BioSignalML

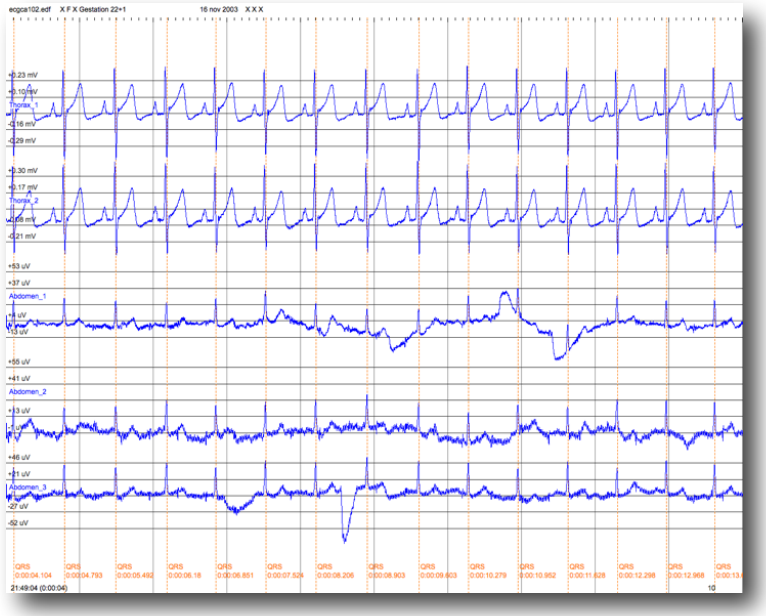
Putting biosignals onto the Semantic Web

David Brooks

AUCKLAND BIOENGINEERING INSTITUTE
THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

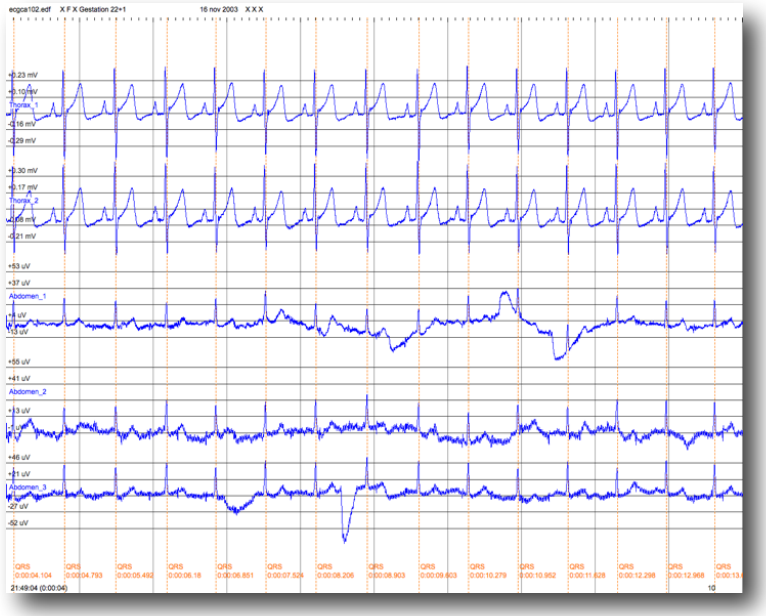
Background

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



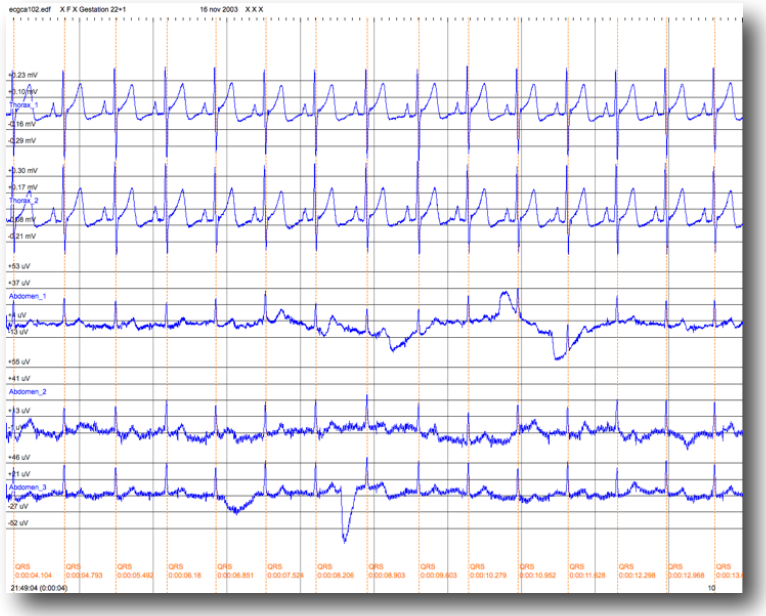
Background

- 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.



