Open Research Data to Support Sustainable Health Initiatives

24 April 2018, Brussels
WELCOME

Dr. Frederick Fenter
Frontiers Executive Editor
OPENING

Prof. Michel Goldman
Director Institute for Interdisciplinary Innovation in Healthcare - I3h
Chief Editor, *Frontiers in Medicine*
Top 10 Causes of Death

- Ischemic heart disease
- Stroke
- Pneumonia
- COPD
- Cancer
- Diabetes
- Alzheimer
- Diarrhea
- Tuberculosis
- Road Injury

Deaths in millions

WHO [link](http://bit.ly/1c9a3vO)
The determinants of non-communicable diseases

Genome

Microbiome

Exposome

Tobacco
Alcohol
Diet
Sleep quality
Drugs
Environmental exposures

.....
From precision medicine to precision health

Life-long risk assessment
Genome
Microbiome
Exposome

Data analytics
Machine learning
Predictive modeling
Digital twin

Recommendation
Coaching
Self care
Telemedicine
Doctor visit
Hospital care

Adapted from
Gambhir et al., Science Transl Med 2018, 10: eaao3612
Open Science is Critical to Move Health Research Forward

Transparency and dissemination of data and methods

Results Reproducibility

Translation in Health Outcomes
Supporting sustainable health research initiatives: education is key!

Educating for the 21st-Century Health Care System: An Interdependent Framework of Basic, Clinical, and Systems Sciences
Gonzalo, Jed D. MD, MSc; Haidet, Paul MD, MPH; Papp, Klara K. PhD; Wolpaw, Daniel R. MD; Moser, Eileen MD, MHPE; Wittenstein, Robin D. EdD; Wolpaw, Terry MD, MHPE

Academic Medicine: January 2017 - Volume 92 - Issue 1 - p 35–39
doi: 10.1097/ACM.0000000000000951

FIELD GRAND CHALLENGE ARTICLE

Education in Medicine: Moving the Boundaries to Foster Interdisciplinarity

Michel Goldman*
Institute for Interdisciplinary Innovation in healthcare, Université libre de Bruxelles, Brussels, Belgium
Thank you

www.i3health.eu
SESSION 1

Breakthroughs in data-intensive health research

Moderator: Charlotte Geerdink – SwissCore
SESSION 1

Prof. Paolo Vineis
Professor of Environmental Epidemiology
Imperial College London, UK;
Field Chief Editor, Frontiers in Public Health
Paolo Vineis
Imperial College London

“Big data”: integrating omic technologies into public health research
The first international beauty contest judged by “machines” was supposed to use objective factors such as facial symmetry and wrinkles to identify the most attractive contestants.

After Beauty.AI was launched this year, roughly 6,000 people from more than 100 countries submitted photos in the hopes that artificial intelligence, supported by complex algorithms, would determine that their faces most closely resembled “human beauty”.

But when the results came in, the creators were dismayed to see that there was a glaring factor linking the winners: the robots did not like people with dark skin.

The Guardian 27 January 2017
Why omic technologies? An exploration of the universe of known and unknown exposures that influence human health
The concept of the exposome refers to the totality of environmental exposures from conception onwards, and includes both external and internal components.

