ULICE WP7 update

Interim report

Steve Harris, University of Oxford

WP7 recap

- established metadata services
- recommended fundamental standards
  - terminologies
  - follow-up variables and form design
  - evaluated external caBIG standards
- looked at how these standards matched with practice
- looked at how standards could be deployed as services for trial harmonisation

ULICE

What we learned (1)

- terminology is an issue
  - comprehensive disease terms lacking, *or*
  - terms we want to use are not in the same terminology
  - organisation of the terminology may not be suitable
    - Snomed CT vs BFO/OBO

ULICE

	Proposed Question	Value domain	Candidate question(s) from caDSR		comments
1			Randomisation/on study: Patient Identification		
1.1	Centre ID	Identifier (text)	Participating Organization ID, Organization ID number	C	No direct match, Queried with Dianne Reeves. They have suitable definitions in CTEP but they are overly specific
1.2	Patient ID	Identifier (text)	Subject No. Study Participant Identifier	A	Exact match. Fulfils D.MGT/JRA/NA/TA 7.2 table 3 question 1
1.3	Initials	Text	Participant initials	A	
2			Randomisation/on study: Review of Inclusion and Exclusion Criteria		
2.1	CoMéTHE agreement	Y/N	(no match)	N	What is CoMeTHE?
2.2	Acceptable delay of treatment (delay to be defined)	Y/N	(no match)	N	
2.3	Age > 18 years	Y/N	Age >= 18 years	A	Exact match
2.4	Signed informed consent	Y/N	Has written informed consent been obtained?	A	Exact match
3			Randomisation/on study: Inclusion visit		
3.1	Inclusion date (= randomisation date)	Date: dd/mmm/yyyy	Randomization Date Enrollment date	A	Exact match on randomisation data. American format date for 'Enrollment Date' (sic). Fulfils D.MGT/JRA/NA/TA 7.2 table 3 question 2
3.2	Sex	M/F	Patient gender	N	Value domain also includes 'unknown' and 'unspecified'. Is this a problem?
3.3	Date of Birth	Date: dd/mmm/yyyy	Patient's Date of Birth	N	Partial match: American date format - need one that is used in France/Europe.
3.4	ECOG performance status	0/1/2/3/4	Performance Status (ECOG) ECOG Performance Status	A	Includes a 'Dead' status. Obviously not appropriate for prospective trial. Need to clarify which is preferred for our application
3.5	Weight	Number (kg)	Body Wt (Kg) (most widely used exact match)	A	The standard elements are a weight number and enumeration for the units: Patient weight; Weight unit of measure
3.6	Smoking status	Smoker / ex-smoker / never smoked	Smoking History	N	Close match: a little more complicated, ex-smokers are subdivided into those who have stopped smoking >15 years ago
3.7	History of cardiovascular disease (coronary artery disease, MI, stroke, peripheral arterial disease)	Y/N	Cardiac	N	Should be checked: definition is 'yes/no/unknown indicator that asks whether there was a history of cardiac comorbidities'.
3.8	Diabetes	Y/N	Diabetes	A	The definition is slightly confused because (I think) it is generated from the concepts that annotate the variable. The

What we're doing (1)

- collate and present terminology developed in this project (WP2 especially) in appropriate formats, together with other sources such as parts of Radlex
- impetus from the Maastro clinic to develop a 'Radiotherapy Oncology Ontology' as a part of the Open Biomedical Ontology foundry – we intend to develop this
- looking for collaborators and content …

ULICE

what we've learned (2)

