Health Data IG Minutes
"Health Data Privacy & Security issues"
RDA 8th Plenary, Denver, 17/09/2016

Chairs & Speakers
Yannis Ioannidis (*introduction and moderator*)
Anthony Chang (*first speaker*)
Edwin Morley-Fletcher (*second speaker*)

Attending individuals - Broad spectrum of backgrounds:

- IPSR depository social sciences researcher, niocady project, US project to index health data
- Agricultural sciences information specialist
- Health sciences information specialist
- Researcher from bioscience centre Japan
- Data management librarian
- Astronomer, liaison technical advisory board of RDA
- Software developer of medical data
- Ex-astronomer, data manager
- Security researcher
- Computer scientist
- Infrastructure epidemiologist, data release from birth
- Head of data publishing at springer nature, open access publishing
- Council member
- Swedish national bioinformatics structure researcher
- Legal officer from the European Commission
- Athena research center researcher
- Public health care privacy rights specialist
- Physician, researcher in bioinformatics, MD for a company
- Medical student, UK

Objectives of the meeting

1. Summary of discussions at BoF sessions at 6th and 7th Plenaries (Yannis Ioannidis)
2. Overview of health data (Yannis Ioannidis)
3. Data driven medicine (Anthony Chang and Edwin Morley-Fletcher)
4. Open discussion (all)
5. Next steps (Yannis Ioannidis)

1. Summary of discussions at BoF sessions at 6th and 7th Plenaries - Background

HD IG idea rooted in several EU projects in biomedical informatics (Health-e-Child www.health-e-child.org; Sim-e-Child www.sim-e-child.org; Cardioproof http://www.cardioproof.eu; MD-Paedigree www.md-paedigree.eu; p-medicine www.p-medicine.eu; Avicenna www.avicenna-isct.org)

Preparatory BoFs during Plenaries 6 and 7 were attended by over 35 individuals from equally diverse composition as today's session.

P6 and P7 discussion results included an emphasis on privacy, security, ethical issues, and to relate data management and processing issues.

Scientific activities relating to IG are particularly life sciences, specifically trust, IPR protection, privacy and personal protection.

Health Data IG charter discussed in the previous BoF sessions addressed the following topics:

- Data access and protection
- Data-based healthcare for personalised medicine
- Data literacy in health care
- Patient data repositories

- In-silico drug development and clinical trials
- Policy making

2. Overview of Health Data - How does the Health Data IG fit into RDA?
HDIG is unique in dealing with vertical health data.
Nevertheless there might be connections with many related groups, for example bioSharing registry, big data, ELIXIR.
Therefore, in the next Plenary (P9) in Barcelona there may be links or joint sessions with these related groups, with a special focus on the resolution of overlapping areas.
A first output is expected to be delivered 12 months after prioritisation.

3. Data Driven Medicine
3.1. First speaker, Anthony C. Chang, "AI in medicine (medical intelligence)"
The first question arises from the consideration that artificial intelligence is already utilised in many areas. So how can we apply AI to clinical care for patients?
There are many problems in medicine that need to be tackled:

- Missing data: not uncommon for patient data to have 50-70% missing.
- Unstructured data: 90% of HC data is unstructured. For example, unstructured note taking, difficult for computers to interpret.
- Escalating data: medical literature doubles every five years, physicians cannot obtain that volume of information.
- Future data: thanks to wearable technologies and interpretation, we will have a "big data tsunami".

Health care is mired by the data management. We need to get better at utilising and collecting the data here. Furthermore, there is an increasing volume of healthcare data, wearable technologies, processing powers, cloud computing. Data mining is gaining popularity, instead of randomised control trials. There is not a big data problem, we do not need more data, there is a *structuring problem.*

AI in medicine
AI in medicine is not using robots, it is instead a new "energy" that can be harnessed. Computers need to be able to understand communication between individuals. However, we need to mimic the brain but not copy it exactly. For example, a bird flies by flapping its wings: humans tried this initially but we were unsuccessful. Instead we tried something slightly different and now we can fly faster. Cloud computing and big data are tremendous opportunities for HC, but it's a very hard attempt to change the culture and encourage sharing. Medical intelligence could include image recognition, machine learning, natural language processing, and decision support systems.

Possibilities once these problems are overcome

- Decision support and hospital monitoring: ICU data and being able to predict adverse events live rather than post hoc analysis. Hospitals pool data so that multicenter data can be interpreted. Need for pro active initiative.
- Medical image and biomedical diagnostics: it is under-utilised. Combining AI + EHR + genomic sequencing could lead to very precise interpretation and personalised therapeutics (additionally when a patient has investigations). Has there been a change in imaging over time?
- Precision medicine and drug discovery: it involves deep learning techniques, genomics, transcriptomics, proteomics, drugs. We shouldn't take a drug and test the drug on a patient, we need to tailor treatments to patients.
- Digital medicine and wearable technology: for example iClinics, online clinics which incorporate wearable technology, telemedicine and virtual assistants. Good for high maintenance patients, can decrease number of visits by leveraging technology.
- Robotic technology and virtual assistants: virtual assistants can answer frequently asked questions.