A MULTIPLICITY OF EXPOSURES - IARC Exposome-Explorer

- **350 environmental pollutants**
- **147 dietary compounds**

**PBDEs**
- BDE-17
- BDE-47
- BDE-99
- BDE-100
- BDE-153
- BDE-154
- BDE-183
- BDE-189
- BDE-190
- BDE-209

**PAHs**
- 1-Hydroxypyrrene

**PCDDs**
- OCDD
- 2,3,7,8-TCDD
- 2,3,7,8,9-HpCDD

**PCDFs**
- 2,3,4,6,7,8-HpCDF
- 2,3,4,7,8,9-HpCDF
- 2,3,7,8,9-HpCDF

**PCBs**
- PCB-126
- PCB-130
- PCB-153
- PCB-156
- PCB-170

**FATTY ACIDS**
- Oleic acid (cis-9)
- Linoleic acid (cis-9)

**CAROTENOIDS**
- Beta-Carotene
- Lutein

**POLYPHENOLS**
- Lycopene
- Quercetin
- Resveratrol

**Biomarkers for environmental pollutants**
- 350 biomarkers
- 7,342 concentration values
- 265 publications analyzed

**All biomarkers**
- 497 biomarkers
- 10,480 concentration values

- Chemistry
- Cohorts where measured
- Biospecimens
- Analytical methods
- Concentrations
- State of validation
- Correlations with exposures
- Confounding factors
- Available on-line
- Linked to other databases
«Pathway perturbation» - The 1940s and 1950s saw a burgeoning of the **randomized clinical trial** as a paradigm for causality, with emphasis on **single causal factors**. Today etiology is better understood as a **process** involving the complex interplay of numerous agents that act along several mechanistic pathways.  
**Casals-Casas and Desvergne, 2011**
<table>
<thead>
<tr>
<th>Study</th>
<th>Transcriptomics</th>
<th>Epigenetics</th>
<th>Proteomics</th>
<th>Metabolomics</th>
<th>Adductomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEM Basel</td>
<td>89 Raw</td>
<td>128 Raw/ Proc.</td>
<td>90 Raw/ Proc.</td>
<td>127 Raw/ Proc.</td>
<td>798 Raw</td>
</tr>
<tr>
<td>PEM Norwich</td>
<td>56 Raw</td>
<td>60 Raw/ Proc.</td>
<td>42 Raw/ Proc.</td>
<td>61 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>PEM Turin</td>
<td>76 Raw</td>
<td>127 Raw/ Proc.</td>
<td>85 Raw/ Proc.</td>
<td>127 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>PEM Utrecht</td>
<td>63 Raw</td>
<td>90 Raw/ Proc.</td>
<td>87 Raw/ Proc.</td>
<td>89 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>ENVIRONAGE</td>
<td>193 Raw</td>
<td>200 Raw/ Proc.</td>
<td>198 Raw/ Proc.</td>
<td>204 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>Piccoli+</td>
<td>N/A</td>
<td>99 Raw/ Proc.</td>
<td>97 Raw/ Proc.</td>
<td>100 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>Rhea</td>
<td>N/A</td>
<td>100 Raw/ Proc.</td>
<td>100 Raw/ Proc.</td>
<td>100 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>INMA</td>
<td>N/A</td>
<td>600 Raw/ Proc.</td>
<td>97 Raw/ Proc.</td>
<td>100 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>EPIC Turin</td>
<td>N/A</td>
<td>172 Raw/ Proc.</td>
<td>187 Raw/ Proc.</td>
<td>382 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>EPIC Varese</td>
<td>N/A</td>
<td>143 Raw/ Proc.</td>
<td>192 Raw/ Proc.</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>MCC</td>
<td>N/A</td>
<td>406 Raw/ Proc.</td>
<td>405 Raw/ Proc.</td>
<td>591 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>Asthma SAPALDIA</td>
<td>N/A</td>
<td>604 Raw/ Proc.</td>
<td>402 Raw/ Proc.</td>
<td>405 Raw/ Proc.</td>
<td>566 Raw</td>
</tr>
<tr>
<td>Asthma ECHRS</td>
<td>298 Raw</td>
<td>80 Raw/ Proc.</td>
<td>80 Raw/ Proc.</td>
<td>80 Raw/ Proc.</td>
<td>N</td>
</tr>
<tr>
<td>TAPAS</td>
<td>117 Raw</td>
<td>N/A</td>
<td>N/A</td>
<td>120 Raw/ Proc.</td>
<td>146 Raw</td>
</tr>
<tr>
<td>Oxford Street</td>
<td>316 Raw/ Proc.</td>
<td>N/A</td>
<td>N/A</td>
<td>360 Raw/ Proc.</td>
<td>404 Raw</td>
</tr>
<tr>
<td>PISCINA</td>
<td>86 Raw/ Proc.</td>
<td>N/A</td>
<td>120 Raw/ Proc.</td>
<td>120 Raw/ Proc.</td>
<td>134 Raw</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1,294</strong> (Raw)</td>
<td><strong>2,809</strong> (Raw &amp; Proc. samples)</td>
<td><strong>2,182</strong> (Raw &amp; Proc. samples)</td>
<td><strong>2,966</strong> (Raw &amp; Proc. samples)</td>
<td><strong>1,914</strong> (Raw files)</td>
</tr>
</tbody>
</table>
SAPALDIA

MWAS on adult-onset asthma

Mummichog’s pathway enrichment tests

Pathways enriched for adult-onset asthma

Search for MITM pathways

Linoleate metabolism for PM2.5 and UFP
Glycerophospholipid metabolism for UFP

Laboratory confirmation of chemical identities within the MITM pathways

Linoleate (m/z=281.2464; RT=7.283) was confirmed

EPIC

MWAS on air pollution

Mummichog’s pathway enrichment tests

Pathways enriched for air pollution

Search for MITM pathways

Fatty acid activation for PM2.5
Linoleate metabolism for PM2.5
Glycosphingolipid metabolism for UFP
Carnitine shuttle for NO2

Laboratory confirmation of chemical identities within the MITM pathways

Carnitine (m/z=162.1128; RT=0.601) and Stearoylcarnitine (m/z=428.373; RT=6.479) were confirmed

Validation**

Validation**
miRNA work in relation to air pollution shows that air pollutants impact several pathways via miRNA activation that in turn are relevant to the multi-organ toxicity of air pollution.

Pollutant-specific cmiRNAs associated with TRAP exposure. The figure shows the overlap as well as the specificity of the pollutant-specific cmiRNAs associated with exposure to NO2, UFP, PM2.5, BC and PM10 of the included subjects in Hyde Park and Oxford Street. Julian Krauskopf et al, submitted
Water quality in a swimming pool: metabolites from metabolomics show overlap, unlike for air pollutants
Fingerprints of exposures: exposures may leave characteristic fingerprints in body molecules, and facilitate risk assessment.

**CANCER ETIOLOGY**

**Mutational signatures associated with tobacco smoking in human cancer**

Ludmil B. Alexandrov,1,2,3,* Young Seok Ju,4 Kerstin Haase,5 Peter Van Loo,5,6 Iñigo Martincorena,7 Serena Nik-Zainal,7,8 Yasushi Totoki,9 Akihiro Fujimoto,10,11 Hidehiko Nakagawa,10 Tatsuhiko Shibata,9,12 Peter J. Campbell,7,13 Paolo Vineis,14,15 David H. Phillips,16 Michael R. Stratton7*

Tobacco smoking increases the risk of at least 17 classes of human cancer. We analyzed somatic mutations and DNA methylation in 5243 cancers of types for which tobacco smoking confers an elevated risk. Smoking is associated with increased mutation burdens of multiple distinct mutational signatures, which contribute to different extents in different cancers. One of these signatures, mainly found in cancers derived from tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of DNA editing by APOBEC cytidine deaminases and of an endogenous clocklike mutational process. Smoking is associated with limited differences in methylation. The results are consistent with the proposition that smoking increases cancer risk by increasing the somatic mutation load, although direct evidence for this mechanism is lacking in some smoking-related cancer types.

