



To Become Fit for the IoT Data Game Change

Peter Wittenburg

Max Planck Society, Max Planck Computing & Data Facility

RDA Europe Director

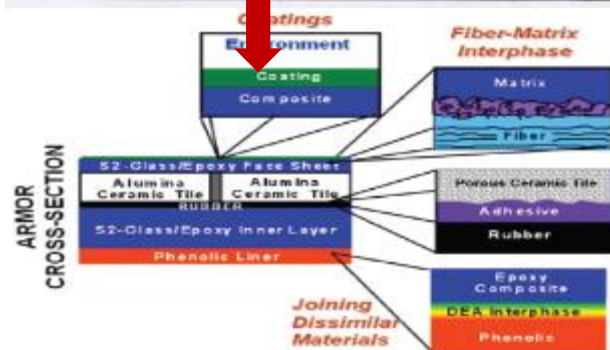
NoMaD - Material Science



- many Labs create data about materials and compounds (experiments + simulations)
 - space of chemical compounds is endless
 - let's categorise this space to quickly find useful compound materials?
 - > 3 Mio aggregated entries now



Periodic table of elements with color-coded groups: Alkali metals (blue), Alkaline earth metals (orange), Transition metals (green), Post-transition metals (yellow), Nonmetals (purple), Halogens (red), Noble gases (grey).



- categorisation via Machine Learning etc.
 - **Revolution: writing paper is not the only scientific goal anymore**
 - **it's about repurposing**

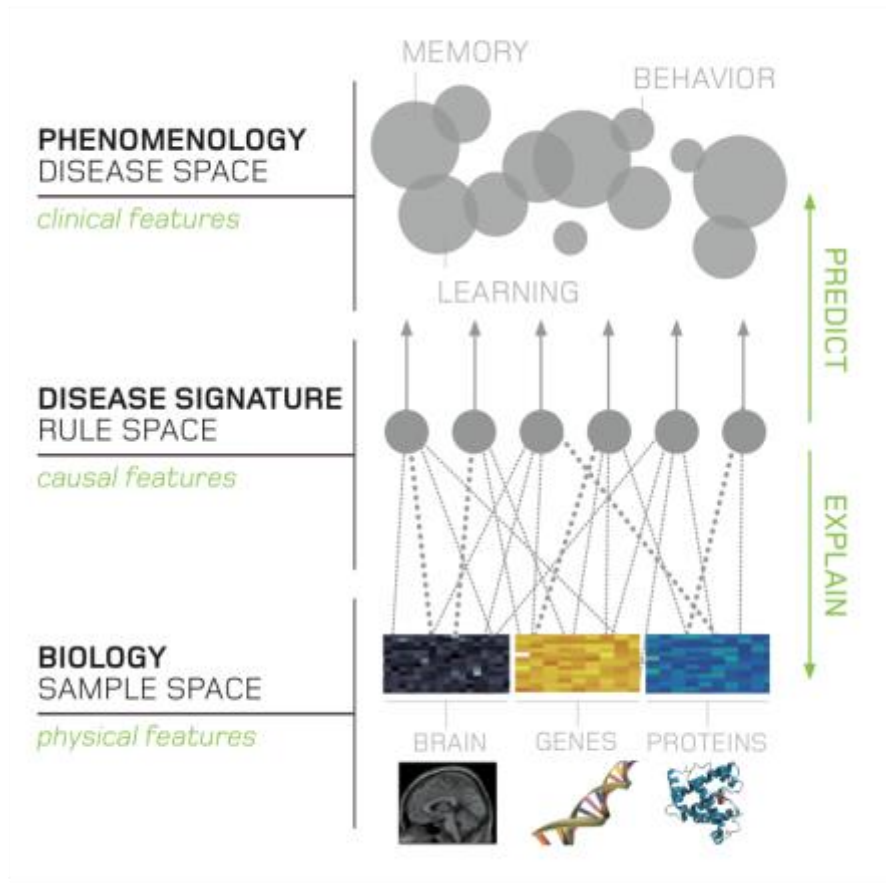
DOBES – Humanities/Languages



- ~70 global teams
- ~80 TB in online archive
- 4 dynamic external copies
- remote archives

- how can one use data to validate theories about the evolution of languages (and cultures) over thousands of years
- how to understand which languages are more "economic" than others
- *Revolution in humanities: scientific paper is not only goal anymore*
- *it's about repurposing*

Brain Research – Detect Disease Patterns



- early detection of causal basis of brain diseases
- machine learning to correlate patterns in data with phenomena
- much data from various specialized labs and hospitals is required

- **Revolution in medical world:**
 - *sharing data outside of the hospitals - solving r & e*
 - *it's about repurposing data*

What do they have in Common?

- › addressing new questions by aggregating Big Data
- › re-usage of data for different purposes
- › huge effort to do the “integration” with complexity as most crucial factor
- › why?
 - › huge fragmentation of the solutions space, insufficient data management, lack of quality curation
 - › rights & ethics difficult to sort out
 - › lack of trained people at all layers
- › results hardly reproducible (disaster for science)
- › **data intensive work does not scale**
- › **too many are excluded**
- › **data challenges in industry not really different or?**

IoT Changes the Game

- › billions of smart devices create continuous and highly granular data streams
- › in many cases application of crowd sourcing model (who owns data?)
- › massive data re-use for different purposes – optimisation, decision taking, etc.
- › choices
 - › monopolies will hampering innovation at the end
 - › www with its pros and cons – www not made for DM
 - › concepts being discussed in RDA may stimulate
 - › Digital Object Architecture, Global Digital Object Cloud
 - › Type-Triggered Automatic Processing
 - › etc.

What do we urgently need?

- › dramatic reduction of costs through interoperability
 - › global interop is expensive – where to settle?
- › global forum of openly visible & sealed Digital Objects
 - › change basic principles of exchanging data
 - › separation of collection and transaction processes
- › a domain of typed DO to allow automation
 - › allow crawlers to find suitable data
- › new models for rights and ethics to govern open forum
 - › enable participation of owners and checks
- › neutral bottom-up platform as a trusted entity to drive evolving data ecosystem
 - › use RDA (constitution is similar to IETF)

Towards an Open Forum of Sealed Data

Sealing means:

- resolvable unique identifier for each DO (Handle)
and associating crucial information with it such as
- unique identifier for owner (person, organisation)
 - unique identifier for devices (IPv6, etc.)
 - a fingerprint of the DO (checksum, etc.)
 - a type ID (Handle)

**It's all there, let's try it out systematically
how to organize such a forum (RDA Plenary Berlin)**

**RDA is ready to
open up to other
sectors (industry)
beyond research!**



RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn -

www.linkedin.com/in/ResearchDataAlliance

Slideshare -

<http://www.slideshare.net/ResearchDataAlliance>

RDA Europe

Email - info@europe.rd-alliance.org

Twitter - [@RDA_Europe](https://twitter.com/RDA_Europe)

RDA US

Twitter - [@RDA_US](https://twitter.com/RDA_US)