



Software Engineering Conference Russia 2018

October 12-13
Moscow

О решении проблем интеграции данных
в медицинской информационной системе
в НМИЦ им. В.А. Алмазова

Ирина Радченко¹, **Антон Яркин¹**, **Александр Чистяков¹**, **Игорь Николаев²**,
Ольга Лисицына³

¹Университет ИТМО, ²СПбГУ, ³Университет Турку

Цель и задачи

Разработка системы обработки диагностических данных Медицинского центра им. В.А.Алмазова с целью обеспечения научной обработки данных.

Решаемые задачи:

- интеграция с медицинской информационной системой;
- деперсонификация или анонимизация данных;
- хранение данных в аннотированной форме;
- обеспечение сохранности и целостности данных;
- доступ к данным при помощи браузера.

Перечень проблем

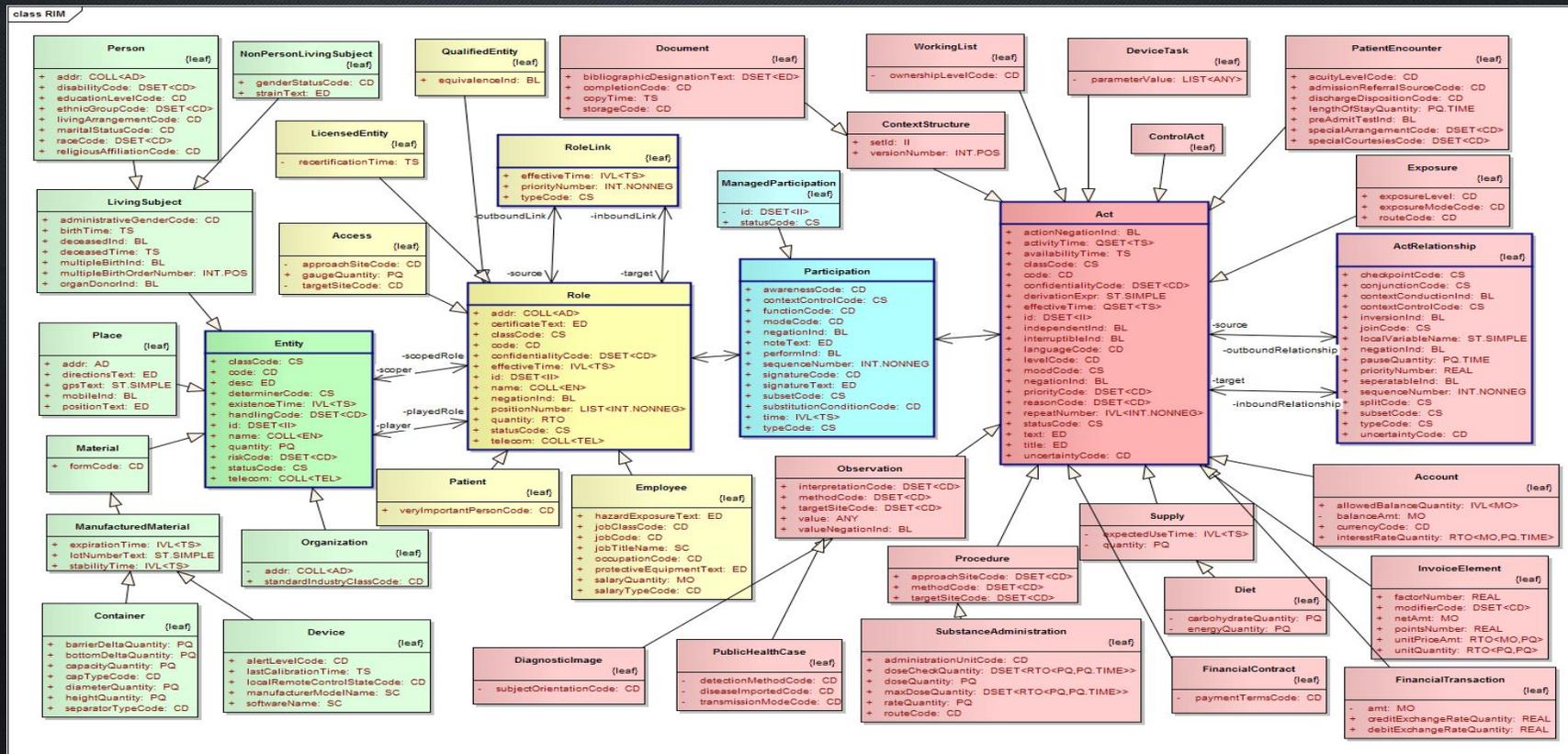
- HIS (Hospital Information System) используют специфические медицинские стандарты для обмена данными: DICOM и HL7.
- Интеграция HIS разных производителей затруднена как технологически, так и административно.
- Аннотирование данных и доступ к ним осуществляется в ручном режиме.
- Деперсонификация данных не предусмотрена.