Tobacco smoking has been associated with cancer genome sequencing, we recently described.

Tobacco smoke as a mixture that leaves different signatures depending on the cancer site and possibly on the chemicals involved – e.g. PAH for lung cancer: signature 4

*Science, 4 November 2016*
Metabolomics in Young Finns: low vs high education (fully adjusted model)
The end of theory: the data deluge makes the scientific method obsolete

All our research is not really “agnostic” (though it uses untargeted omics).

There are explicit theories behind the study design and the observations (e.g. a theory of ageing; the theory of hallmarks of cancer ...)

Sometimes there are implicit theories that are taken for granted (e.g. the idea that richness “trickles-down” by virtue of the market economy, or the theory of the “rational agent” – Jean Tirole).
Big Data

The early emphasis was on the “3 Vs”: *volume* (increasingly large size of the data), *variety* (the diversity of types of data, e.g. from free text to remote sensing), and *velocity* (the rapid generation and flow of data).

Normandeau (2013) and others have added the growing importance of *veracity* (issues of data bias and cleanliness), *validity* (appropriate data for the intended use), and *volatility* (how long the data are valid, how long need to store) (Khoury 2014).
The distinction prediction vs explanation is crucial for informed consent. The issue is not that I do not want to release my personal information (to the extent that Zuckerberg stated that privacy no longer exists) but that I want to know what is the theoretical and practical framework in which my information is used.

I want to avoid manipulation (commercial, political and ideological). For example personal data may lead to false discoveries based on wrong premises such as racism.

It is complete wrong to believe that data speak by themselves (prediction is thin, explanation is thick!)
Thank you
SESSION 1

Dr. Lee Wen Hwa
Director, Disease Foundations Network, Strategic Alliances, Structural Genomics Consortium, Oxford, UK
Radical Open Science to accelerate drug discovery

Wen Hwa Lee PhD
Director, Patient & Disease Foundations
Strategic Alliances
Biomedical research is risk averse.

The (un)targeted cancer kinome.
Fedorov O, Müller S, Knapp S.

Too many roads not taken.
We don’t know enough human biology
Secrecy leads to redundancy and inefficiency

100
63
39
25
14
8
3.5
2
<2


Does it work?

$500-700 M, 5-7 years
A general model for open access/open source in early stage drug discovery

**CREATIVE COMMONS**

- Public–Private Partnership
- Public Domain

**Tools & Basic Knowledge**

- NOVEL Proteins only!
- Structure
- Chemistry
- Antibodies
- Screening
- Cell Assays

**DISCOVERY AND EXPLORATION**

- No patent
- No restriction on use
- Open access to tools and data.
- Target identification & validation

**DRUG DISCOVERY AND DEVELOPMENT**

- (re)Screening
- Lead Optimisation
- Pharmacology
- Metabolism
- Pharmacokinetics
- Toxicology
- Chemical development
- Clinical development
- Etc.

Facilitated by access to increased amount of information in the public domain


wenhwa.lee@thesgc.org
SGC at a glance

- Operations started in June 2004
- Government agencies, Wellcome Trust, charities & leading pharma companies
- In excess of $400M ($200M by industry)
- +300-strong team in Oxford, Toronto, Stockholm, Campinas, Chapel Hill & Frankfurt (plus extended network)
- Open Access Policy:
  - Promptly placing results, reagents and know-how in the public domain
  - SGC scientists never file patents
GSK informs SGC about Mitsubishi compound

Oxford and Harvard start collaboration

Co-publication of JQ1 probe (SGC; cancer) and I-BET probe (GSK; inflammation)

JQ1 distributed to 100+ labs

Brd4 linked to AML (Nature) MM (Cell)

Filippakopoulos et al., Nature 2010 (Dec)
Progressing towards patients!
Bromodomain inhibitors in clinical trials

Presently (November 2016):
- 22 clinical trials registered
- 12 Companies (3~4 began as startups, were acquired by big pharma)
- 13 different molecules
- Total investment by industry: $60,000
- Avg. cost per patient in clinical trial for oncology: $50
- Avg. number of patients in Phase I trial: 22

<table>
<thead>
<tr>
<th>Company</th>
<th>Compound</th>
<th>Named Indications</th>
<th>Stage</th>
<th>Initiated</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbvie</td>
<td>ABBV-075</td>
<td>Advanced Cancer; Breast Cancer; Non-Small Cell Lung Cancer; Acute Myeloid Leukemia; Multiple Myeloma</td>
<td>Phase I</td>
<td>2015</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Bayer</td>
<td>BAY1238097</td>
<td>Neoplasms</td>
<td>Phase I</td>
<td>2015</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Gilead</td>
<td>GS-5829</td>
<td>Solid Tumors; Lymphomas</td>
<td>Phase I</td>
<td>2015</td>
<td>Ongoing</td>
</tr>
<tr>
<td>GSK</td>
<td>GSK525762</td>
<td>Elapsed, Refractory Hematologic Malignancies; NUT Midline Carcinoma (NMC) and Other Cancers</td>
<td>Phase I/II</td>
<td>2013</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Merck (Oncoethix)</td>
<td>OTX015</td>
<td>Acute Myeloid Leukemia; NUT Midline Carcinoma; Triple Negative Breast Cancer; Non-small Cell Lung Cancer With Rearranged ALK Gene/Fusion Protein or KRAS Mutation; Castrate-resistant Prostate Cancer (CRPC); Pancreatic Ductal Adenocarcinoma</td>
<td>Phase I</td>
<td>2014</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Constellation</td>
<td>CPI-0610</td>
<td>Acute Leukemia, Myelodysplastic Syndrome, or Myelodysplastic/Myeloproliferative Neoplasms; Previously Treated Multiple Myeloma; Progressive Lymphoma</td>
<td>Phase I</td>
<td>2013</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Tensha</td>
<td>TEN-010</td>
<td>Acute Myeloid Leukemia and Myelodysplastic Syndrome; Solid Tumors</td>
<td>Phase I</td>
<td>2014</td>
<td>Ongoing</td>
</tr>
</tbody>
</table>

Box 1 – The Consortium’s high-quality outputs have catalysed the translation of basic discoveries into clinical trials.