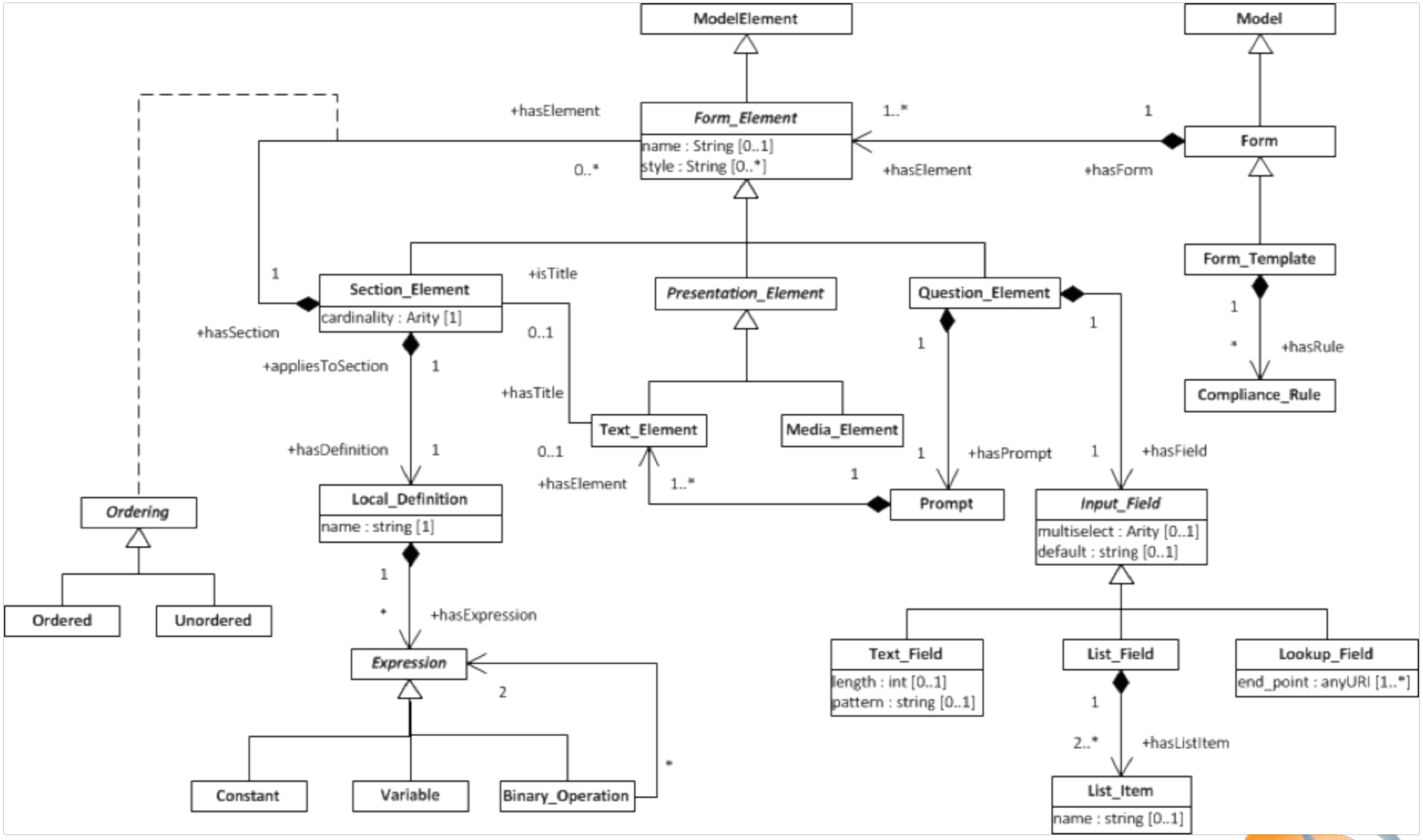
- not simply enough to set standards for the variables on case report forms …
- … need to develop mechanisms to standardise template forms, form sections, whole forms
- … need mechanisms to exchange forms between trials management systems
- working out how to map section and section modifiers to Snomed CT expressions