HL7

- Health Level 7 - обмен клиническими и административными медицинскими данными.
- Появился в 1987 году, сейчас HL7 V3 использует XML-файлы.
- Лабораторные приборы любят формат HL7 V2.x ('|' - разделитель) вида:
OBX|3|NM||y^Lmps^Lcl^736-9^Lmps^LN||20|%|11-47|||F|207|b|12^T|
Где-то в инструкции прибора может быть написано что всё это значит.
- HL7 определяет RIM (Reference Information Model).
- HL7 велик, но производители приборов реализуют его по-своему.
- Тем не менее, HL7 не определяет, что именно должен отвечать прибор, например, давление можно измерить и в mmHg, и в Pa, размер -- в футах и см.

<http://www.hl7.org/>

Тяжелый мир HL7 Reference Information Model



Статус HL7 в России

Подкомитет ТК468 росстандарта “Информатизация здоровья”, стандарты HL7

- ГОСТ 21731-2013 эталонная информационная модель (RIM)
- ГОСТ 27931-2015 протокол обмена данными (protocol V2.5)
- ГОСТ 27932-2015 архитектура клинических документов (CDA)
- ГОСТ 27951-2016 общие терминологические службы (CTS)

Тексты из ГОСТ'ов о правах на стандарт противоречат нормативно-правовым актам РФ

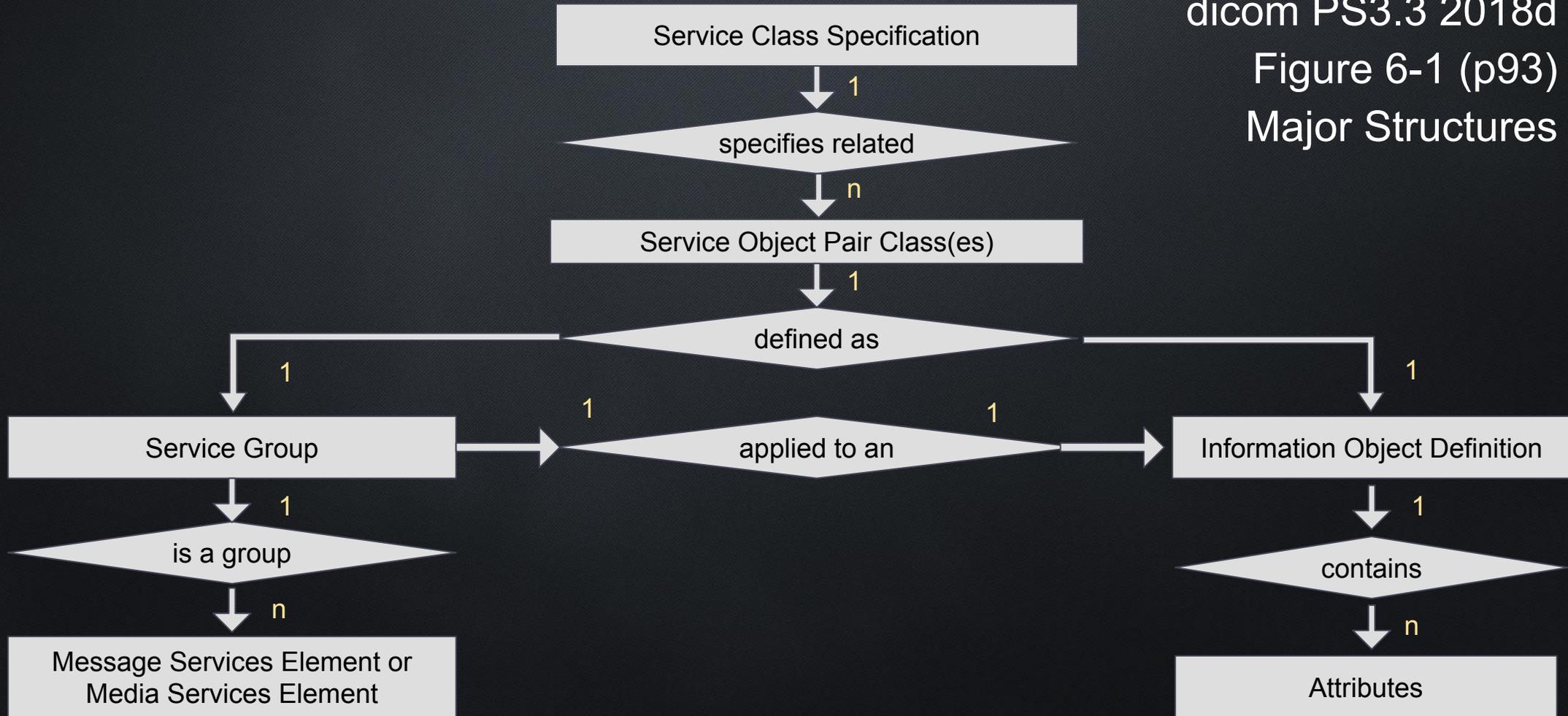
DICOM

- Digital Imaging and Communications in Medicine - стандарт для передачи, хранения, печати, обработки и отображения медицинских изображений.
- Создан в 80-х годах, имеет 16 редакций, 21 часть, тысячи страниц.
- В DICOM стандартизировано ВСЁ. От определений до сетевого взаимодействия.
- Реализации поддерживают лишь части какой-то версии стандарта.

<https://www.dicomstandard.org/current/>

DICOM Information Model

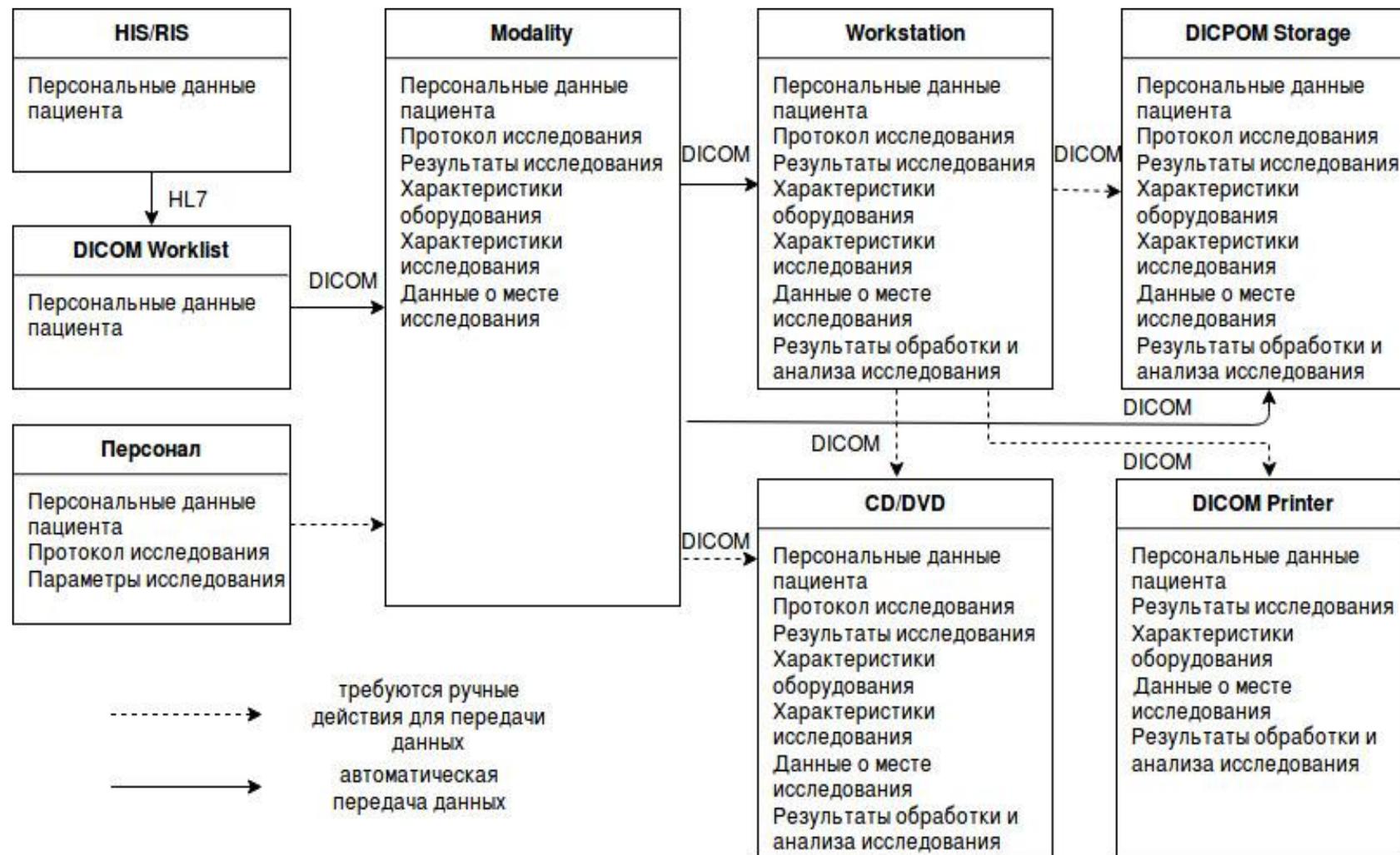
dicom PS3.3 2018d
Figure 6-1 (p93)
Major Structures