<table>
<thead>
<tr>
<th>Company/Organisation</th>
<th>No. of trials registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbbVie</td>
<td>1</td>
</tr>
<tr>
<td>Bayer</td>
<td>1</td>
</tr>
<tr>
<td>Boehringer-Ingelheim</td>
<td>1</td>
</tr>
<tr>
<td>Bristol-Myers Squibb</td>
<td>1</td>
</tr>
<tr>
<td>Constellation/Roche</td>
<td>3</td>
</tr>
<tr>
<td>FORMA Therapeutics/Celgene</td>
<td>1</td>
</tr>
<tr>
<td>Gilead Sciences</td>
<td>3</td>
</tr>
<tr>
<td>GlaxoSmithKline</td>
<td>5</td>
</tr>
<tr>
<td>Hadassah Medical Organization*</td>
<td>1</td>
</tr>
<tr>
<td>Incyte Pharmaceuticals</td>
<td>1</td>
</tr>
<tr>
<td>Merck Sharp &amp; Dohme</td>
<td>2</td>
</tr>
<tr>
<td>OncoEthix/Merck</td>
<td>4</td>
</tr>
<tr>
<td>Orion</td>
<td>1</td>
</tr>
<tr>
<td>Plexxikon/Daiichi Sankyo</td>
<td>1</td>
</tr>
<tr>
<td>Resverlogix</td>
<td>7</td>
</tr>
<tr>
<td>Tensha Therapeutics/Roche</td>
<td>3</td>
</tr>
<tr>
<td>University of Texas*</td>
<td>1</td>
</tr>
<tr>
<td>Zenith Epigenetics</td>
<td>3</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>40</strong></td>
</tr>
</tbody>
</table>

List of organisations with clinical trials registered since the Consortium’s seminal work in 2011. Academic trials are marked with * (source: Thomson Reuters Cortellis)
Open Access enables **faster** target validation

Closed model
- Knowledge accumulation is linear, repetitive & slow
- Longer to de-risk
- Limited access to global academia
- Small, non-pooled funds

Open model
- Open access & freedom to operate ensures much quicker adoption
- Knowledge accumulation is exponential, less repetitive and fast
- Unlimited access to global academics
- Larger pooled funds

Lee WH, PLOS Biology 2015
Open to involve non-traditional partners and stakeholders

That’s not the right ring!

Erythromycin
I’M PATIENT
Open access enabling early-stage drug discovery by patient & disease foundations
Fibrodysplasia Ossificans Progressiva (FOP): From gene to clinical candidate

**Gene**
- **Apr 2006**: Gene causing FOP was discovered

**Mechanism**
- **Apr 2009**: Structure solved; mutations mapped

**Pre-clinical**
- **2012-2013**: New tool compounds & validation in animal models

**Clinical candidate**
- **2015**: Clinical study being designed

6 years

**Figures**
- A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva.
- Potent and selective activity.

**Alex N Bullock**
Only Open can enable different communities helping each other

Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma

Kathryn R Taylor, Alan Mackay, Nathalie Truffaux, Yaron S Butterfield, Olena Morozova, Cathy Philippe, David Castel, Catherine S Grasso, Maria Vinci, Diana Carvalho, Angel M Carcaboso, Carmen de Torres, Ofelia Cruz, Jaume Mora, Natacha Entz-Werle, Wendy J Ingram, Michelle Monje, Darren Hargrave, Alex N Bullock, Stéphanie Puget, Stephen Yip, Chris Jones & Jacques Grill

Affiliations | Contributions | Corresponding authors

Nature Genetics 46, 457–461 (2014) | doi:10.1038/ng.2925
Received 30 April 2013 | Accepted 21 February 2014 | Published online 06 April 2014

ACVR1 Mutations in DiPG: Lessons Learned from FOP

Kathryn R. Taylor1,2, Maria Vinci1,2, Alex N. Bullock3, and Chris Jones1,2

Abstract

Whole-genome sequencing studies have recently identified a quarter of cases of the rare childhood brainstem tumor diffuse intrinsic pontine glioma to harbor somatic mutations in ACVR1. This gene encodes the type 1 bone morphogenic protein receptor ALK2, with the residues affected identical to those that, when mutated in the germline, give rise to the congenital malformation syndrome fibrodysplasia ossificans progressiva (FOP), resulting in the transformation of soft tissue into bone. This unexpected link points toward the importance of developmental biology processes in tumorigenesis and provides an extensive experience in mechanistic understanding and drug development hard won by FOP researchers to pediatric neurooncology. Here, we review the literature in both fields and identify potential areas for collaboration and rapid advancement for patients of both diseases. Cancer Res. 74(17), 4565–70. © 2014 AACR.
Open Science creating economic benefits

“Stone-Man Syndrome”

Open Science Clinical Trial

Brennan, Bullock

Brain Tumour Charity

Open Science Pharma

M₄K Pharma
Making Medicine, not money

❖ world’s first open science drug discovery company M4K Pharma (Meds4Kids)
❖ the beneficiaries are ‘open science and the public good’
❖ aims to aggregate and align diffuse academic, foundation and industry research into a traditional drug discovery and development programme

Owen Roberts  Aled Edwards  M4K Founders

http://m4kpharma.com
Welcome to SGC’s Open Lab Notebooks

In a groundbreaking, "extreme open science" initiative, SGC scientists around the world are starting to share their laboratory notebooks live, online. We believe that making our research, data, and protocols available on a day-to-day basis will generate scientific ideas and discussions, avoid redundancy, foster collaborations, and accelerate progress. Our hope is that this

We introduce our experiments at opennotebook.thesgc.org with links to experimental details.