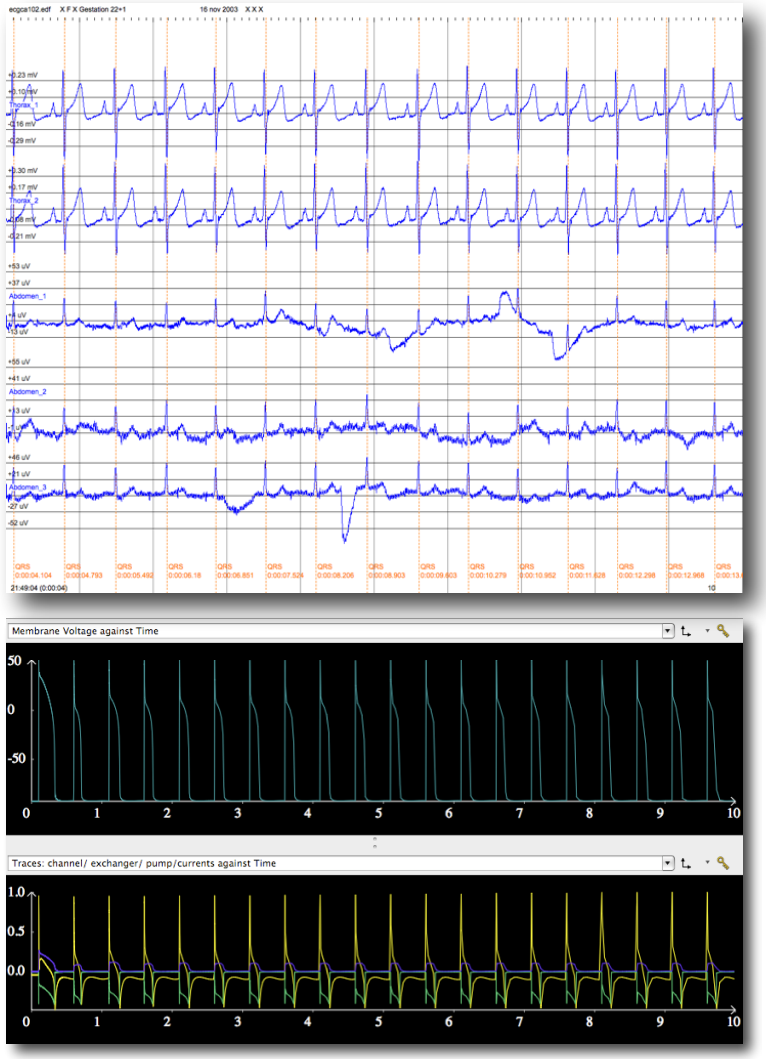
Background

- 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; …

BDF	24 bit version of EDF
EDF	European Data Format
EDF+	European Data Format plus
FDAXML	FDA standard for ECG
GDF	General Data Format (an EDF derivative)
MFER	Medical waveform Format Encoding Rules (ISO)
SCP	Standard Communication Protocol for ECG (CEN)
WFDB	WaveForm DataBase

Background

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

BDF	24 bit version of EDF
EDF	European Data Format
EDF+	European Data Format plus
FDAXML	FDA standard for ECG
GDF	General Data Format (an EDF derivative)
MFER	Medical waveform Format Encoding Rules (ISO)
SCP	Standard Communication Protocol for ECG (CEN)
WFDB	WaveForm DataBase

Background

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

BDF	24 bit version of EDF
EDF	European Data Format
EDF+	European Data Format plus
FDAXML	FDA standard for ECG
GDF	General Data Format (an EDF derivative)
MFER	Medical waveform Format Encoding Rules (ISO)
SCP	Standard Communication Protocol for ECG (CEN)
WFDB	WaveForm DataBase

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.

BDF	24 bit version of EDF
EDF	European Data Format
EDF+	European Data Format plus
FDAXML	FDA standard for ECG
GDF	General Data Format (an EDF derivative)
MFER	Medical waveform Format Encoding Rules (ISO)
SCP	Standard Communication Protocol for ECG (CEN)
WFDB	WaveForm DataBase

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.

BDF	24 bit version of EDF
EDF	European Data Format
EDF+	European Data Format plus
FDAXML	FDA standard for ECG
GDF	General Data Format (an EDF derivative)
MFER	Medical waveform Format Encoding Rules (ISO)
SCP	Standard Communication Protocol for ECG (CEN)
WFDB	WaveForm DataBase

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]

[1] D. Rapoport, I. Ayappa, R. Norman, and S. Herman, “NPSG data interchange-dealing with the Tower of Babel.” Sleep, vol. 29, no. 5, p. 599, 2006.

[2] D. M. Rapoport, email correspondence, November 2011.

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.

[1] D. Rapoport, I. Ayappa, R. Norman, and S. Herman, "NPSG data interchange-dealing with the Tower of Babel." Sleep, vol. 29, no. 5, p. 599, 2006.

[2] D. M. Rapoport, email correspondence, November 2011.

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, $V \times 10^{-6}$??

[1] D. Rapoport, I. Ayappa, R. Norman, and S. Herman, "NPSG data interchange-dealing with the Tower of Babel." Sleep, vol. 29, no. 5, p. 599, 2006.

[2] D. M. Rapoport, email correspondence, November 2011.

Semantic Web

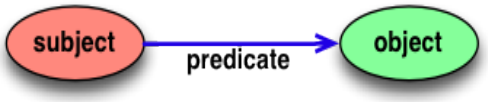
- Web content that is meaningful to computers.
  - Knowledge representation, ontologies, reasoning, intelligent agents, …

