Design and implementation of a generic DICOM archive for clinical and pre-clinical research

J. Lamy, R. Lahaxe, J.-P. Armspach, F. Heitz
Contact: lamy@unistra.fr

ICube, University of Strasbourg-CNRS

Introduction

Clinical trials versus clinical practice

- Unrelated patients vs. groups of subjects
  - Patient vs. subject: de-identification
  - Group: e.g. based on demography, pathology or treatment
- Single institution vs. multiple sites
  - Data gathering (CD-ROM, SFTP, DICOM, Shanoir, Keosys, etc.)
  - Harmonization of acquisition parameters
  - Normalization of data (name of entities, e.g. subjects and series)
- Longitudinal studies: time points
  - Fixed in time (e.g. baseline, 6 months after start of treatment)
  - Event-based (e.g. conversion of CIS to MS)

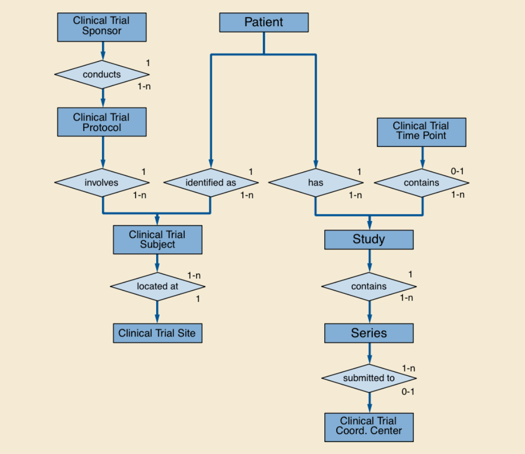
Introduction

Medical imaging on the data side

- Technically, one standard: DICOM
  - File format
  - Data model
  - Query and transfer protocol
  - De-identification, worklist, printing, etc.

Introduction

DICOM model for clinical trials



Introduction

In real-life™: DICOM is not everywhere

- Not every application speaks DICOM
  - Bruker scanners output mostly Bruker format
  - Image processing packages input other formats (NIfTI, NRRD, MHA, etc.)
- Not every DICOM application speaks the same dialect
  - Vendors use and sometimes abuse private elements
  - Philips scanners: TS changes for private elements
  - Bruker: some elements are missing (DTI)

Introduction

In real-life™: guidelines are not always followed

- Not every acquisition conforms to the guidelines
  - Technicians mis-type subject IDs (e.g. `AWE2345_01-4321_W40`)
  - Subjects not undergoing the full acquisition
  - Series performed multiple times (e.g. subject motion)
  - Local naming conventions are preferred over study guidelines (e.g. study with 29 centers, 58 different names for a T1 3D)
- Incompatible constraints (clinical practice vs anonymization)
  - Modalities require a birth date for SAR computation, anonymized version is age-only
  - Anonymizers rename every series to “`anonymized`”

Introduction

Challenges created by clinical trials

- Import/export to and from DICOM
  - Meta-data alignment from Bruker, NIfTI, etc.
- Post-hoc modification of data
  - External to internal data alignment, retain original information
- User-defined queries on standard and private fields
  - Storage of every element in SQL tables borders on impossible
  - Switch to document-oriented model
- Fine-grained access rights
  - Authentify users, customize access
    (e.g. publicize existence of data, but restrict access)

Introduction

Our contributions

- Routing
  - Convert to and from DICOM, align data sets to a model
  - Easily customizable for local constraints
- Storage
  - Store any current DICOM document, be future-proof
  - Scalable: new modalities mean more data (fMRI, DTI)
  - Fine grain access rights: easily give and revoke rights

Routing

Basic concepts

- All file formats can be represented as a list of named fields
- Intra-format routing is a list of rules based on the field names
  - Transform Patient's Birth Date and Study Date to Patient's Age
- Inter-format routing is a mapping between field names
  - Bruker's VisuStudyId corresponds to DICOM's Study Description