We provide all experimental details at zenodo.org

https://opennotebook.thesgc.org
Contact

Wen Hwa Lee, PhD
Director
Disease Foundations Network, Strategic Alliances
Structural Genomics Consortium
wenhwa.lee@thesgc.org

ABOUT THE SGC
The SGC is an international public–private partnership (UK charity number 1097737) that aims to carry out basic science of relevance to drug discovery, placing all information, reagents and know–how into the public domain without restriction.
SESSION 1

Dr. Michael Rebhan
Senior Investigator, Novartis Institutes for BioMedical Research, Basel, Switzerland
Text & Data Mining
& “health state” modeling

What will be the next wave in “human-centric” health (innovation)?

Role of “systems approaches” & agility?

Michael Rebhan, Ph.D. (Senior Investigator)
Novartis, Basel, Switzerland (Research)

Open Research Data to Support Sustainable Health Initiatives, Brussels, 24-Apr-2018
Health(care) in 10 years = ?

Pressures on the current health systems (sustainability despite aging populations)

But: we also have much progress in:
- New kinds of interventions
- Technology / AI / big data
- Agility culture in “digital” innovation
- Public-private partnership excitement ↑
- Value-based models & tools

Who really benefits from intervention X? When?

Science + Economy (real life!)

Patient heterogeneity
Personalization
Human-centric design
Global burden is shifting

The **Global Burden of Disease study** has recently released new data on global disease burden shifting, using IHME methodology for comprehensive, quantitative study of global disease trends that impact patients, providers etc. (WEF report)

- Since 1990, significant progress has been made on infectious diseases
- However, advances in life expectancy and well-being will be compromised, unless we learn how to cope with non-communicable, **chronic diseases (NCDs)**
- Almost \( \frac{3}{4} \) of deaths are now from NCDs, not even talking about HC resources, **quality of life, impact** on families / society, provider burn-out...

**Question:**
Are we set up to deal with it well, **across** institutions & disciplines?

Open access to models from real world data
[Web visualizations](#)
"Data Science" as "glue" between worlds

- Work with (big) data
- Programming skills
- Put data "in order"

Computer science

Math and statistics

Data science

Domain knowledge

Science

Medicine

Economics

Humanities

What is the problem?

Enable solid data-driven decisions
What is the ‘best time’ to intervene?

When is it too **early** or too **late** for intervention X, in a particular patient subgroup?

State 1 (healthy)  
$p_{1,2}$  
$p_{2,1}$

State 2 (disease)

State 3 (disease)

State 4 (disease)

**HSM = “health state modeling”**

**Genetic** stratification - limited value here? (not “dynamic”)

How can we learn faster, ‘above’ the **data silos**? (open)

**NEW:** Role of patient-centric data ecosystems, patient empowerment?

Example: model **clusters** of similar **patient journeys**
Open Data & Tools

Patient stratification
- biomarkers
- Preclinical learning (translatability!)
- Map interventions & value
- Tools for innovators

Open Science & the new ecosystem

Module integration ⇒ solutions

Re-usable Modules
therapies, biomarkers, outcome / digital tools

Feedback Loops!

Solutions
(locally deployed, continuously adapt/learn!)

FDA Pre-Cert
(risks / systems)

It’s a co-creation problem...

with all this complexity, how agile can it be?
Co-creation: can we agree on principles?

1. Focus on outcomes & **value**, e.g. pay only, if it works!
2. Build **feedback loops** for learning
3. Transparency
4. More **patient-centric**, earlier in R&D
5. Learn faster: **agile**, fail faster

**Emerging Consensus!** (the “new system”)

Current reality, on the ground:

- Territorial behaviors (“us and them”)
- Incentives are mis-aligned
- Career paths favor silo mindset & action

*build on IMI etc. (data & tools!)*
Agile loops: a design example

Questions / Hypoth.
e.g. biomarkers improve risk stratification

3-6 months / loop = agile
(Scrum principles ⇒ fit?)

Digital tools to track patient status
- Multi-dimensional value capture incl. QoL
- E.g. using smartphones / tablets (skin)
- Physician monitors clin. data meaning
- Patients are involved, understand/get value

MVP concept (startups)

b biomarkers in human fluid

Ulcer outcomes

Loop design 8-12 weeks!
Visual: “me”, on the HS map?

HSM-based “map”
- where am I?
- QoL of “next stage”? What can I do?

Chronic kidney disease
(stages 1-5, 5 = end-stage)

Diabetes complications
(e.g. foot, eyes, kidney)

Text mining pilot (OpenMinTeD / H2020)
• Extract initial maps from text
• Facilitate crowd-based refinement
• Enables agile loop learning (crowd)?