Существующая ситуация

- Используется ручная выгрузка данных из MRI (Magnetic Resonance Imaging) и CT (Computer Tomography).
- Возможна лишь ручная деперсонафикация данных.
- Используемые процессы совершенно не масштабируются.
- Осуществляется выборочное хранение данных, имеющаяся система обладает недостаточным объёмом хранения.
- Интеграция информационных потоков в единую систему производится при помощи ручных процедур.

Исходная схема потоков данных



Как сделать ещё лучше?

- **Переделать HIS**
Закупка и введение в эксплуатацию новой HIS.
- **RIS - Радиологическая Информационная Система**
Переделка информационных систем для каждого прибора.
- **DICOM proxy**
Дополнительная система в разрыв информационных потоков.

При выборе вариантов решения необходимо обеспечить постоянную работоспособность имеющейся HIS.

Вариант: “Переделатъ HIS”

- Высокая стоимость закупки новой HIS.
- Замена HIS затруднена административно.
- Потребуется замена не поддерживаемого оборудования.
- Большой объем тестирования (на живых людях?).
- Высокая трудоёмкость ввода новой системы в эксплуатацию.

Этот вариант практически нереализуем в работающей крупной медицинской организации. Для замены HIS необходима остановка деятельности клиники и переучивание персонала.

Вариант: “RIS”

- Простота в интеграции по сравнению с HIS
- Всё равно дорого
- Потребуется замена не поддерживаемого оборудования.
- Большой объем тестирования (на живых людях?).
- Высокая трудоёмкость ввода новой системы в эксплуатацию.
- Проблема тестирования на эксплуатируемом оборудовании.

Этот вариант возможен, но обладает высокой трудоёмкостью и требует постоянной поддержки при эксплуатации. Он реализуем на уровне компании, занимающейся техподдержкой RIS по контракту.

Вариант: “DICOM proxy”

- Дополнительная вычислительная система, встраиваемая между источниками данных и старым DICOM PACS.
- Используется существующая введённая в эксплуатацию HIS.
- Управление не на уровне приборов, а потоков данных.
- Требуется достаточная квалификация разработчиков в различных областях (it, медицина, управление данными, микробиология).

Этот вариант наиболее консервативен, так как не вносит эксплуатационных изменений в эксплуатируемую HIS.

