BioSignalML
Putting biosignals onto the Semantic Web

David Brooks
Background

- Biosignal -- time series data resulting from a biological process.
Background

- Biosignal -- time series data resulting from a biological process.
- Sampled, usually at a regular rate, which is usually much greater than the highest frequency of interest.
Background

- Biosignal -- time series data resulting from a biological process.
- Sampled, usually at a regular rate, which is usually much greater than the highest frequency of interest.
- Electrical, pressure, concentration, ...
Background

- **Biosignal** -- time series data resulting from a biological process.
- Sampled, usually at a regular rate, which is usually much greater than the highest frequency of interest.
- Electrical, pressure, concentration, ...
- Simulation time series data.
Background

- A lot of file formats:
  - manufacturer; research; regulatory; ...

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>24 bit version of EDF</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EDF+</td>
<td>European Data Format plus</td>
</tr>
<tr>
<td>FDAXML</td>
<td>FDA standard for ECG</td>
</tr>
<tr>
<td>GDF</td>
<td>General Data Format (an EDF derivative)</td>
</tr>
<tr>
<td>MFER</td>
<td>Medical waveform Format Encoding Rules (ISO)</td>
</tr>
<tr>
<td>SCP</td>
<td>Standard Communication Protocol for ECG (CEN)</td>
</tr>
<tr>
<td>WFDB</td>
<td>WaveForm DataBase</td>
</tr>
</tbody>
</table>
Background

- A lot of file formats:
  - manufacturer; research; regulatory; ...
- Often developed for a specific problem domain.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>24 bit version of EDF</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EDF+</td>
<td>European Data Format plus</td>
</tr>
<tr>
<td>FDAXML</td>
<td>FDA standard for ECG</td>
</tr>
<tr>
<td>GDF</td>
<td>General Data Format (an EDF derivative)</td>
</tr>
<tr>
<td>MFER</td>
<td>Medical waveform Format Encoding Rules (ISO)</td>
</tr>
<tr>
<td>SCP</td>
<td>Standard Communication Protocol for ECG (CEN)</td>
</tr>
<tr>
<td>WFDB</td>
<td>WaveForm DataBase</td>
</tr>
</tbody>
</table>
Background

- A lot of file formats:
  - manufacturer; research; regulatory; ...
- Often developed for a specific problem domain.
- All generally good at storing time series data.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>24 bit version of EDF</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EDF+</td>
<td>European Data Format plus</td>
</tr>
<tr>
<td>FDAXML</td>
<td>FDA standard for ECG</td>
</tr>
<tr>
<td>GDF</td>
<td>General Data Format (an EDF derivative)</td>
</tr>
<tr>
<td>MFER</td>
<td>Medical waveform Format Encoding Rules (ISO)</td>
</tr>
<tr>
<td>SCP</td>
<td>Standard Communication Protocol for ECG (CEN)</td>
</tr>
<tr>
<td>WFDB</td>
<td>WaveForm DataBase</td>
</tr>
</tbody>
</table>
Background

- A lot of file formats:
  - manufacturer; research; regulatory; ...
- Often developed for a specific problem domain.
- All generally good at storing time series data.
- Metadata format is file specific.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>24 bit version of EDF</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EDF+</td>
<td>European Data Format plus</td>
</tr>
<tr>
<td>FDAXML</td>
<td>FDA standard for ECG</td>
</tr>
<tr>
<td>GDF</td>
<td>General Data Format (an EDF derivative)</td>
</tr>
<tr>
<td>MFER</td>
<td>Medical waveform Format Encoding Rules (ISO)</td>
</tr>
<tr>
<td>SCP</td>
<td>Standard Communication Protocol for ECG (CEN)</td>
</tr>
<tr>
<td>WFDB</td>
<td>WaveForm DataBase</td>
</tr>
</tbody>
</table>
Background

- A lot of file formats:
  - manufacturer; research; regulatory; ...
- Often developed for a specific problem domain.
- All generally good at storing time series data.
- Metadata format is file specific.
- Metadata content tends to be domain specific.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>24 bit version of EDF</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EDF+</td>
<td>European Data Format plus</td>
</tr>
<tr>
<td>FDAXML</td>
<td>FDA standard for ECG</td>
</tr>
<tr>
<td>GDF</td>
<td>General Data Format (an EDF derivative)</td>
</tr>
<tr>
<td>MFER</td>
<td>Medical waveform Format Encoding Rules (ISO)</td>
</tr>
<tr>
<td>SCP</td>
<td>Standard Communication Protocol for ECG (CEN)</td>
</tr>
<tr>
<td>WFDB</td>
<td>WaveForm DataBase</td>
</tr>
</tbody>
</table>
Difficulties

- Polysomnography:
  - “Currently, digital data from most PSG systems can only be viewed if one utilizes the system with which it was collected.” [1]
  - “Unfortunately, not much has happened since … no consensus for data sharing has taken root.” [2]


Difficulties

• Polysomnography:
  – “Currently, digital data from most PSG systems can only be viewed if one utilizes the system with which it was collected.” [1]
  – “Unfortunately, not much has happened since … no consensus for data sharing has taken root.” [2]

• Metadata terms:
  – Different groups may have different meanings for a term.