ULICE

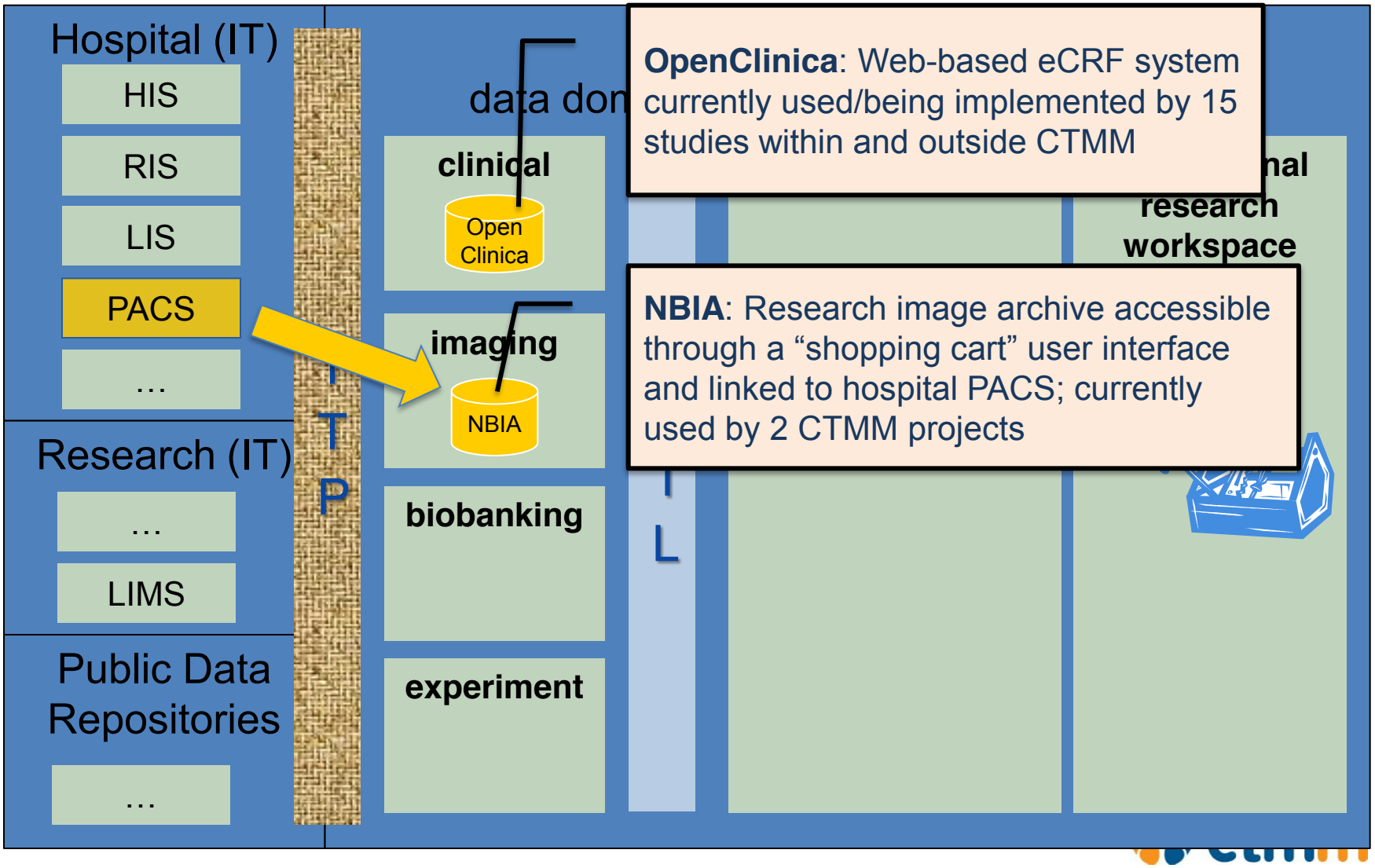
what we're doing (2)

- developed excel plugin to author OpenClinica and RedCap forms - working with TraiT to roll out OpenClinica form design services to Dutch clinical research
- developing an international standard for the representation of the semantics of forms under ISO JTC1 SC32 WG2 – ISO19763-13