Contact: Mappet Walker (Frontiers, Open Science publisher)
Agile Loop Design Principles

- Avoid “waterfall” issues (80% of time invested in document writing, updating, ...), reduce to 1%

- **Simplicity** / clarity of goal of next iteration (3-6 months max., reduce the problem)

- **Human**-centric design

- **Value** is captured at the end of each iteration (multi-dimensional, clear to stakeholders)
  - For **patients**, e.g. QoL, mobility
  - For **providers**, e.g. pain point addressed
  - (For **payors**, e.g. cost efficiency)
Towards a systems approach for chronic diseases, based on health state modeling [version 1; referees: 3 approved, 1 approved with reservations]

Abstract

Rising pressure from chronic diseases means that we need to learn how to deal with challenges at a different level, including the use of systems approaches that better connect across fragments, such as disciplines, stakeholders, institutions, and technologies. By learning from progress in leading areas of health innovation (including oncology and AIDS), as well as complementary indications (Alzheimer's disease), I try to extract the most enabling innovation paradigms, and discuss their extension to additional areas of application within a systems approach. To facilitate such work, a Precision, P4 or Systems Medicine platform is proposed, which is centered on the representation of health states that enable the definition of time in the vision to provide the right intervention for the right patient at the right time and dose. Modeling of such health states should allow iterative optimization, as longitudinal human data accumulate. This platform is designed to facilitate the discovery of links between opportunities related to a) the modernization of diagnosis, including the increased use of omics profiling, b) patient-centric approaches enabled by technology convergence, including digital health and connected devices, c) increasing understanding of the pathobiological, clinical and health economic aspects of disease progression stages, d) design of new interventions, including therapies as well as preventive measures, including sequential intervention approaches. Probabilistic Markov models of health states, e.g. those used for health economic analysis, are discussed as a simple starting point for the platform. A path towards extension into other indications, data types and uses is discussed, with a focus on regenerative medicine and relevant pathobiology.
Basel as a co-creation hub for agile innovation

Precision Medicine / Digital Health Convergence

* Increasing focus on *agility* in health innovation, under the umbrella of [DayOne](#).swiss


* [DayOne](#) facilitates catalyst projects:
  - Health hacking (devices, patient empowerment)
  - Applications of blockchain / distributed ledgers
  - DigitalCenter (Open innovation for chronic diseases, systems approach, build loops)
  - pediatrics

Contact: [Thomas Brenzikofe](#)r (leads DayOne)
DayOne DigitalCenter: OI for NCDs

A few key points from our gathering in Brussels yesterday

- **Find win-win** (e.g. Basel & Bochum): Help each other learn on building “agile loops”
- Each hub has a special mix of strengths ⇒ **best practice**
- Learn how to better involve the public, co-create!
- What will **Europe’s** model for “big data” & society be
- Keep it simple, at least initially – each loop
- Invite the **IMI** network into co-creation
- Apply systems approach to **prevention**, incl. economics
- Continuous improvement of systems, risks? Engage with pre-Cert (FDA) on new models

**Initial focus:** a) diabetes complications incl. CKD, skin ulcers, b) early stages of dementia, c) pediatrics
Contributors

At yesterday’s workshop in Brussels

• Mappet Walker (Frontiers, Open Science movement)
• Lutz Groh (InDIG, Germany – new health market)
• Thomas Brenzikofer (DayOne.swiss, Basel – a hub)
• Armin Furtwängler (Boehringer, new models)
• Marco d’Angelantonio (eHealth & chronic diseases)
• Hilde Stevens (i3health, Brussels – education)
• Grigorios Tzortzis (iASIS, big data for PM / public health)
Panel discussion:
Breakthroughs in data-intensive health research

Moderator: Charlotte Geerdink – SwissCore

Prof. Paolo Vineis – Imperial College London
Dr. Lee Wen Hwa – Structural Genomics Consortium
Dr. Michael Rebhan – Novartis Institutes for BioMedical Research
CLOSING REMARKS

Dr. Frederick Fenter
Frontiers Executive Editor
SESSION 2

Dr. Samuel Kerrien
Section Manager Data and Knowledge Engineering
Blue Project/EPFL, Lausanne, Switzerland
Blue Brain Nexus

An open science solution for comprehensive FAIR data and knowledge management
Why Blue Brain Nexus?

- Rise of trans-disciplinary science
- Issue: domain specific data integration platforms
- Challenge: multiple domains, distributed data, enable discovery
- Implement FAIR data principles
What is Blue Brain Nexus?

Application Programming Interface

Security

FAIR Metadata Store

Data Management

FAIR Data Store

IT Infrastructure

3rd Party Data Stores

EUDAT

amazon S3

3rd Party FAIR Metadata Stores

UniProt

figshare

zenodo
What is a Knowledge Graph?

- It is a graph
- It is semantic
- It is smart
- It is alive
What is a Knowledge Graph?

- It is a graph
- It is semantic
- It is smart
- It is alive
What is a Knowledge Graph?

- It is a graph
- It is semantic
- It is smart
- It is alive

Rule: parent of parent is grand parent
What is a Knowledge Graph?

- It is a graph
- It is semantic
- It is smart
- It is alive

How to design your own Knowledge Graph with Nexus?

- An entity is defined through a schema (W3C SHACL)

- Property types can be:
  - Literals (int, float, string, ...)
  - Structured objects
  - Link to Ontology terms
  - Link to other Entities (e.g. provenance)

- Semantic constraints
  - Cardinality
  - Enumeration
  - Range
  - Entity Type/Ontologies
What is provenance?

Source: https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/
Why provenance?

