CERN openlab

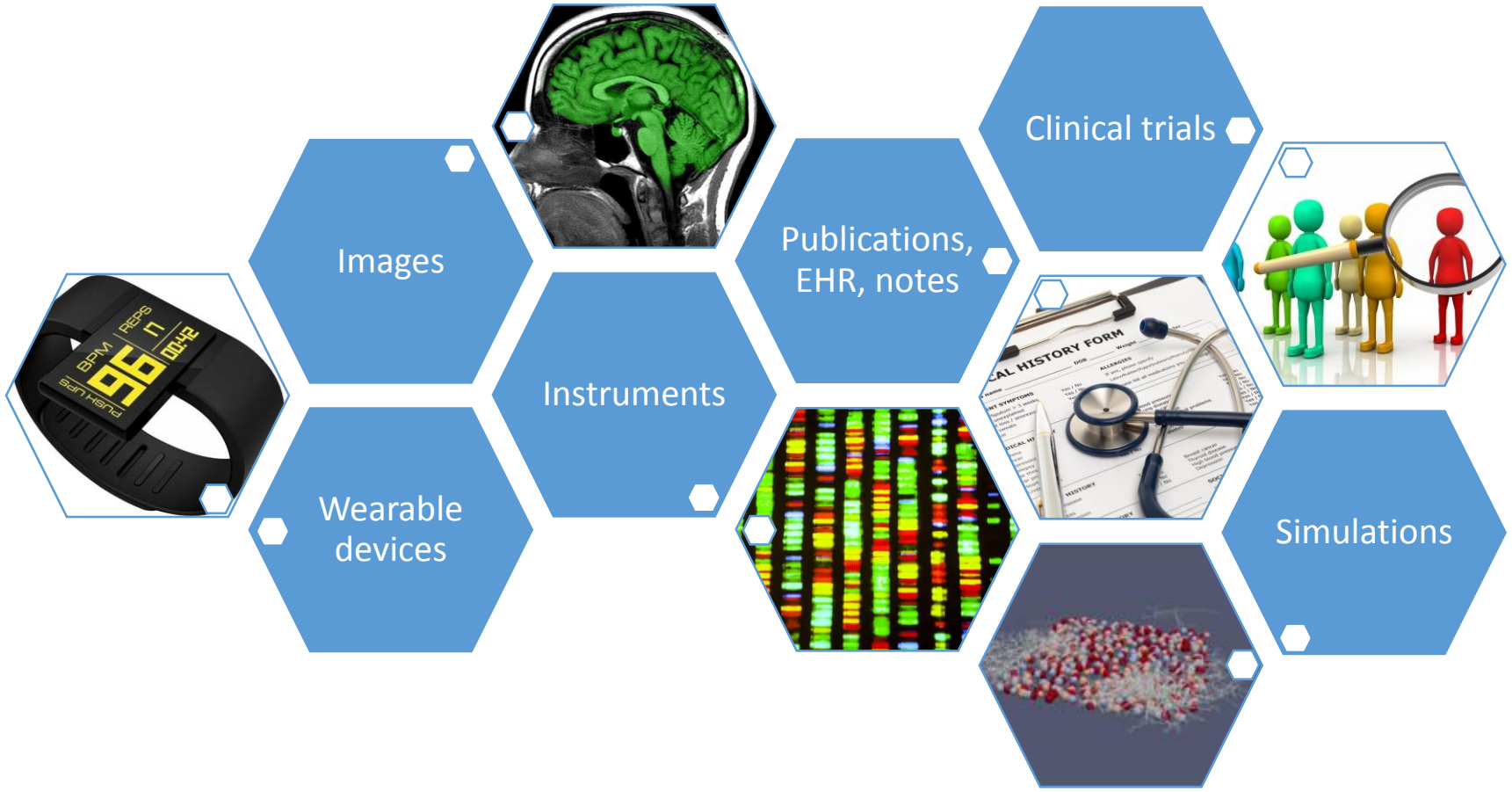
LivingLab

A Socio-Technological Platform for Making Sense of (Medical) Data

LivingLab Kick-Off Meeting

Alberto Di Meglio – CERN openlab Head

Medical Data Deluge



- "**150 EBytes** of medical data in the US, growing 48% annually" [1]
- Cost of instruments and laboratory equipment decreasing fast (e.g. **sub-1k$ genomic sequencers**)
- Medical and fitness wearable devices on the rise, projected data produced in 2020 **335 PB/month** [2]

[1] Esteva A. et al., A Guide to Deep Learning in Healthcare, in Nature – Medicine, Vol. 25, Jan 2019, 24-29
[2] https://www.statista.com/statistics/292837/global-wearable-device-mobile-data-traffic/

