephone Number	A number, including any exchange or location code, at which a person or organisation can be contacted in the UK by telephonic means.	35 characters, no spaces	Mobile or landline
Gender	A person's gender currently	0 = Not Known 1 = Male 2 = Female 9 = Not specified	Based on ISO/IEC 5218:2004
Disability	Indicates whether the person considers themselves to have a disability	Yes/No	Self defined
Person Ethnicity <sup>20</sup>	A person's ethnic group	For England and Wales: ----- ----- WHITE British Irish Other White ----- ----- MIXED White & Black Caribbean White & Black African White & Asian Other Mixed ----- ----- ASIAN OR	Self defined. Required for equal opportunities monitoring.

<sup>20</sup> For further details see [http://www.statistics.gov.uk/about/ethnic\\_group\\_statistics/](http://www.statistics.gov.uk/about/ethnic_group_statistics/)

		ASIAN BRITISH Indian Pakistani Bangladeshi Other Asian ----- ----- BLACK OR BLACK BRITISH Caribbean African Other Black ----- ----- CHINESE ----- ----- OTHER ETHNIC GROUP Any Other (please describe)	
--	--	--	--

### Research Project class

Definition: Any research project provisionally or actually supported by NIHR.

Includes: all types of research taking place in the NHS

Excludes: studies not adopted by NIHR

### Attributes

Identifier	Definition	Field syntax	Notes
Primary Register	The name of the primary register holding information about a study	Default value = NIHR	
UK Internal Code	A unique code representing the 'country of origin' of a research study.	Max 2 characters: 1 England 2 Scotland 3 Wales 4 Northern Ireland 5 Isle of Man	Use to distinguish between NIHR portfolio (England only) and UKCRN portfolio (UK wide)

		6 Guernsey 7 Sark 8 Jersey 9 Alderney 10 Channel Islands	
ICTRP	Indicates whether a study is intended for ICTRP	Yes/No	The purpose of this field is to distinguish between NIHR adopted studies and clinical trials to be included in ICTRP.
Portfolio	Indicates the status of a study in relation to the UKCRN Portfolio	0 = not part of portfolio 1 = provisional 2 = adopted	
Study ID#	The system identifier for a research study that has been included in the NIHR portfolio	Alpha-numeric. NIHR DOI is being piloted at this time	NIHR provisional DOI prefix is 10.3333
Date of Registration	The date at which a study was added to the NIHR portfolio	10 Characters in the format CCYY-MM-DD	
Secondary ID#s	Any other identifiers useful to study management	Alpha-numeric	This field should be stored as pairs of data {scheme, id} in a separate table to allow for expansion.
Source(s) of monetary support	The organisation providing the majority of direct funding		Store organisation ID
Primary Sponsor	The organisation which is sponsoring the study		Store organisation ID
Secondary Sponsors	The organisation which is also sponsoring this study		If it is important to store several secondary sponsors, hold this in a separate table
Contact for public queries	The name and e-mail of the person		

	who can be contacted to answer general questions about the study		
Contact for scientific queries	The name and e-mail of the person who can be contacted to answer scientific questions about the study		
Public title	A short (and plain English if possible) title for the research study		
Scientific title	The full descriptive title of the study		
Acronym	Study Acronym if any		
Countries of recruitment	Countries in which patients are being recruited		List England, Wales, Scotland, Northern Ireland separately
Health condition(s) or problem(s) studied	The main health condition being studied	SNOMED code	Code to the greatest level of detail available. SNOMED codes can be aggregated to match UKCRC Health Categories.
Intervention name(s)	The name of the intervention (if any) being studied		If no intervention, write n/a. Use dm+d code for a drug if it is available
Other details	Any other information about the intervention that will be of use to readers		
Control intervention	The name of the control intervention (if any)		
Other details	Any other		

