MyData4COVID19 Workshop Minutes
Friday, June 12th, 2020,
14:00 - 17:00 CEST

HOW TO CONTRIBUTE TO THIS DOC:

- To suggest edits, use suggest mode
- To comment or ask a question, use the commenting tool

If in doubt ask help from Slack from the #mydata-4-covid19 channel

Links

Workshop slides [ref]

Human-centric Data Platform for managing COVID-19 pandemic

- EIT Digital proposal presentation [ref]
- White paper (WIP) [ref]
- MyData Commons - Prototype Update 4th April - Iain Henderson [ref]
- MyData Commons Prototype - Blog - Iain Henderson [ref]

Theory of change - COVID related [ref]

Corona Project Management team (not active)

- Meeting minutes [ref]
- Projects - Trello Board [ref]

Table of Contents

Section 1: Introduction (Karolina)

Participating:

- Isabelle de Zegher (MyData Belgium)
- Casandra Grundstrom (MyData Oulu)
- Iain Henderson (MyData Scotland)
- Davide Calvi (MyData UK)
- Dixon Siu (MyData Japan)
- Sherry Chung (MyData Taiwan)
- Adrian Gropper (MyData US)
- Frederik Lindén (MyData Sweden)
- Karolina (MyData Staff and Moderator)
- Lal Chandran
- Crt Ahlin
- Hessie Jones
- Lucy Yang
- Paul Theyskens
- Alessandro Carelli

Apologies

- Louise Helliksen
- Marcelle Ngounou
- Ioannis Goumenopoulos

What is our hope? clarity, momentum, build something concrete, unite, contribute, human-centric, standardize interfaces, interoperability, integration, individuals control, privacy, collaboration, implementation, design the future, address needs proactively

Section 2: Setting the Scene (Isabelle & Casandra)

Key Takeaways

1. We want a concrete & actionable to guide the thematic group
2. The data commons needs trust to overcome consent and sensitive data
3. A data commons definition

Discussion

1. **Do we agree on the problem to be solved?**
    a. We need a bottom-up solution that covers ***medical, economic and societal problems*** caused by COVID 19
        - it is not only healthcare, but it is also about supply chains and value chains - and *How can the economy start back up? (back to work)*
        - impacts is changing and will keep changing over time
          => this is a ***moving target***
        - Need to change because we are moving to digital (equal to going from prescription to electronic prescription) – ***BIGGER than COVID***
        - More than just mobility, more than just confinement, more than just the immediate medical crisis
    b. We don't have the right data
    c. The data commons is a fantastic solution, but how can we deal with sensitive health data?
        - Need to solve TRUST => management of data commons in a TRUST / foundation (cannot give the date to the government).
          Need GOVERNANCE / next generation governance model
    d. ***Sustainability point of view***: (1) time to act is NOW (pandemics is going ahead) - but need to think at how the solution can be implemented for other challenges (mobility)

2. **Leveraging forces of MyData (are we missing something/how can we optimize)?**
    a. MyData brings two attributes that are pretty unique in this situation.

        - the human-centric root and all that implies for richer, more complete data, and
        - an approach that is inherently global (like the virus is)
    b. Can the solution be thought about in different contexts?
    c. Can it be 'sold' as something that is missing from the world?
3. **Others point:** We are missing a shared understanding of what a data commons is - it is a buzzword without clear boundaries
        - Commons is a term used in jargon terminology but is used and confused like many buzzwords are. But what we actually mean is a data *trust*: a legal entity put together in order to tackle a particular problem with specific terms and conditions - - purpose, period of time - and requires careful crafting.

Section 3a: Proposed Solution - Pillars (Iain) & MyData Operator - Architecture (Davide)

Key Takeaways

1. Existing MyData commons Prototype uses DID anonymized data
2. Question: Who would be trustworthy to hold the data of the commons/trust? Answer: MyData. Why? Because you need human-centricity and you need global solutions.

Discussion

1. **The dilemma: Centralised or Decentralised data storage? Can it be a mixed approach to the third approach?**
    a. Decentralized
        i. Sustainable
        ii. Controlled by the individual
        iii. The way forward ?
    b. Centralised is a collection of decentralised
    c. Mixed
        i. Most of the time the data is decentralised, such as in PDS and is more flexible

d. Physical versus logical ?
   From the data commons perspective, it is a logical view of the whole thing - one operator isn't running everything
e. With decoupled storage, the operator's "brand" could still serve the purpose of pulling in users
f. Local PDS would be better at building local distribution and dealing with nuances in specific locales - different models for different uses
g. The aggregate view of the subset of data is the most important perspective
h. For this to be of use, it needs to be adopted
i. Should this be part of the specification of architecture when there are detailed functional requirements for the data set
j. We need to clarify definitions

2. **Can Data Commons data model and MYSA container be the cornerstone for MyData Operator interoperability?**
   a. MYSA is about setting up a smart contract between MyData and the Individual
   b. CommonAccord in Stanford - digitalizing contracts
   c. From the DG Justice in the EU Commission, there is a working group for establishing digital contracts, there is a solution to this for this from the legal point of view : Modelling the EU Economy as an Ecosystem of Contracts
   d. No one outright objected to this being the approach we will use, but we should keep in mind that it is very complicated - we want to move forward on this topic and provide a 'lessons learned'. We will show it could work!
   e. One thing to consider if local PDS is preferred distribution — We need to consider performance implications (battery, capacity), and security (reliability)
   f. Personium would decide later whether it is easy to support data commons and MYSA without big modifications in our core development. Therefore I would not say now those are the cornerstone for MyData Operator interoperability
3. **Which are the technologies/interfaces that can be used both for identity, permission and consensus management and enhancing interoperability?**
   a. What we are planning in MyData Taiwan leverage decentralized ledger as the permission and consensus management. If users

consent, the history can be recorded in the ledger such as IBM hyperledger or IOTA. This can avoid the use of the centralized server for consensus management and we hope that it will be easier to gain cross-organization trust.