Rising Interest in ML/DL

In the past 6-8 years applications of ML/DL techniques to medical data have rapidly developed. For example:

- **Supervised Learning** for classification of skin lesion images
- **Reinforced Learning** for robotic-assisted surgery
- **CNN and Transfer Learning** for complex cancer diagnostics from scans
- **Data augmentation and GAN** for training histopathology models on limited datasets or unlabeled sets
- **NLP/RNN and auto-encoders** to analyse EHR, predict diagnosis from temporal sequences of events, or to perform automatic transcription and summarization of doctor-patient conversations
- **Generalized DL** methods applied to genomic analysis, GWA, or phenotype prediction, combining genomic data, images and other sources

System Biology Principles

- The availability of large amounts of data of many different types fosters a new approach to research of complex biological systems, including the human system
- A “holistic” approach where interactions between different parts are also considered, rather than a “reductionist” approach where single parts are studied and specialised clinical solutions are adopted
- A natural field of application of advanced data analytics and deep learning methods
- Wide range of applications from large-scale statistical studies to “personalized medicine” where holistic models are applied to individual systems (persons)

Health Studies and “social simulations”

- One of the major points coming out of the workshop on Big-Data in Medicine two weeks ago was the need for open platforms to test and simulate interactions among cross-disciplinary activities and medical/healthcare scenarios
- Although the scale needed for this is ambitious and won’t be in LivingLab from day one, our platform could provide one of the few examples of this type of specialized collaboration platforms where different types of “things” are integrated where thing means data, people, policies, processes, ideas, etc.

Many Challenges Ahead

Many different types of data (structured, unstructured, images, PDFs, etc.) of widely different quality

Lack of dominant standards

Privacy and data protection

Social, cultural, ethical opinions

LivingLab Main Objectives

R&D

Investigate ML/DL tools and techniques to process heterogeneous data sets, perform anomaly detection, mitigate noise
Build expertise in block chain technologies for end-to-end data integrity validation

Legal and Policies

Build expertise on data protection regulations and tools

Assess state-of-the-art EHR management systems

Knowledge Sharing

Share knowledge and expertise about ML/DL from the many ongoing investigations on methods, frameworks, tools, platforms at CERN and in prominent medical research facilities

Medical and Social

Investigate challenges and solutions of providing preventive, personalized medical care services in collaboration with medical doctors, support standardization and data protection efforts, organize education and training events

LivingLab Members and Roles

CERN: CERN openlab, project management, design and implementation of the analysis platform (IT-DB, KT/MA-funded resources), ML/DL expertise, organization of events, seminar, training (technical, legal), communication, human networking, **specific use cases**

Be-Studys: protected data management tools, legal and compliance expertise, ML/DL expertise, funds for 1 DOCT + secondments of relevant skills, training, human networking

ScimPulse: translation/clinical use cases, medical expertise, human networking, 1 person seconded to CERN