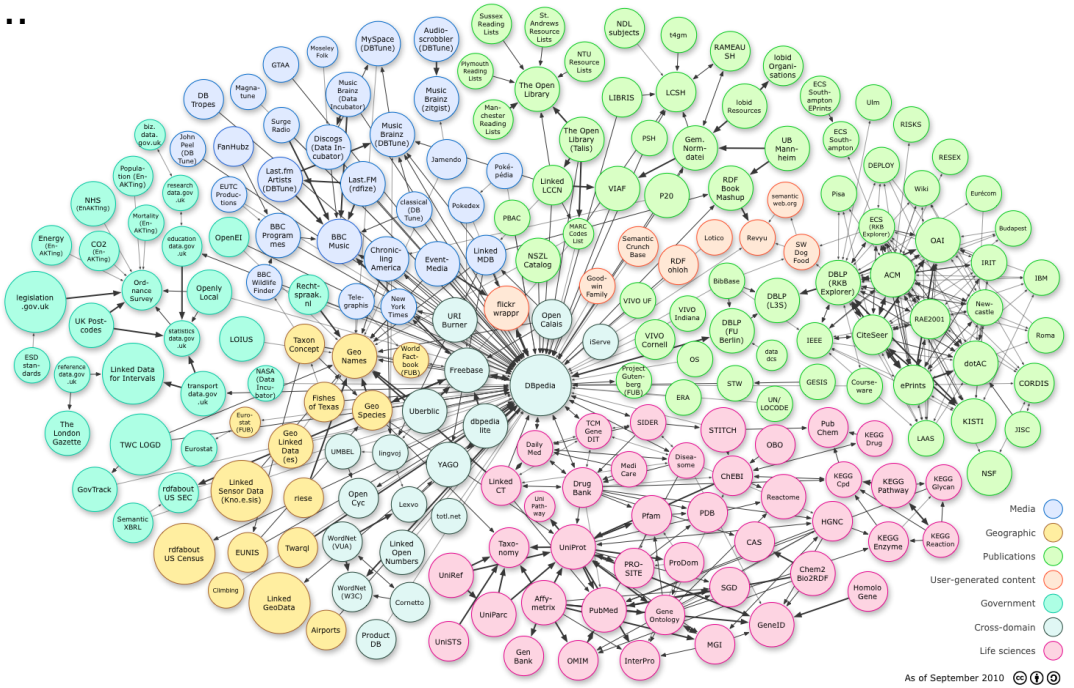
Semantic Web

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

Semantic Web

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

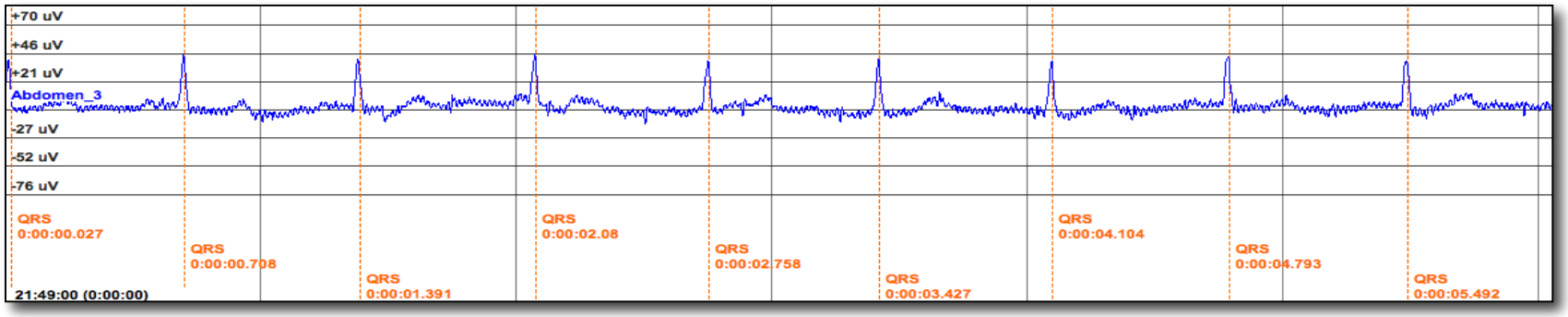
BioSignalML

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

BioSignalML

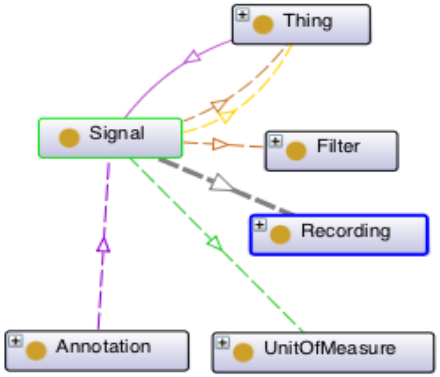
- Abstract common elements of storage formats.
- Use Semantic Web standards/technologies.
  - Objects have web identifiers.
  - Ontologies define terms, properties, relationships.
- Time series data is in native format; everything else is available as RDF metadata.

http://repository.biosignal.org/recording3/signal/4



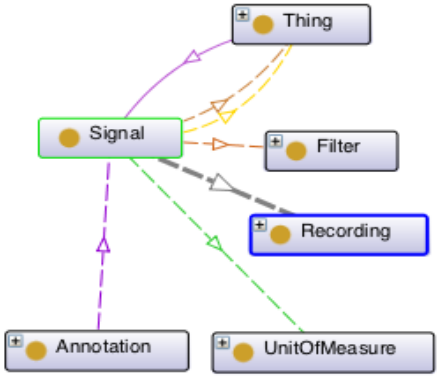
BioSignalML as RDF

- Core concepts:
  - Recordings
  - Signals
  - Events and Annotations.