- Tracking origin of data
- Tracking how data is used (derivation)
- Assessing the quality of data
- Trust
- Reproducibility
- Data and algorithm attribution
Using Blue Brain Nexus – step by step

1. Design Your Domain
   - Define your entities
   - Leverage provenance
   - Share your domain

2. Build Your Knowledge Graph
   - Import ontologies
   - Create and link entities
   - Grow your knowledge graph

3. Upload Data
   - Attach raw data to your entities
   - Manage data revisions

4. Set Privacy Settings
   - Manage platform users/groups
   - Secure your dataset and metadata

5. Search & Discover
   - Traverse your knowledge graph
   - Find relevant data
   - Discover new facts

6. Federate
   - Connect Nexus instances
   - Share data
   - Search seamlessly
What is FAIR data?

- Framework to ensure research data is effectively re-used

- A set of 4 principles to ensure that data is made:
  - Findable, Accessible, Interoperable, Re-usable

- Not a goal in itself but a key factor to:
  - Enable knowledge discovery for both humans and machines
  - Accelerate innovation

- Blue Brain Nexus implements all FAIR principles

2016: Nature Scientific Data / doi:10.1038/sdata.2016.18
Comprehensive FAIR data and knowledge management

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Re-usable</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Persistent identifiers</td>
<td>✓ High quality metadata</td>
<td>✓ Provenance</td>
<td>✓ Semantic search</td>
</tr>
<tr>
<td>✓ Provenance</td>
<td>✓ Knowledge Graph exploration</td>
<td>✓ Notification</td>
<td></td>
</tr>
</tbody>
</table>

Coming soon ...

- Federation
- Inference
- Notifications
Comprehensive FAIR data and knowledge management

Findable | Accessible | Interoperable | Re-usable
---|---|---|---
☑ Open API
☑ Secure
☑ Scalable
☑ Versioning
## Comprehensive FAIR data and knowledge management

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Re-usable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>☑ Supports controlled vocabularies</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>☑ Uses open standards for data and knowledge representation</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>☑ Enables 5-star Linked Data</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>☑ Open source &amp; open to extensions</strong></td>
<td></td>
</tr>
</tbody>
</table>
Comprehensive FAIR data and knowledge management

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Re-usable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓ Describe license terms of data and metadata</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓ Open schemas</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓ Capture and search provenance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓ Support data evolution using migrations</td>
</tr>
</tbody>
</table>
Who is using Blue Brain Nexus?
Conclusions

❖ Blue Brain Nexus is a comprehensive solution to support Open Science
  ❖ Data Governance
  ❖ Data Integration
  ❖ Data Management
  ❖ Data Publishing

❖ Blue Brain Nexus is Open Source
❖ We are building a community
SESSION 2

Dr. Christine Durinx
Associate Director
Swiss Institute of Bioinformatics, Lausanne, Switzerland
Identifying what matters in life (sciences)

Christine Durinx, Associate Director SIB & co-lead of ELIXIR Data Platform
Building a sustainable infrastructure for biological information across Europe
66% of databases have no or 1 year assured funding

Survey ELIXIR Preparation phase
201 Resource managers
15 Years - 326 public life science databases

23%

14%

62%

Attwood et al., 2015
Long-term sustainability of databases
Which databases are essential?
What are ELIXIR Core Data Resources?

- Fundamental importance
- Complete collections of generic value
- High levels of usage, scientific quality and service
A carefully chosen basket of indicators, reflecting the multiple facets of bioinformatics resources

Durinx et al. 2017
METHOD ARTICLE

Identifying ELIXIR Core Data Resources [version 2; referees: 2 approved]

Christine Durinx¹, Jo McEntyre², Ron Appel¹, Rolf Apweiler², Mary Barlow², Niklas Blomberg³, Chuck Cook², Elisabeth Gasteiger⁴, Jee-Hyub Kim², Rodrigo Lopez², Nicole Redaschi⁴, Heinz Stockinger¹, Daniel Teixeira¹, Alfonso Valencia⁵

https://f1000research.com/articles/5-2422/v2
## Current set of ELIXIR Core Data Resources

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArrayExpress</td>
<td>Europe PMC</td>
</tr>
<tr>
<td>CATH</td>
<td>Human Protein Atlas</td>
</tr>
<tr>
<td>ChEBI</td>
<td>IMEx Consortium (IntAct &amp; MINT)</td>
</tr>
<tr>
<td>ChEMBL</td>
<td>InterPro</td>
</tr>
<tr>
<td>EGA</td>
<td>PDBe</td>
</tr>
<tr>
<td>ENA</td>
<td>PRIDE</td>
</tr>
<tr>
<td>Ensembl</td>
<td>STRING db</td>
</tr>
<tr>
<td>Ensembl Genomes</td>
<td>UniProt</td>
</tr>
</tbody>
</table>

[https://www.elixir-europe.org/platforms/data/core-data-resources](https://www.elixir-europe.org/platforms/data/core-data-resources)
Life cycle – Data Resources

- **Emerging**
  - Quality standards
  - Guide and inform

- **Mature**
  - Monitoring of usage trends & scientific impact

- **Legacy**
Emerging requirement: Deposition Databases
To provide guidance about repositories for publishing open data in the life sciences
Current set of ELIXIR Deposition Databases

ArrayExpress  IntAct
BioModels     MetaboLights
BioSamples    PDBe
BioStudies    PRIDE
EGA
ENA
EVA

https://www.elixir-europe.org/platforms/data/elixir-deposition-databases
Long-term sustainability of databases
To design and implement an international plan for long-term sustainability

- Determine which data resources are of fundamental importance
- Eligible for shared international support
A global coalition to sustain essential data resources
Infrastructure model

Core Data Resources: funding allocation

Funding Agency A
Research grants

Funding Agency B
Research grants

Funding Agency C
Research grants

Funding Agency D
Research grants

Gabella et al. 2018
5 Funding agencies
0.9% of total spending for life science research grants sufficient for all core data resources

Gabella et al. 2018
RESEARCH ARTICLE

Funding knowledgebases: Towards a sustainable funding model for the UniProt use case [version 2; referees: 3 approved]