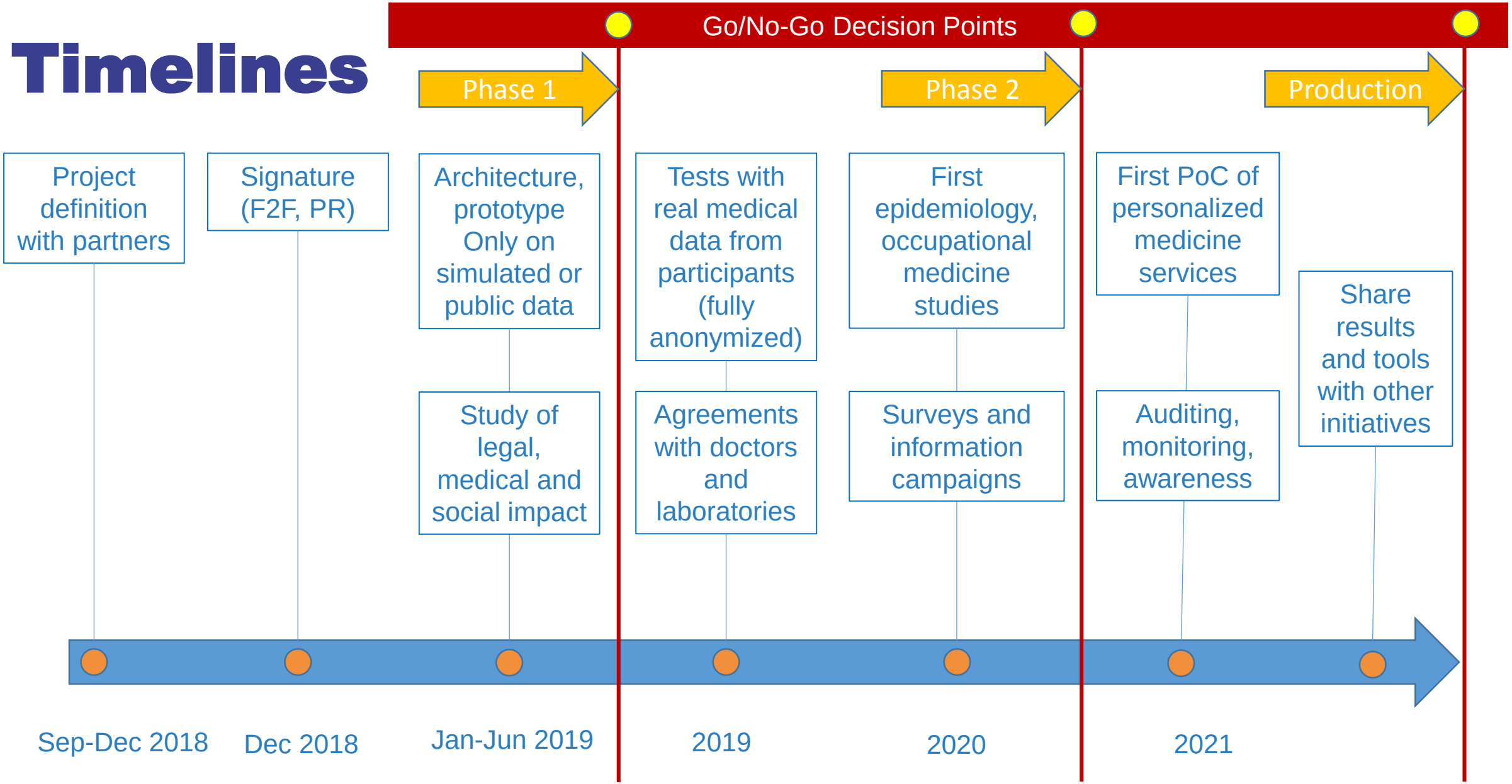
Intel: not a direct member, but support with specialized hardware and software

Community: not direct members, any entity (hospitals, doctors, labs, etc.) with interest and relevant data/skills

CERN Stakeholders/Interested Parties

- **CERN openlab**: general coordination, communication, participation in selected projects, contributions to fund resources
- **IT-DB**: data analysis platform operations and support (best effort)
- **KT/MA**: responsible for the CERN MA strategy, exploitation and outreach of results, internal and external networking, contributions to fund resources
- **Medical service (HSE)**: participation in selected projects, interest in occupational medicine, EHR handling
- **Data Protection Office:** oversight of CERN DP compliance, interest in DP discussions
- **Legal service**: general interest in legal aspects and respect of CERN status, privileges, immunities
- **CERN community**: specific interests in proposing activities on the platform and evaluating technologies

Timelines


Go/No-Go Decision Points
Phase 1
Phase 2
Production
Project definition with partners
Signature (F2F, PR)
Architecture, prototype Only on simulated or public data
Study of legal, medical and social impact
Tests with real medical data from participants (fully anonymized)
Agreements with doctors and laboratories
First epidemiology, occupational medicine studies
Surveys and information campaigns
First PoC of personalized medicine services
Auditing, monitoring, awareness
Share results and tools with other initiatives
Sep-Dec 2018
Dec 2018
Jan-Jun 2019
2019
2020
2021

Kick-Off Objectives

- Get to know each other
- Agree on broad definitions of scope and goals for LivingLab
- Provide a general overview of necessary background information (resources, policies, roles, skills, etc.)
- Define a few initial technical and organizational objectives and the activities (“projects”) to achieve them
- Work out more detailed timelines and milestones

CONTACTS

ALBERTO DI MEGLIO
CERN openlab Head
alberto.di.meglio@cern.ch

MARIA GIRONE
CERN openlab CTO
maria.girone@cern.ch

FONS RADEMAKERS
CERN openlab CRO
fons.rademakers@cern.ch

FEDERICO CARMINATI
CERN openlab CIO
federico.carminati@cern.ch

ANDREW PURCELL
CERN openlab Communications Officer
andrew.purcell@cern.ch

KRISTINA GUNNE
CERN openlab Administration/Finance Officer
kristina.gunne@cern.ch

www.cern.ch/openlab

Thanks!

alberto.di.meglio@cern.ch
@AlbertoDiMeglio