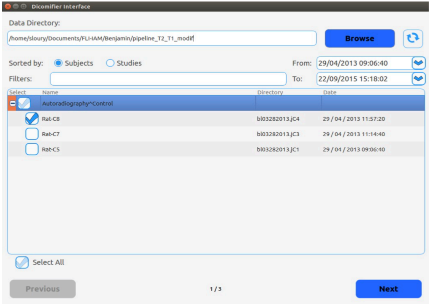
Implementation

- Paper mentions XML: impractical, replaced by Javascript
- Based on V8: functions can be wrapped from C++ to JS (e.g. input/output, handling of binary data)

Routing

Available software: Dicomifier

- GUI to convert from Bruker to DICOM
- CLI to convert from Bruker to DICOM and from DICOM to NIfTI
- Full Javascript/C++ access for the tinkerers
- Available on GitHub: https://github.com/lamyj/dicomifier



Data Directory: /home/sloury/Documents/FLI-IAM/Benjamin/pipeline_T2_T1_modif

Sorted by: ● Subjects ○ Studies

From: 29/04/2013 09:06:40

Filters:

To: 22/09/2015 15:18:02

Select	Name	Directory	Date
☑	Autoradiography^Control		
☑	Rat-C8	bl03282013.jC4	29 / 04 / 2013 11:57:20
□	Rat-C7	bl03282013.jC3	29 / 04 / 2013 11:14:40
□	Rat-C5	bl03282013.jC1	29 / 04 / 2013 09:06:40

☑ Select All

Routing

Example: DICOM to DICOM

```
// readDICOM: C++ function wrapped in JS
var dataSet = readDICOM('foo.dcm');
var birthDate = dataSet['00100030'].Value[0];
var studyDate = dataSet['00080020'].Value[0];
dataSet['00101010'] = {
    'vr': 'AS', Value = [getAge(birthDate, studyDate)] };
delete dataSet['00100030'];
writeDICOM(dataSet, 'bar.dcm');
```

Routing

Example: DICOM to NIfTI

```
var dicomDataSets = [
    readDicom('foo.dcm'), readDicom('bar.dcm')];
// dicomifier.dicom2nifti.convert: mostly pure Javascript
var niftiDataSet = dicomifier.dicom2nifti.convert(
    dicomDataSets, './');
```

Storage

Requirements

- Store any DICOM file
  (human/animals/samples, modalities, vendors, etc.)
- Be future-proof: accept unknown modalities or elements
- Be scalable: data repository can grow quickly
- Allow queries on arbitrary fields
- Fine-grained access control

Implementation

- SQL is out of the way: entity-relationship model is too complex
- Use document-oriented database (MongoDB)

Storage

Storing DICOM data sets in MongoDB

- MongoDB documents are in BSON, a JSON-like format: direct use of the JSON model of DICOM
- Original file stored in GridFS (MongoDB storage for binary data)

Features

- No rigid model: store arbitrary DICOM files, even future ones
- MongoDB was built for big data: scalability is included
- Every field can be queried, or processed through map/reduce

Storage

Access control

- Three orthogonal criteria
  - The action performed (echo, store, query, retrieve)
  - Who performs the action (user or service)
  - What is transmitted (subject's id, study name, modality, etc.)
- Stored in a MongoDB collection
- Modular authentication backend (file, LDAP, Kerberos, etc.)

Storage

Access control

- Everybody can perform an echo request:

```
{ "service": "Echo", "user": "*" }
```

- Everybody can get information about some study:

```
{ "service": "Query",
  "user": "*",
  "dataset": {"00081030": {"vr": "LO", "Value": ["Some study"]}}}
```

- A subset of users can retrieve the corresponding datasets:

```
{ "service": "Retrieve",
  "user": "myself",
  "dataset": {"00081030": {"vr": "LO", "value": ["Some study"]}}}
```

Storage

Available software: Dopamine

- Available on GitHub: https://github.com/lamyj/dopamine
- In C++
- Still beta: for the braves only

Conclusion

- Routing and storage software available on GitHub
  - Free software (CeCILL-B license, cf. MIT or LGPL)
  - Dicomifier is mostly stable, Dopamine still beta
  - Testers and contributors are welcome!
- Perspectives
  - Specify new DICOM services: modification and deletion
  - Last piece of the puzzle for a DICOM-based image-processing platform: software pipelines

DICOM archive for clinical & pre-clinical research – J. Lamy et al. Conclusion 18 / 18