Section 3b: Integrating Existing Applications (Lal & Dixon)

Key Takeaways

1. We need to be wary of building solely medical focused solutions when the needs have moved on to economic and societal. That's not to ignore medical, it remains a big challenge in many places, and there will be further waves and further viruses. But let's not get pigeon-holed into medical.
2. It is difficult to answer many of these questions without detailed user stories covering a whole range of medical and non-medical scenarios. The user stories will drive the requirements that will shape the functional and then technical requirements.
3. We need some sessions, ideally led by the design team, to take the work on customer journey and persona forward.

Questions to be Addressed

1. Global interoperability of COVID app
2. Standardised permission management
3. Standardisation data sharing agreement handling
4. Reduced development time when it comes to human-centric and compliant data exchange capabilities
5. Standardised logging of consents and data exchange
6. Standardised data exchange JSON-LD for data sources and data using services

1. App: What is the proposed metadata schema for the proposed COVID app?
2. App: What is the type of the attribute? (Mandatory/Recommended/ Optional). Explanation on the rationale of the choice made.
3. App: Where is the data stored?

    a. Proposal is to store the data in the application, in the device or the server.
    b. The server service could be provided by the application provider or the data operator
4. Data operator: Do we enable granular access control at purpose and attribute level?
5. Data operator? Do we address Pseudonymisation and anonymisation of personal data?
6. Where should the anonymisation process be performed? Within the PDS or at the places that collect the data?

Discussion

These are the minimal things that need to be addressed from the application point of view for the data commons. Having a common infrastructure can be beneficial for different actors in the COVID context for modular contributions and value-based outcomes.

What do we really need in order to make it happen? A common integration and adaptation layer.

1. **Global interoperability of COVID app**
    a. global interoperability: with decentralized storage, the access to (properly anonymized) data could not be blocked or under control of one entity (nation). (And when I talk about decentralized storage, I don't mean on-device data, but a globally accessible data in Swarm storage)
2. **Standardised permission management**
3. **Standardisation data sharing agreement handling**
    a. ODPI creates open source standards to help you use and understand data across all platforms.
4. **Reduced development time when it comes to human-centric and compliant data exchange capabilities**
5. **Standardised logging of consents and data exchange**
6. **Standardised data exchange JSON-LD for data sources and data using services**
7. **App: What is the proposed metadata schema for the proposed COVID app?**

a. How does a key look like?

8. **App: What is the type of the attribute? (Mandatory/Recommended/ Optional). Explanation on the rationale of the choice made.**
9. **App: Where is the data stored?**
    a. **Proposal is to store the data in the application, in the device or the server.**
    b. **The server service could be provided by the application provider or the data operator**
10. **Data operator: Do we enable granular access control at purpose and attribute level?**
    a. If we agree on a core model across different applications, steps can be taken to follow a specific protocol or architecture to have interoperable/portable data ex. MyLog, Data4Life, Personium
    b. Verified Claims and similar needs granularity at the attribute level
    c. If you manage permissions at the lowest level they can all be rolled up
    d. The choice of granular control is an opportunity for design implementation, but the architecture should be able to support all consent management
    e. We don't need to be micromanaging the data, there is a limbo level that will be ideal for the purposes of data commons
11. **Data operator? Do we address Pseudonymisation and anonymisation of personal data?**
    a. Depends on the sharing purpose, for the application itself we should be able to choose
    b. The purpose drives the chosen approach of personal data pseudo or anonymization
    c. Depends on use case
12. **Where should the anonymisation process be performed? Within the PDS or at the places that collect the data?**

Section 4: Towards a RoadMap: Milestones (Isabelle)

Key Takeaways

1. Mobility in the context of a crisis

Questions to be Addressed

1. What defines success?
2. Workstream and phases:
    a. OK?
    b. Participation in deliverables/ target dates
3. Funding: targets - what is realistic?

Discussion

1. **What defines success?**
    a. Concrete results
    b. Show value of MyData operator model
    c. Reusability for other scenarios
    d. Library of user stories that impact human-centric application
    e. Reflecting the current pandemic-implications and providing a user solution
2. **Workstream and phases:**
    a. **Participation in deliverables/ target dates**
        i. If you see something you want to participate in please indicate that
        ii. If you know of experts who might be interested please invite them in
3. **Funding: targets - what is realistic?**
    a. Crowdsourcing capital?

Section 5: Conclusions (Karolina & Casandra)

- A common understanding of data commons
- Solutions should be easily reusable in other contexts
- MyData is the trustworthy data commons source for a human-centric global solution
- Opportunity to benefit from reuse of existing structures, applications, and data sources
- The TG is a sandbox to pragmatically have expert-selected solutions working towards a common problem (credit Davide)

- ZOOM