Именно он использован в данной работе

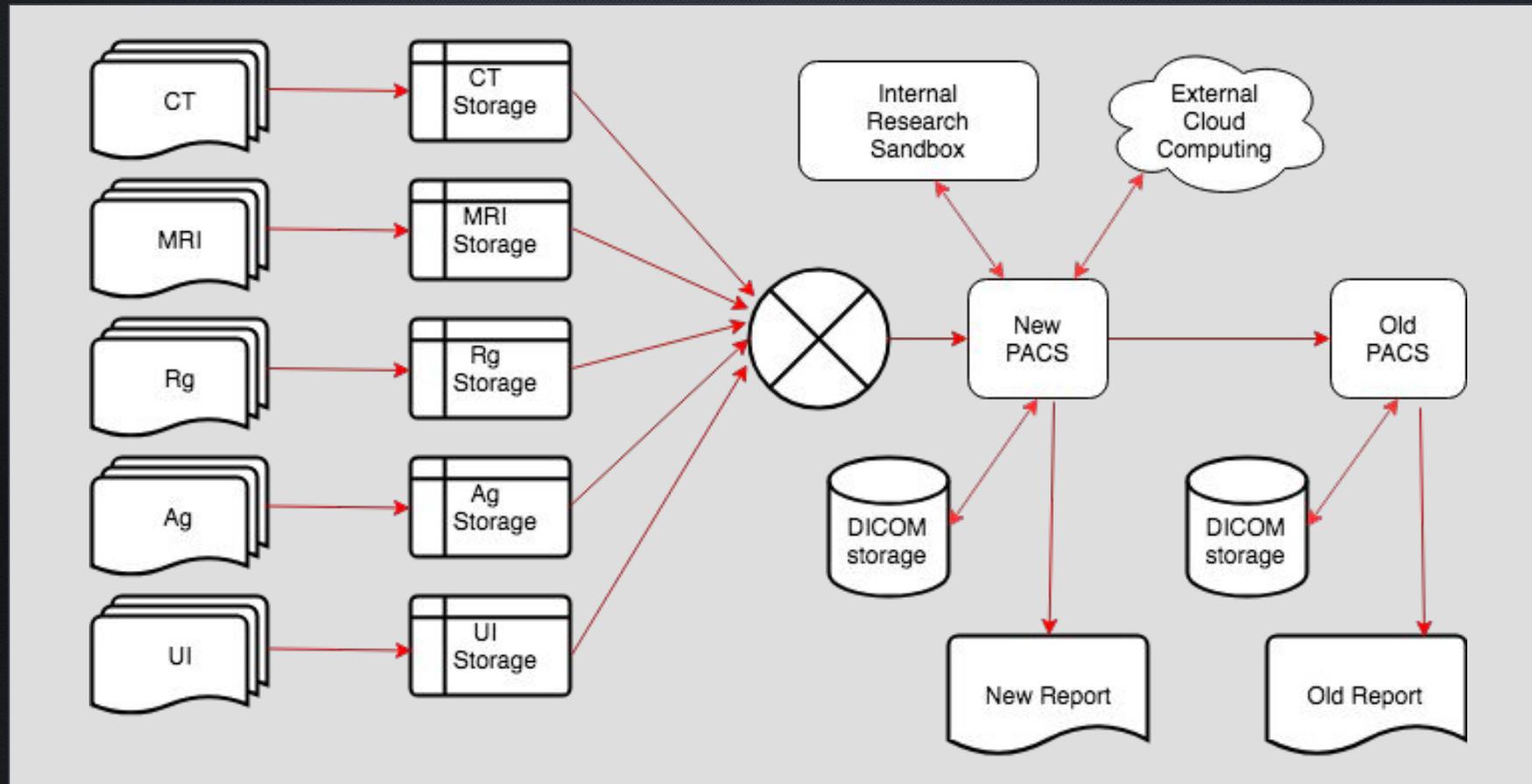
Функции DICOM proxy

- Выделение данных для обработки.
- Прозрачная передача остальных данных.
- Разветвление потоков данных.
- Хранение данных.
- Деперсонификация данных.
- Интеграция с research sandbox
(песочница для отладки алгоритмов обработки данных).
- Унификация данных.

Унификация данных

- Перекодировки (например, ломаный CP1251 - UTF8).
- Единая система измерений (например, $\mu\text{mol/L}$ и mmol/L в mol/L).
- Единые нормативные диапазоны результатов лабораторного анализа КДЛ (клинико-диагностической лаборатории), исходя из приказов минздрава, а не из внутренних соображений изготовителя прибора.
- Использование для идентификации пациентов единых идентификаторов, деперсонализирующих данные.