ULICE


ModelElement
Model
+hasElement
Form_Element
name : String [0..1]
style : String [0..*]
1..*
1
Form
0..*
+hasElement
+hasForm
Form_Template
1
Section_Element
cardinality : Arity [1]
+isTitle
Presentation_Element
Question_Element
1
*
+hasRule
+hasSection
0..1
1
1
+appliesToSection
1
+hasTitle
Compliance_Rule
Text_Element
Media_Element
+hasDefinition
1
0..1
1
+hasPrompt
1
+hasField
+hasElement
1..*
1
Prompt
Ordering
Local_Definition
name : string [1]
Input_Field
multiselect : Arity [0..1]
default : string [0..1]
1
*
+hasExpression
Ordered
Unordered
Expression
2
Text_Field
length : int [0..1]
pattern : string [0..1]
List_Field
Lookup_Field
end_point : anyURI [1..*]
1
2..*
+hasListItem
*
Constant
Variable
Binary_Operation
List_Item
name : string [0..1]

Where are we now: ready for use



this year's deliverables

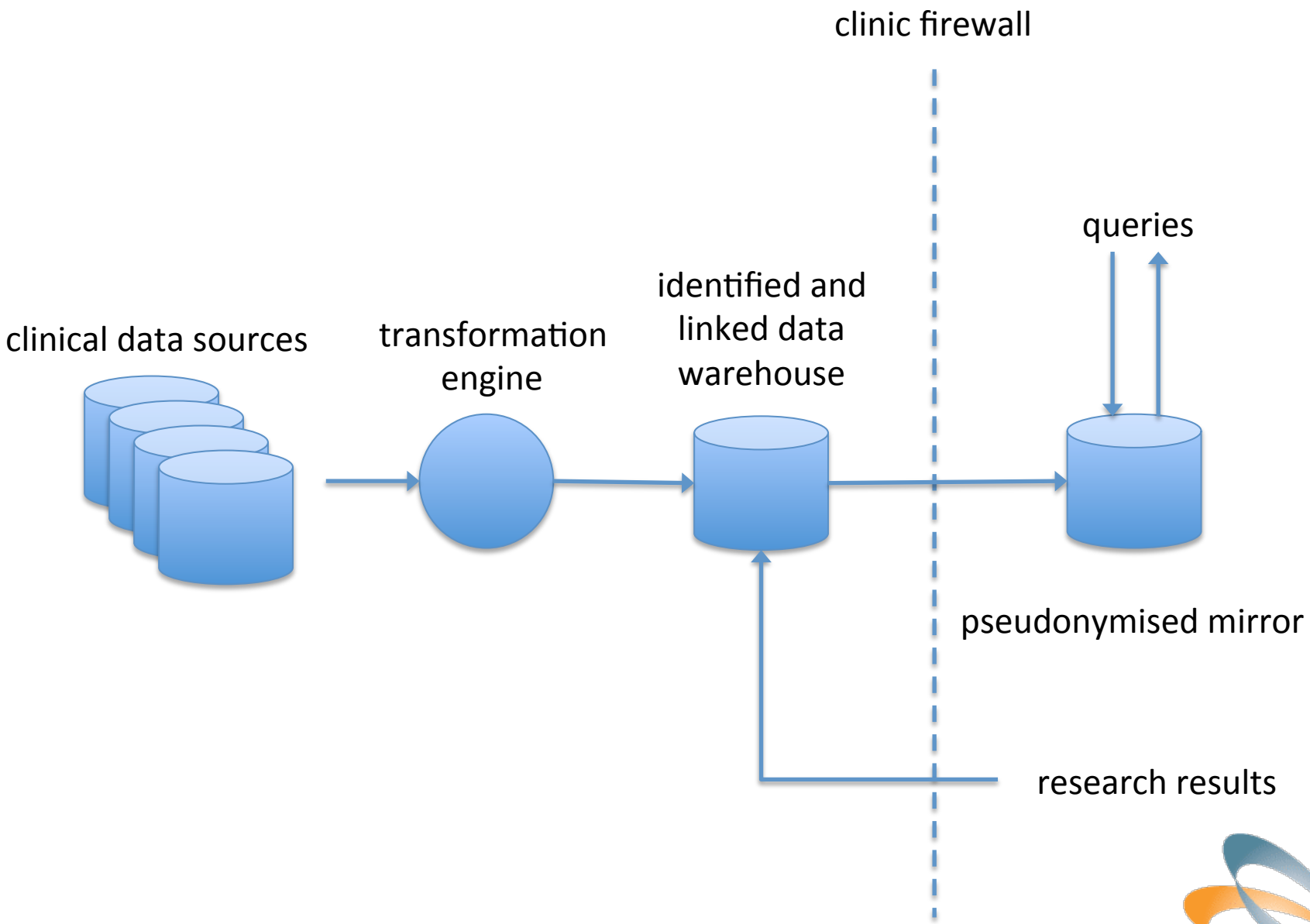
- 7.3 prototypical services for the capture, cleaning, anonymisation and federation of case history data
- 7.4 prototypical services that allow a clinician to locate case histories, to view summary information, and to request detailed information in support of treatment planning
- … support for records based research