Data scientist can tell us what we can't see, they can make "...the visible invisible" and the "...visible invisible". In the future, thinking about "super AI" in the cloud, it could be accessed

by physicians and patients around the world. We will shift from evidenced based medicine, which is prone to bias, to *intelligence based medicine*.
How can we do this?
Humans need to work closely together, collaboration can help us with this. To achieve data mining there needs to be shared data. Therefore, we need to be comfortable with cyber security. Approximately, 50% have had issues with data security, and it is surprising there haven't been more breaches.
Blockchains may be a good way to achieve security, through encryption and decentralisation, we could achieve greater security.

For additional information see: aimed-mi3.com

4. Discussion
- Q.: The speaker described a tsunami of data, whereas what we actually have is multiple reservoirs with high fences that are difficult to access. Due to management difficulties, it is hard to get this data out. How can we address this?
A.: Key people/barriers

- Legal - how are you going to use the data
- IT - data security
- Executive suite

How to overcome these barriers

- Group-think mentality: once a small number of hospitals have adopted it, others will see that it works. We have to be patient and persistent.
- We need to be able to share data in a standardised way: at the end of the program, every medical record will have to be from a data-stamped open system.
- We need criteria and validation.
- For rare paediatric diseases, there's a need for international consortium.

- Q.: What about errors in AI diagnostics, who is liable?
A.: AI supercedes physicians, AI is faster, better and more accurate. AI is not perfect, it can't make creative diagnoses. Instead the question will be, how liable will you be if you don't use AI. False positive rates are unacceptably high, especially for prostate cancer.
The legal discussion shouldn't distract us from development but must be considered. Policy and ethical discussions are lagging behind development. Additionally, physicians are disconnected from the data world, we need to force the exchange.

- Q.: How do we protect data interpretation from misuse and misinterpretation that can occur in evidence based medicine?
A.: We need to move away from published data, we need to access the information that is not published. Once that is unlocked there will be a tremendous difference in medicine.

- Q.: Where will we see the most movement in healthcare AI?
A.: Medical imaging and diagnostics including genomics and data interpretation. There is a data interpretation and volume issue in these areas.

3.2. Second speaker, Edwin Morley-Fletcher, "MyHealthMyData"
MyHealthMyData is a data collection project encompassing 15 participants: 5 SMEs, 4 clinical partners, 4 Research centers and academia, 1 legal consultancy, and 1 industry partner.
€4m grant, to start 1st November 2016.

What is the role of MHMD?

Aggregate personal data from disparate sources (social media accounts, clinical data repositories, personal drives, wearable devices, etc.) in a centralised, user-owned account (PDA).

Questions MHMD will deal with

What are users and patients willing to share? How much data is necessary? Who will use the data? Who owns the data? What is the value of the data? What are the benefits and risks of sharing?
MHMD aims at assigning data access rights in an intuitive and efficient workflow, with exhaustive but simple questions:

- Type of data requested,
- Intended use,
- Data that will be retained,
- Data that will be shared with 3rd parties and intended use,
- Implementation of the Right to be forgotten.

MHMD shall enable to revoke data access rights, or extend them.
A "privacy compass" was developed to summaries all privacy access.

- Stay informed of, and enquiry on, relevant data transactions after access has been granted.
- Be able to receive requests from stakeholders for data access permissions. Requests may also include incentives offered by stakeholders in exchange for data.
- Defined post-mortem usage or donation of personal data.
- Defined user entitlements

System architecture addressing what where how and why for partner contributions.
Blockchain, started with bitcoin, allows non-third party based transactions. It is a secure non-editable record of all transactions taking place. Stakeholders can be provided with a 'wallet' containing an encrypted identifier and dynamic consent. Smart contracts bind peoples to specific actions and outcomes. Data can be accessed anywhere without the need for a centralised secure hub.

There are two layers of data flow

- Semi-automated data layer: ensures data profiling and quality.
- Privacy preserving and security layer: anonymisation and encryption preserve privacy.

Dynamic Consent
It is necessary to acknowledge dynamic consent, this can be done through smart contracts, which allow users to track, monitor and transparently document HC data.

Dynamic Consent allows to extend traditional consents, combining them into a user workflow in which patients may or may not allow access to their data based on a range of key parameters:

- What will data be used for
- What will be done with the data
- What data will be retained
- What data will be shared with 3rd parties and for what purpose
- How will the right to be forgotten be implemented

4. Discussion
FAIR data should be incorporated.

- Feasible
- Accessible
- Interoperable
- Reproducible

Mapping exercise is important to prevent duplication of tool or recommendations.
How does health related data, but not patient data, fit into the IG? This is in scope.

5. Next steps
Looking forward to P9 in Barcelona, a TC meeting will be planned by December to starting organise the next meeting with a major involvement of other group participants and projects.

Assistants: Ludovica Durst, Fotis Kariagiannis
Minutes: Matthew Byrne