Схема интеграции DICOM proxy

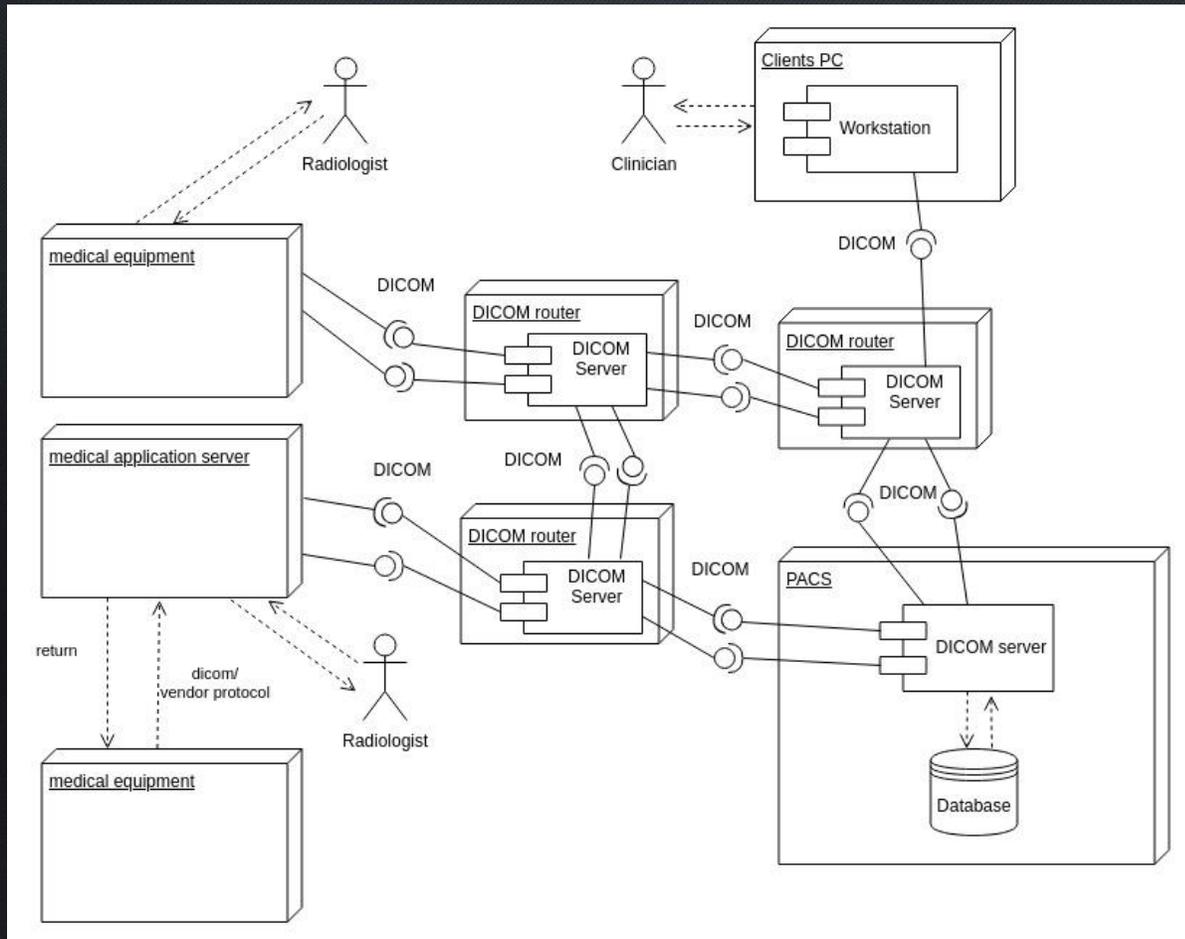


Промежуточные результаты

- Сформирована рабочая группа для выполнения НИР в Институте им. В.А. Алмазова.
- Защищены две магистерских работы в Университете ИТМО в 2018 г.*
- Прошли апробацию результаты выполнения работ по распознаванию гематом на КТ-снимках.
- Реализована первая версия прототипа подсистемы автоматической выгрузки данных в формате DICOM для исследовательских целей (DICOM serverConquest на C++ и скрипт на Lua).
- Подготовлен план-график работ по реализации системы получения и хранения данных из PACS в исследовательских целях.

* https://github.com/iradche/itmo-students-misc/blob/master/2018/226372_mag1_18062018155231.pdf
https://github.com/iradche/itmo-students-misc/blob/master/2018/226370_mag5_18062018155912.pdf

Первый прототип DICOM proxy



Ближайшие планы

- Завершить создание прототипа DICOM proxy.
- Анонимизированные данные перемещать в research sandbox.
- На базе research sandbox создать систему проверки гипотез с использованием нейросетей.

Выводы

- Решение поставленной задачи имеет высокую административную, техническую и научную сложность.
- Проводимая работа с существующими диагностическими данными получила положительное заключение врача-клинициста.
- Появилась возможность организовать новый PACS для DICOM с деперсонализированными сырыми диагностическими данными, пригодными для обработки нейронными сетями.

Контакты

- Ирина Радченко
- Email: iradche@gmail.com
- +7-921-865-26-64
- @iRadche
- <http://iRadche.ru>