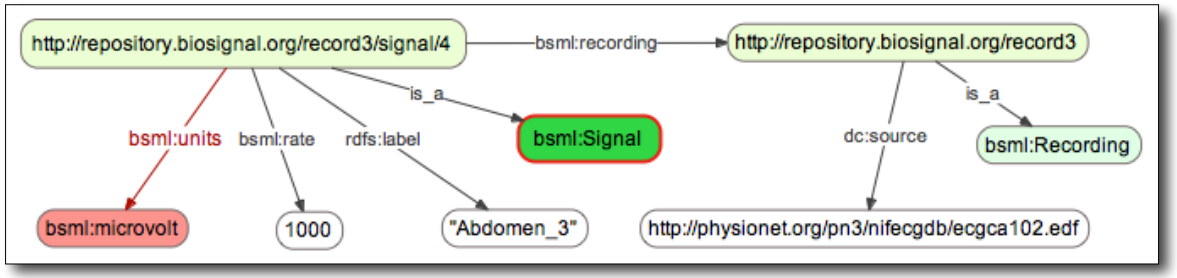


BioSignalML as RDF

- Core concepts:
  - Recordings
  - Signals
  - Events and Annotations.

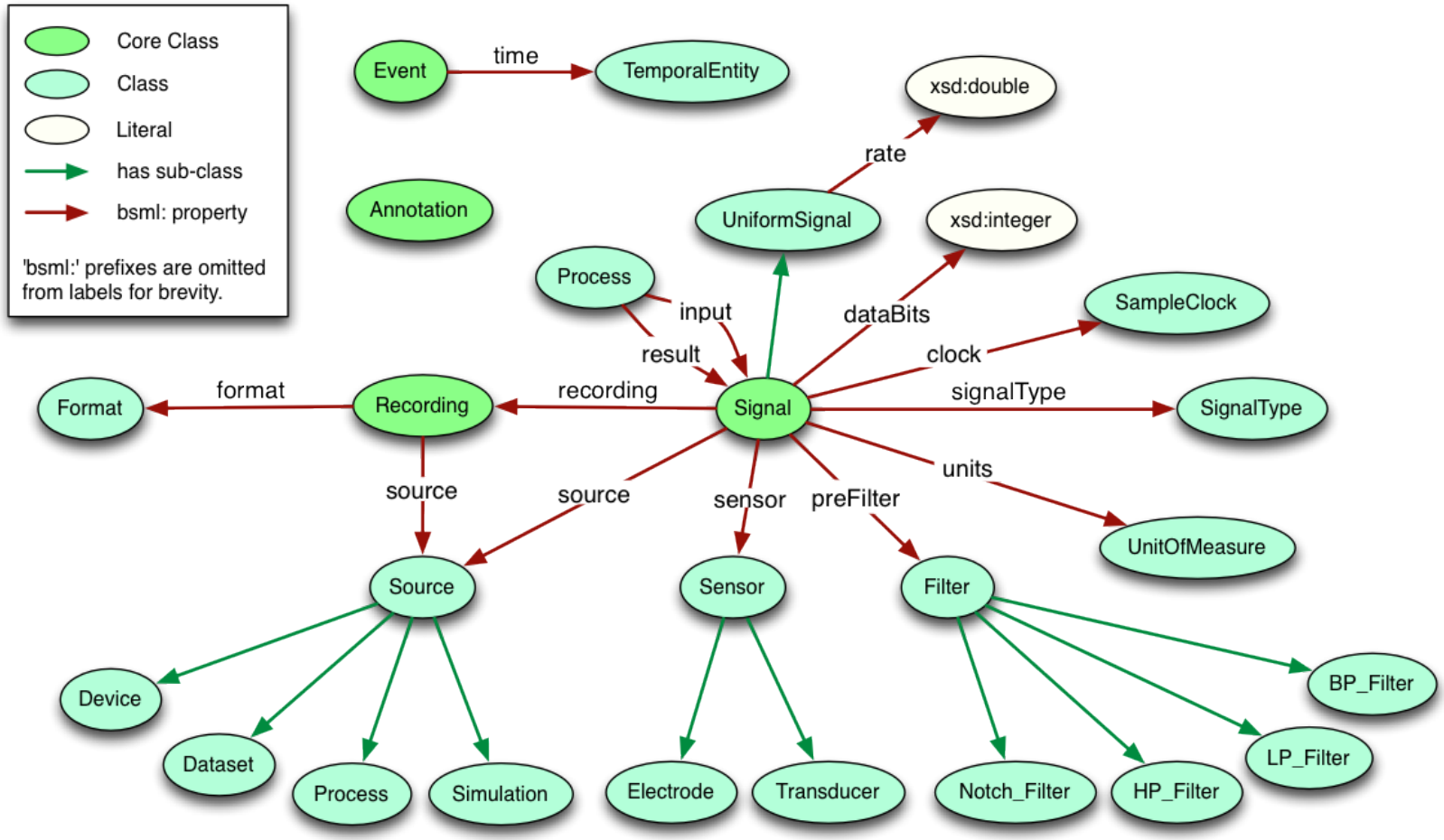


- RDF graph:



BioSignalML as an ontology

- Classes, terms, properties, relationships:

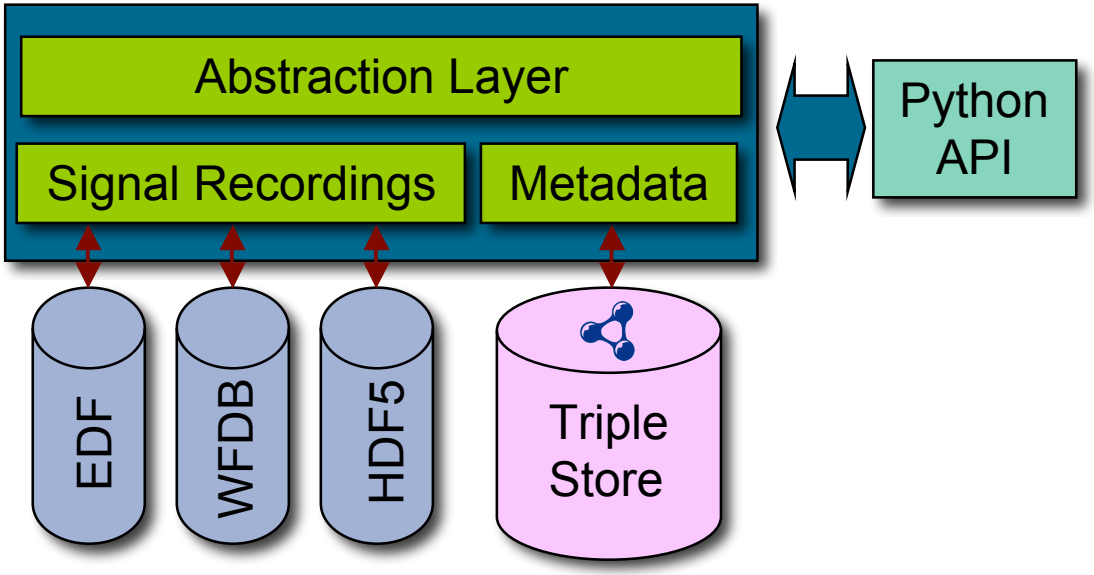


BioSignalML as an ontology

- Classes, terms, properties, relationships:

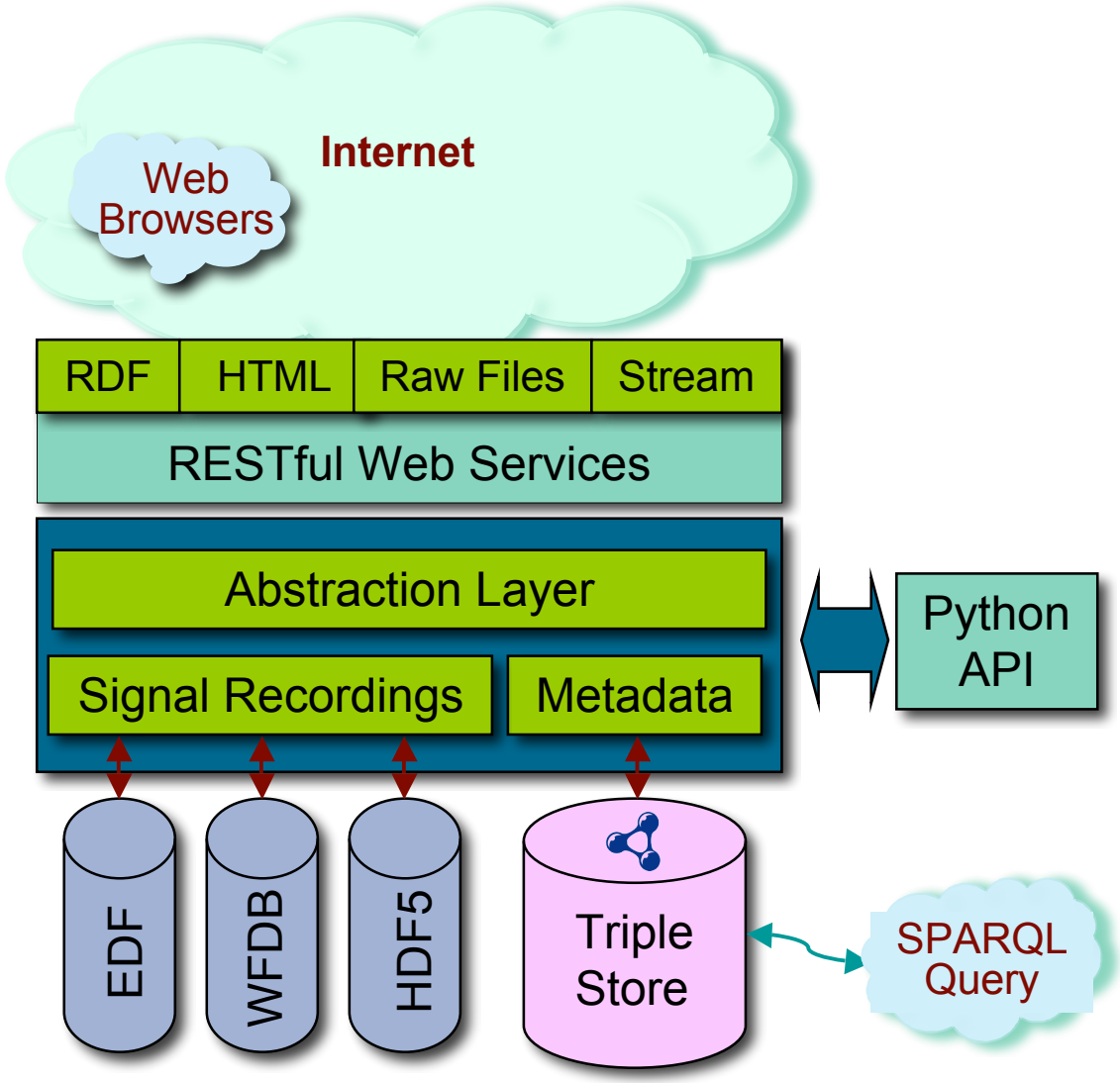
BioSignalML implementation

- Biosignal repository:



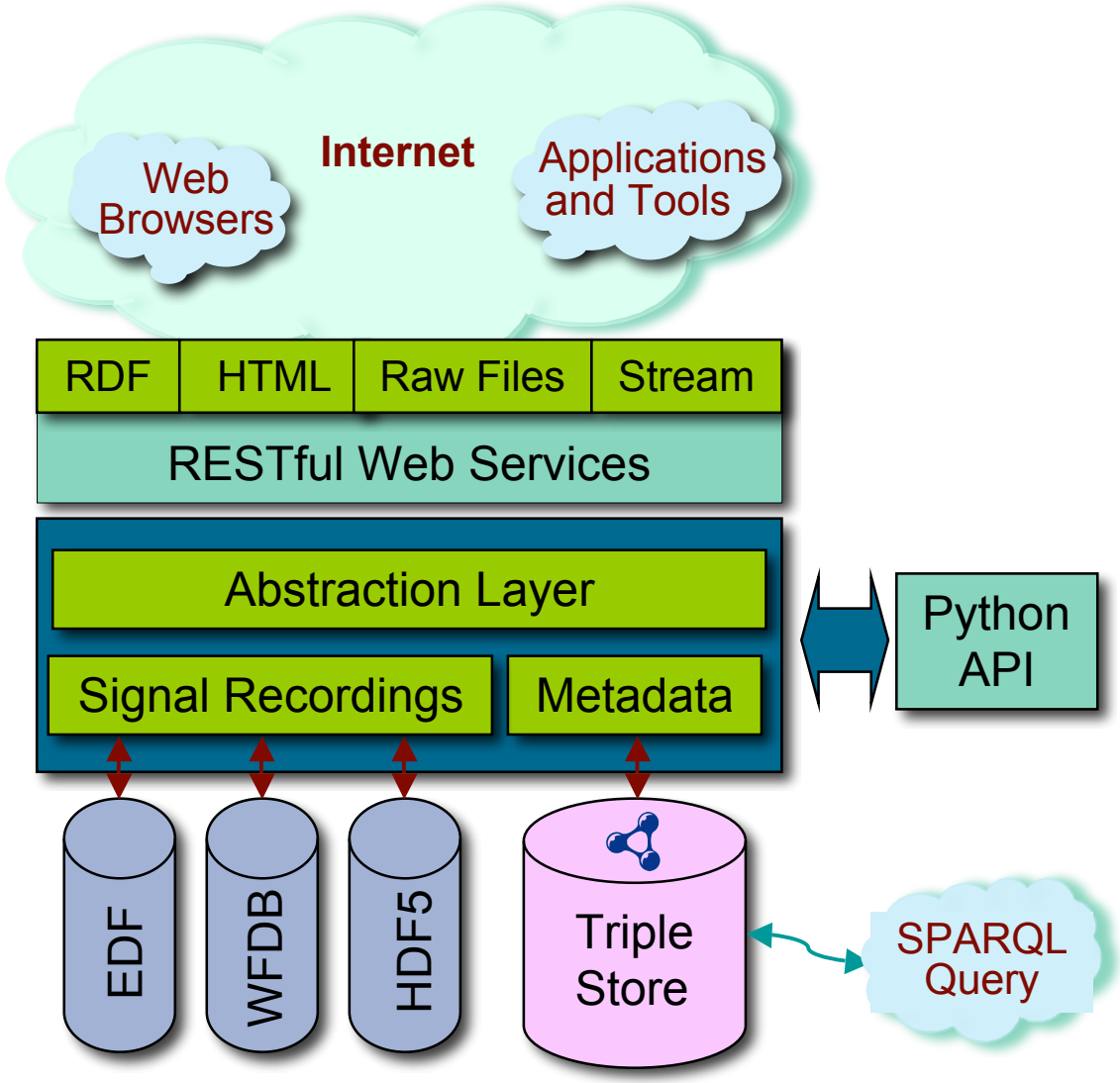
BioSignalML implementation

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



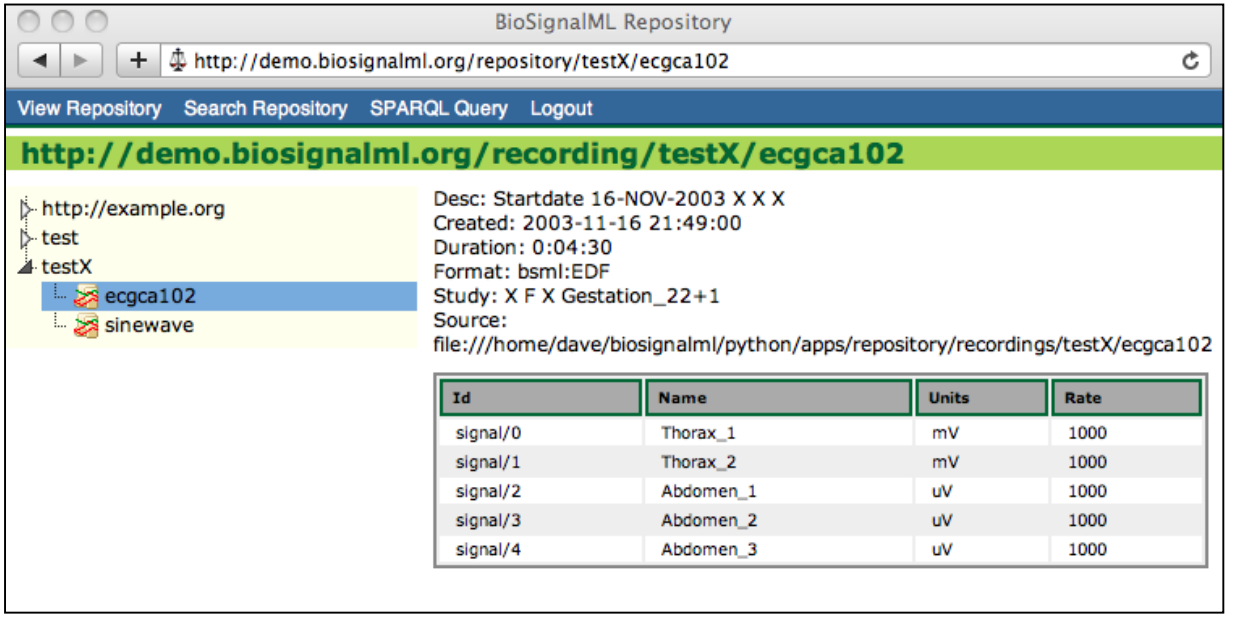
BioSignalML implementation