	information about the control intervention that will be of use to readers		
Key inclusion criteria	The main inclusion criteria for a study	Further work is needed to determine how structured this data item should be	
Key exclusion criteria	The main exclusion criteria for a study	Further work is needed to determine how structured this data item should be	
Study type	The type of study – ie the research methodology	Use the classes in the Portfolio Management System	
Research Activity Code	The area of research according to the UKCRC Research Activity Code scheme <sup>21</sup>		
Date of first enrolment	The date (planned or actual) on which the study is open to recruit patients	CCYY-MM-DD	
Target sample size	The total number of study subjects to be recruited to a study	Numeric	
Recruitment status	The status of a study in relation to recruitment	<p>Recruitment status of this study.</p> <ul style="list-style-type: none"> <li>• <u>Pending</u>: participants are not yet being recruited or enrolled at any site</li> <li>• <u>Active</u>: participants are</li> </ul>	These are the WHO recommended categories

<sup>21</sup> <http://www.ukcrc.org/PDF/Health%20Classification%20system%202014%20June%20.pdf>

		<p>currently being recruited and enrolled</p> <ul style="list-style-type: none"> <li>• <u>Temporary halt</u>: there is a temporary halt in recruitment and enrolment</li> <li>• <u>Closed</u>: participants are no longer being recruited or enrolled</li> </ul>	
Primary Outcome Name	The main outcome measure (if any) that the study is measuring		
Primary Outcome time points	The time points at which the primary outcome will be measured		
Key secondary outcomes	Other outcome measure (if any) that the study is measuring		
Key secondary outcomes time points	The time points at which the secondary outcome will be measured		

### Facilities class

Definition: Any organisation of relevance to the NIHR.

Includes: Organisations and parts of organisations. Funding bodies (including research charities); NIHR networks; NHS organisations; specific facilities within NHS bodies; HE/FE bodies; Regulatory agencies; Commercial organisations; etc

Excludes: organisations of no relevance to NIHR

### Attributes

Identifier	Definition	Field syntax	Notes
Facility Identifier	A unique identifier for a facility	NIHRFxxxxxx	Can be generated from an RDBMS

			unique ID
Name	The name of a facility		
Organisation	The identifier of an organisation of which a facility is part	For an NHS body, use NAC code. For education providers, use UKPRN <sup>22</sup> For a commercial organisation, use the company number <sup>23</sup> For a charitable body, use Charity Registration Number <sup>24</sup> For other bodies, use a simple alpha numeric code	Since an organisation may include several facilities there should be a separate organisation table
UK Postal Address	The address of a postal delivery point for a Facility. This is usually a building and usually comprises a name/number, street, town and county/state/province and sometimes country.	Max. 35 characters per address line. Minimum 2 lines, no maximum although a de facto standard exists for 5 lines for presentation purposes.	
Postcode	The code allocated by the Post Office to identify a group of postal delivery points.	Max 8 characters. Use a space between first and second part.	
Contact	The name of a person who can be contacted with reference to this facility	Use ID from person class	
CLRN	The name of the CLRN in which		Use a standard pick-list of CLRN

<sup>22</sup> UKPRNs are available from <http://www.ukrlp.co.uk/ukrlp/ukrlp.first>

<sup>23</sup> available via an XML gateway <http://xmlgw.companieshouse.gov.uk/index.shtml>

<sup>24</sup> available via <http://www.charity-commission.gov.uk/>

	this facility resides		
Notes	Any unstructured information about a facility of relevance to system users		

### **Finance class**

The Finance data dictionary will be developed during 2008.

Definition: Any financial resource provided by NIHR..

Includes: tbc

Excludes: tbc

### **Attributes**

<b>Identifier</b>	<b>Definition</b>	<b>Field syntax</b>	<b>Notes</b>

## Appendix 1 Conceptual diagram of systems

The following diagram shows the essential features of the overall NIHR information system, including IS Programme systems (Red) other systems (orange) and linkages (green arrows).