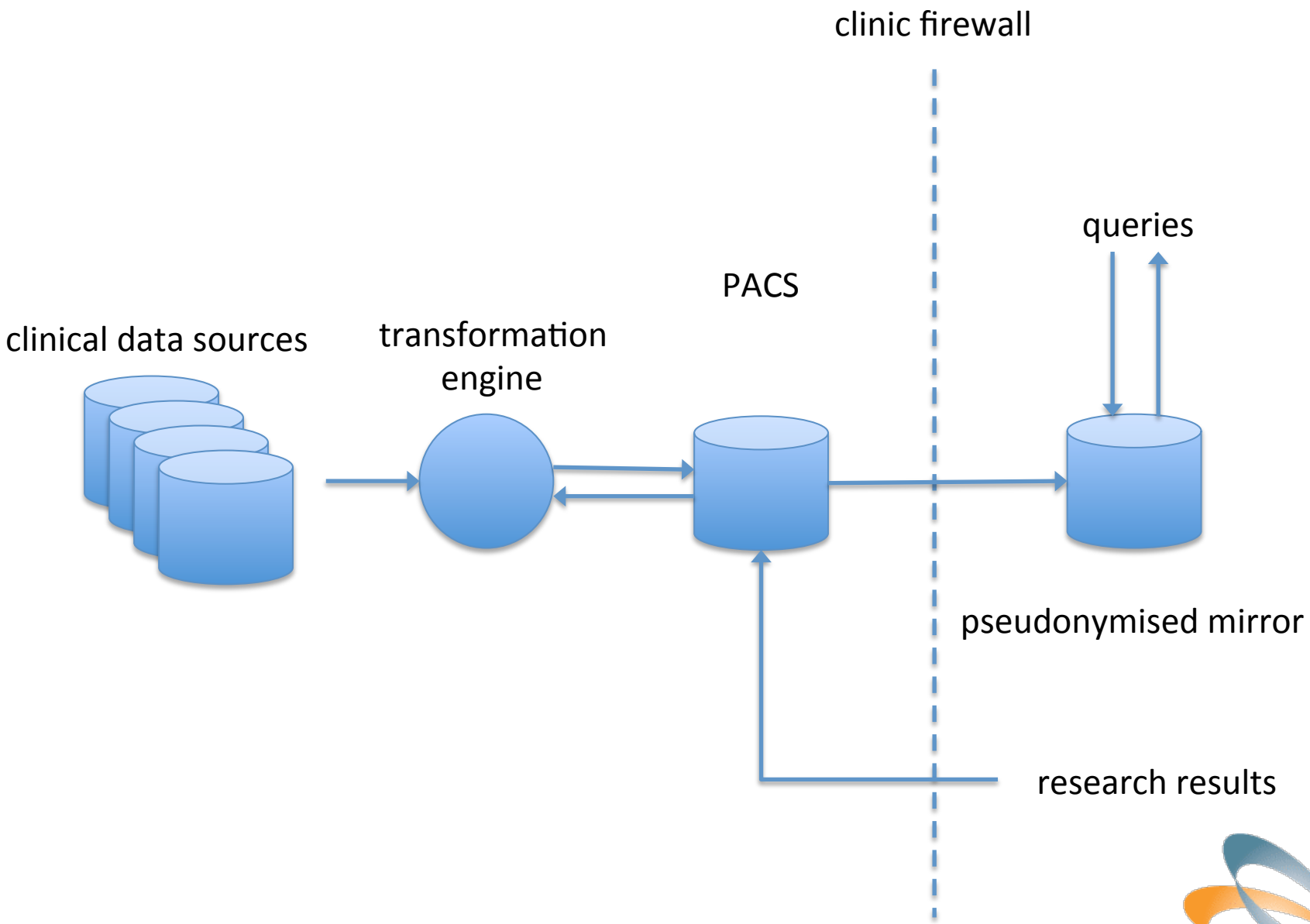
ULICE

records based research

- formulating trial questions in radiotherapy is difficult thus there is a continuing need for case-based research
- environment improving
  - health records likely to be electronic
  - concerted efforts (in UK at least) to facilitate use of clinical records in research
- how do we develop research databases alongside clinical practice?

ULICE


clinic firewall
queries
identified and
linked data
warehouse
clinical data sources
transformation
engine
pseudonymised mirror
research results


clinic firewall
queries
PACS
clinical data sources
transformation
engine
pseudonymised mirror
research results

transformation

- standard clinical information formats unsuited to research
  - wrong terminologies
  - obtuse data formats
  - message based
- access to line of business systems problematic
  - administrative hurdles, policy
  - reliability of your software
- why not just use the integration engine that already exists
  - Mirth Connect

ULICE

data services

- linkage
  - joining based on common identifiers and attributes
  - validating joins by cross checking data
- annotation
  - translation of clinical terminologies into research terminologies – crosswalks, mappings, NCI meta
  - contextualisation – annotations to describe common properties
- anonymisation/pseudonymisation
- blurring

ULICE

query services

- EL++ subsumption queries on Snomed CT terms
  - difficult, even the best reasoners cannot cope with all of SCT
  - need to isolate relevant portions of the tree – e.g. 55342001|neoplastic disease| for diagnosis
- dataset sensitivity – what do you have to do to qualify to receive this data?

ULICE

anonymisation/reidentification

- flipside of record linkage
- how do we publish data that is both useful and difficult to join to other, unknown data sources
- obviously we remove identifiable data – names, addresses, demographics
- however, some demographics are important: geo-location; referral patterns; unique instruments

ULICE

sensitivity calculations

- a score for ‘re-identifiability’ can be calculated
  - the attributes selected: can any be used to reason about identity
  - the uniqueness of the record in the dataset and in the query result
  - the combination of successive query results by same person
  - calculate a weighted product
  - refer all requests over a certain score to oversight

ULICE

blurring

- frequently absolute information is not required for research
  - Kaplan Meier plots are of time elapsed rather than absolute time – so pre-compute relative times
  - statistical analysis bins data so tell the system about the bins
  - no need to obtain data to a precision higher than the accuracy of the observation

the deliverables

- radiotherapy test dataset to be supplied - with images - generated by Addenbrookes (Jena, Rajesh)
- prototypical services implemented as REST based XQuery implementations – easy to translate into Java/.Net based services
- source code, compiled implementations and designs available to project

ULICE