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



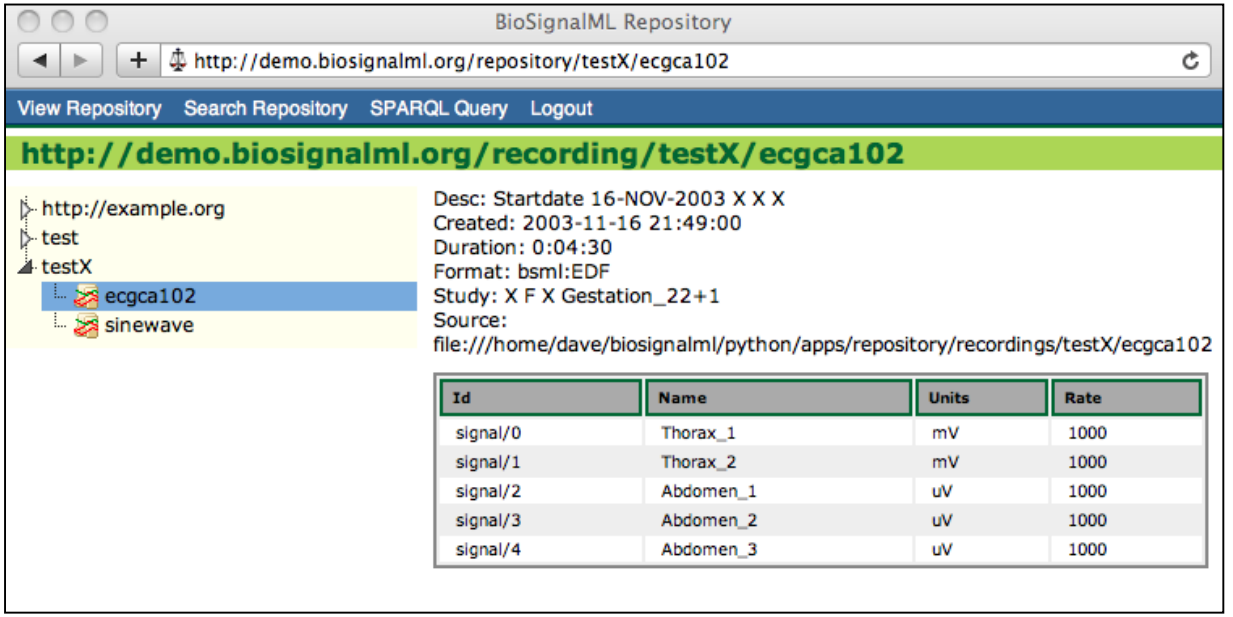
BioSignalML clients

- Web browser:

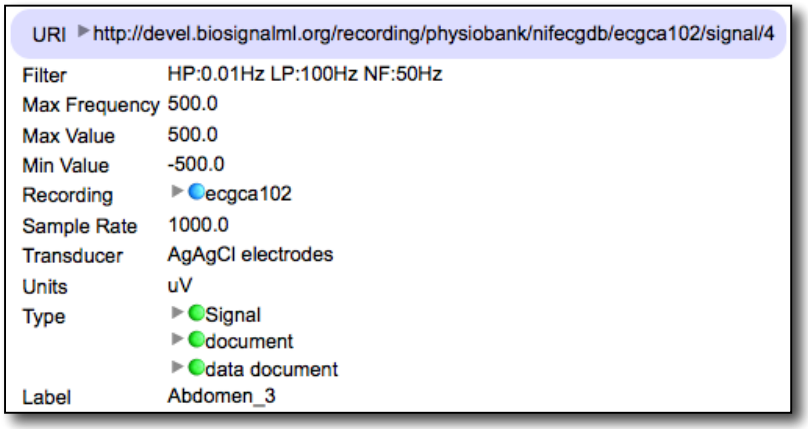


BioSignalML clients

- Web browser:



- RDF browser:



BioSignalML clients

- Python code:

```
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:

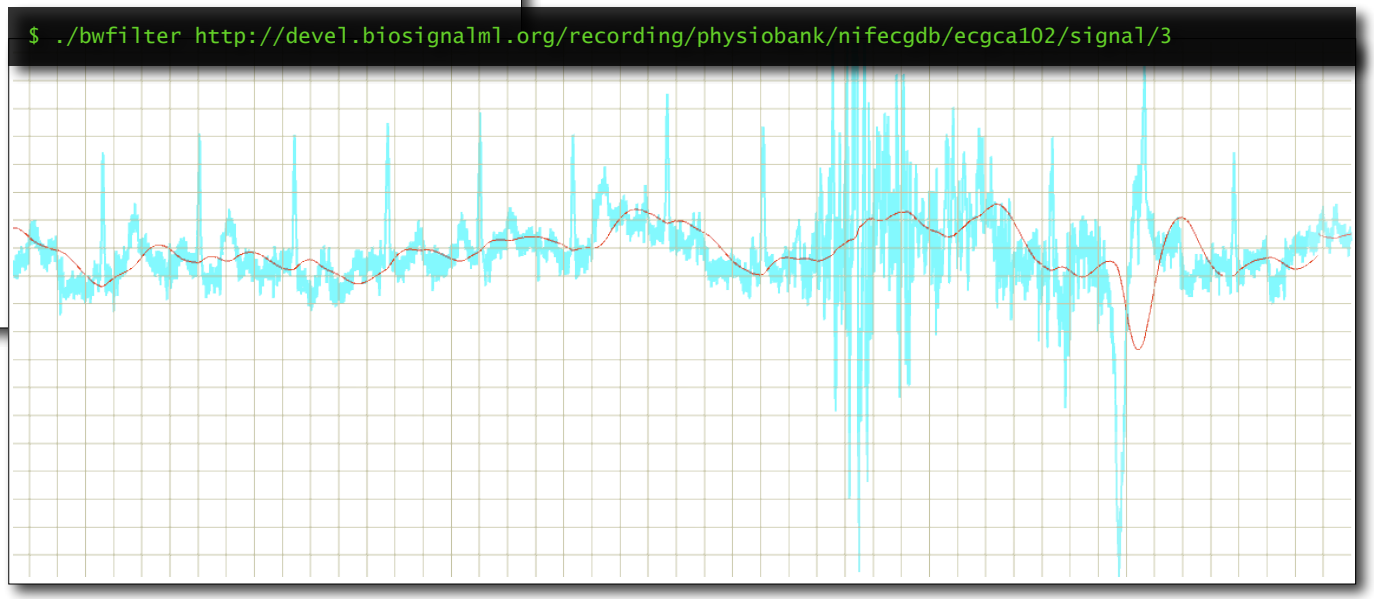
```
<model name="bwfilter" cmeta:id="bwfilter" xmlns="http://www.cellml.org/cellml/
   <component name="filter">
      <variable cmeta:id="time" name="t" units="second"/>
      <variable cmeta:id="input" name="v_i" units="volt" initial_value="0"/>
      <variable cmeta:id="output" name="v_o" units="volt"/>
      <variable name="R" units="ohm" initial_value="1"/>
      <variable name="C" units="farad" initial_value="1.333333"/>
      <variable name="L1" units="henry" initial_value="1.5"/>
      <variable name="L2" units="henry" initial_value="0.5"/>
      <variable name="v_c" units="volt" initial_value="0"/>
      <variable name="i_1" units="ampere" initial_value="0"/>
      <variable name="i_2" units="ampere" initial_value="0"/>
      <math xmlns="http://www.w3.org/1998/Math/MathML">
         <apply>
            <eq/>
            <ci>v_o</ci>
            <apply>
               <times/>
               <ci>i_2</ci>
               <ci>R</ci>
            </apply>
         </apply>
         <apply>
            <eq/>
            <apply>
               <diff/>
               <bvar>
                  <ci>t</ci>
               </bvar>
               <ci>i_2</ci>
            </apply>
            <apply>
               <divide/>
               <apply>
                  <minus/>
                  <ci>v_c</ci>
                  <ci>v_o</ci>
               </apply>
```

BioSignalML clients

- CellML modelling:

```
<model name="bwfilter" cmeta:id="bwfilter" xmlns="http://www.cellml.org/cellml/
   <component name="filter">
      <variable cmeta:id="time" name="t" units="second"/>
      <variable cmeta:id="input" name="v_i" units="volt" initial_value="0"/>
      <variable cmeta:id="output" name="v_o" units="volt"/>
      <variable name="R" units="ohm" initial_value="1"/>
      <variable name="C" units="farad" initial_value="1.333333"/>
      <variable name="L1" units="henry" initial_value="1.5"/>
      <variable name="L2" units="henry" initial_value="0.5"/>
      <variable name="v_c" units="volt" initial_value="0"/>
      <variable name="i_1" units="ampere" initial_value="0"/>
      <variable name="i_2" units="ampere" initial_value="0"/>
      <math xmlns="http://www.w3.org/1998/Math/MathML">
         <apply>
            <eq/>
            <ci>v_o</ci>
            <apply>
               <times/>
               <ci>i_2</ci>
               <ci>R</ci>
            </apply>
         </apply>
         <apply>
            <eq/>
            <apply>
               <diff/>
               <bvar>
                  <ci>t</ci>
               </bvar>
               <ci>i_2</ci>
            </apply>
            <apply>
               <divide/>
               <apply>
                  <minus/>
                  <ci>v_c</ci>
                  <ci>v_o</ci>
```

```
$ ./bwfilter http://devel.biosignalml.org/recording/physiobank/nifecgdb/ecgca102/signal/3
```



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
  - Ontology to derive units from other units.
  - An extensible way to automate units validation and conversion.

Thank you