Difficulties

- Polysomnography:
  - “Currently, digital data from most PSG systems can only be viewed if one utilizes the system with which it was collected.” [1]
  - “Unfortunately, not much has happened since … no consensus for data sharing has taken root.” [2]

- Metadata terms:
  - Different groups may have different meanings for a term.
  - Units: µV, uV, V×10⁻⁶ ??


Semantic Web

- Web content that is meaningful to computers.
  - Knowledge representation, ontologies, reasoning, intelligent agents, ...
Semantic Web

- Web content that is meaningful to computers.
  - Knowledge representation, ontologies, reasoning, intelligent agents, ...
- [http://www.w3.org/standards/semanticweb/](http://www.w3.org/standards/semanticweb/)
  - Resource Description Framework (RDF)
  - RDFS, OWL, SPARQL, ...

[Diagram of subject, predicate, object relationship]
Semantic Web

- Web content that is meaningful to computers.
  - Knowledge representation, ontologies, reasoning, intelligent agents, ...
- [http://www.w3.org/standards/semanticweb/](http://www.w3.org/standards/semanticweb/)
  - Resource Description Framework (RDF)
  - RDFS, OWL, SPARQL, ...
- Linking Open Data

[Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch.](http://lod-cloud.net/)
BioSignalML

- Abstract common elements of storage formats.
BioSignalML

- Abstract common elements of storage formats.
- Use Semantic Web standards/technologies.
  - Objects have web identifiers.
  - Ontologies define terms, properties, relationships.
BioSignalML

- Abstract common elements of storage formats.
- Use Semantic Web standards/technologies.
  - Objects have web identifiers.
  - Ontologies define terms, properties, relationships.
- Time series data is in native format; everything else is available as RDF metadata.

http://repository.biosignal.org/recording3/signal/4
BioSignalML as RDF

- Core concepts:
  - Recordings
  - Signals
  - Events and Annotations.
BioSignalML as RDF

• Core concepts:
  – Recordings
  – Signals
  – Events and Annotations.

• RDF graph:
BioSignalML as an ontology

- Classes, terms, properties, relationships:
BioSignalML as an ontology

- Classes, terms, properties, relationships:
BioSignalML implementation

- Biosignal repository:
BioSignalML implementation

- Biosignal repository:
- Web based with HTTP endpoints:
  - File import/export
  - RDF metadata
  - Data streamed via web-sockets.
BioSignalML implementation

- Biosignal repository:
- Web based with HTTP endpoints:
  - File import/export
  - RDF metadata
  - Data streamed via web-sockets.
- C client (plus Python, Javascript, ...)

Internet
Web Browsers
Applications and Tools

RESTful Web Services

Abstraction Layer

Signal Recordings
Metadata

EDF
WFDB
HDF5

Triple Store

Python API

SPARQL Query
BioSignalML clients

- Web browser:
BioSignalML clients

- Web browser:

- RDF browser:
BioSignalML clients

- Python code:

```python
import biosignalml
import biosignalml.units as units

repo = biosignalml.Repository('http://demo.biosignalml.org/
rec = repo.new_recording('http://example.org/recording/test/
sig = rec.new_signal(id='a1', units=units.millivolt)
for data in datasource:
    sig.append(data)
rec.close()

sig = repo.get_signal('http://example.org/recording/test/signal/a1')
print sig.uri, sig.label, sig.units

start = 0.0
end = 10.0
duration = 1.0
while start < end:
    interval = sig.recording.interval(start, duration)
    for data in sig.read(interval):
        print data  # SignalSegment
    start += duration
```
BioSignalML clients

- CellML modelling:
BioSignalML clients

- CellML modelling:
Ongoing work

• Interfacing with simulation tools (OpenCOR, SED/ML)
  - real world applications.
• Adding a Semantic Web layer to PhysioBank.
• Integrate Units of Measurement Expressions:
  - [http://www.sbpax.org/uome/index.html](http://www.sbpax.org/uome/index.html)
  - Ontology to derive units from other units.
  - An extensible way to automate units validation and conversion.
Thank you