Building a service-oriented platform for online physiological data analysis

M. Colom
http://mcolom.info

CMLA, ENS-Cachan
Reproducible Research

- **Redefine** the product of research:
  - Article, source code, data

- **Why** do we need it? → **Trust results**

- **Applicable** to all disciplines? Cosmology, Biology, Computer Science...?

- What if we **combine** RR with **Clinic Research**?
SmartAlgo

- A new **platform** for **RR** in algorithms applied to **Clinical Research**
- A joint project

- Which kind of medical problems/algorithms?
  - Balance and movement
  - Eye tracking (Infantile Nystagmus Syndrome, Spasmus Nutans-type nystagmus)
  - ...

→ Online prototype demo: *Animated Statos*
SmartAlgo: Oculo project

Quantitative assessment in daily clinic

- Tablet
- IR recording
- Data processing center
- Databases
- WiFi, few seconds
Similar projects

IPOL

Run My Code
Some differences...

- **IPOL** is a **full RR journal**, with **peer-reviewed** article and source code. The demos are a valuable tool, but not peer-reviewed.

- The aim of RunMyCode is to give **visibility** to the results of the research. They publish **non-peer-reviewed** source code and data.

- SmartAlgo is somewhere in the middle.
  - **Peer-reviewed**
  - A **platform for clinical research.** Not just a repository of code or demos
  - **Data** is **real** and come from actual physiological signals obtained with sensors
  - Data needs to be **standardized** because of the different kinds of sensors (for example: different sampling rates, formats, etc)
  - **Validated** and **annotated** data
A technical challenge?

• Of course. It's a complex system which includes:
  – Signal preprocessing and standardization
  – Multiple kind of signals
  – Annotation of signals
  – Storing and retrieving efficiently all the information
  – Complex interface interactions (web, tablets)
  – Etc.

• So, the main difficulties are technical? NO
An agreement challenge!

- It's mainly an agreement problem, not just technical
  - Physicians normally are far away from algorithms, mathematics, formal methods
  - Mathematicians and engineers are not familiar with neurological pathologies or diagnostic methods
  - (Of course!)
  - But the problem needs a multidisciplinary approach to apply advanced techniques of signal-processing and machine-learning to obtain results in clinical research.
  - But physicians and mathematicians/engineers usually talk very different languages.
A dual point of view

- It’s the same problems, but seen from different angles
- For example,
  - Physicians interested in: fall assessment, balance of patients, eye tracking, walk of patients, ...
  - Mathematicians/engineers interested in: models, classification, regularization, generalization, automatic learning, ...

- Problem: which kind graphical interface should be show? Something in the middle?
- Solution:
  - Each user has a role (physician, mathematician/engineer)
  - The graphical interface first matches the general role
  - But it must be adaptive: it should be customizable and remember the preferences of the user.
  - Why this way? Two “different worlds”, but the same problem → They should converge.
Only an “agreement problem”?

- Not only!
- Other issues, very particular of this project
Other issues: input data

- **Real data from physiological signals**
  - Sometimes **incomplete**
  - Might be **inaccurate**
  - Characteristics of the sensor might be **undocumented**
  - Many **different** captors and devices
  - Need to **preprocess** the input data
  - Need to **standardize** all data in a **common format**
Other issues: privacy

- It's data from real patients!
- This kind of data cannot be
  - Stored
  - Made public
  - ...
- Very strict usage conditions
- Legal framework:
  - l'article 8 de la convention européenne de sauvegarde des droits de l'homme
  - la directive 95/46ce
  - la loi du 6 janvier 1978
  - le décret n°2006-6 du 4 janvier 2006 sur l'hébergement de données de santé à caractère personnel sur support informatique
  - l'ordonnance n°2010-177 du 23 février 2010 – article 19
  - ...

- So? Any solution?
  - We're within the special case of clinical research:
    - Low-level signals
    - Need to anonymize data, absolutely
Development cycle

- Designing **usable interfaces and proper data visualizators** is difficult:
  - Physicians and mathematicians/engineers have **different interests**
  - It's difficult to have an **idea** of a new system **until you see a usable prototype**
  - Even **designing** and **modifying** a prototype is **expensive** in terms of **time** and **human resources**
Our proposal: User eXperience Design (UXD)

- **Interviews** with the physicians to understand their needs and the particular problems in their field
- The same with mathematicians/engineers
- Imagine **use scenarios**
- Design **wireframe** or **mockup interfaces** → Show ours
- **Discuss** these interfaces with the users
- **Iterate** the prototypes until **agreement**
- When agreement: write **better prototypes** (real HTML5/CSS), **integrate code**, iterate.
At which point are we now?

- Designing **use scenarios**
- Writing **machine-learning** and **signal-processing** algorithms
- Designing **adaptive user interfaces**
- Building a **development team** → Need of a **large team of engineers**, in UX, design, machine learning, integration, coding, … **Big project!**

- **Antecedents**: we have the experience of have been building **IPOL** at **CMLA**. But still very different!
What do we expect of SmartAlgo?

- **Reproducible Research**
- **Provide** Clinical Research with a **platform** with the best machine-learning and signal-processing **algorithms**. And **data**!
- Have **methods** and **data** we can **trust**
- Create a **large network** of clinical and non-medical researcher contributing with data
- Give the **technical means** (platform, data, algorithms) to establish a **Clinical Reproducible Research community**.
Thank you very much for your attention

Miguel Colom

Website: http://mcolom.info
Email: miguel@mcolom.info

CMLA, ENS-Cachan