Chiara Gabella ID, Christine Durinx ID, Ron Appel

ELIXIR-Switzerland, SIB Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559.
SESSION 2

Dr. Ian Potter
Global Business Development Manager, Publishing & Associations
Clarivate Analytics, UK
Cutting-edge technologies & services for data-intensive health research

Knowledge management with the Web of Science™ Data Citation Index℠

Ian Potter

Open Research Data to Support Sustainable Health Initiatives
24th April 2018, Brussels
What kinds of content and data exist in the Web of Science?
The Web of Science content selection process

- ~3,500 journals submitted
- ~60% meet initial criteria
- most accepted titles will initially be added to ESCI for a period of evaluation

Scheduled re-evaluation

Emerging Sources Citation Index

- evidence of peer review
- ethical publication practices
- enriches or complements the Web of Science
- bibliographic information in English
- meets technical requirements for deliverables (xml/pdf)

10–12% meet criteria for SCIE/SSCI/AHCI

1~3,500 journals submitted

- ongoing citation

Web of Science

Arts & Humanities Citation Index

Science Citation Index – Expanded

Social Sciences Citation Index

Clariivate initial evaluation

Scheduled re-evaluation

Clariivate full evaluation

journal meets ESCI criteria

journal fails to meet ESCI criteria

journal remains in ESCI

journal submission

Clariivate initial evaluation

ongoing citation

ESCI selection criteria

~60% meet initial criteria

most accepted titles will initially be added to ESCI for a period of evaluation

Web of Science content selection process

Trust the difference
The Web of Science content selection process

- Evidence of peer review
- Ethical publication practices
- International editorial conventions
- Timeliness of publication
- Bibliographic information in English
- Meets technical requirements for deliverables (xml/pdf)
- Evidence of peer review*
- Ethical publication practices*
- International editorial conventions
- Timeliness of publication
- Bibliographic information in English*
- Meets technical requirements for deliverables (xml/pdf)*
- Is there scholarly demand for this material?
- How does it compare with similar journals in WoS?
- Will it complement or enrich WoS with novel content?*
- Does the journal target an international audience or specifically a regional audience?
- Is international representation among the authorship and editorial board appropriate for such a journal?
- Total citations
- Recent citation activity
- Authors and editorial board have been cited
- Integration of the journal into the literature over time

Web of Science
Trust the difference
What is in a Web of Science record?
Web of Science* record
Includes all key bibliometric information
- title and abstract
- authors
- affiliations
- publication details
- funding information
- subject category
- keywords
- datasets
- cited references

*Web of Science Core Collection
A Web of Science record
From the Core Collection (SCI-E)
Core record details

Associated data
in the Data Citation Index

Citation network
all documents citing this record / all documents cited by this record

Web of Science
Trust the difference
Funding agency acknowledgements, funding statements, etc.

Associated data links to data study, dataset or other data.
Linking data with literature
- > 7 million records in 350 high quality repositories
- Records built from descriptive metadata
- Cited references
- Standardized and robust citation formats
- Connects data with the literature
- Contextualizes data
- Helps measure of the contribution of data
- Credit for data producers
- Reuse of existing datasets / reproducibility
- 1900 onwards
Data Citation Index selection criteria

Data repository

Citation

Data Citation Index

ongoing evaluation

- persistence and stability – active, deposit policies
- funding statements – provenance and funding information
- peer review – not universal, but an indication of data quality
- age of material – long preservation of the data
- links to research literature

Data Citation Index

editorial content

- Science and technology
- Social sciences
- Arts & humanities
- repositories may have a focus or inter- and multi-disciplinary

international diversity

- diversity amongst contributors, producers, editors
- data with an international audience or
- a local or regional focus

Data Citation Index

---

Data Citation Index

Data Citation Index

- a recommended standard data citation format
- DataCite format (adopted by Clarivate)
- ability to link data repository content to the literature

Web of Science

Trust the difference
**Evaluation**

we have evaluated ~1500 repositories for inclusion in the Data Citation Index

**Inclusion**

over half have been selected but at the present time, a significant proportion do not meet our requirements or are not sufficiently advanced for inclusion
Search for data
- topic
- title
- author
- author identifiers (ORCID, RID)
- group author
- editor
- DOI
A Data Citation Index record
from a search for topic=CRISPR
A Data Citation Index record from a search for topic=CRISPR
Streptococcus thermophilus DGCC7710 Genome sequencing

From Repository: European Nucleotide Archive
Group Author(s): DuPont Nutrition and Health

European Nucleotide Archive
Source URL: http://www.ebi.ac.uk/ena/data/view/PRJNA213367
Viewed Date: 16 Jun 2015
Published: 2013
Document Type: Data study
Data Type: nucleotide sequencing information

Abstract
Streptococcus thermophilus is a food-grade lactic acid bacterium widely used in industrial fermentations for the production of dairy products such as yogurt and various cheeses (cheddar, mozzarella, emmental, pizza cheese...). Originally selected for its outstanding functional properties - short generation time, fast milk acidification, casein proteolysis, exopolysaccharide production, natural competence, and robustness to bacteriophage infection- S. thermophilus DGCC7710 (a.k.a. RD534) has become a model organism for the study of multiple CRISPR-Cas systems. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) loci are usually flanked by a heterogenous set of cas (CRISPR-associated) genes that together provide acquired immunity against invasive nucleic acids such as viruses and plasmids. While three types (I, II, III) of CRISPR-Cas systems have been defined throughout the bacterial and archaeal domains, DGCC7710 is a rare strain which has been shown to contain at least one system of each type: type I-E (CRISPR4), type II-A (CRISPR1 and CRISPR3), and type III-A (CRISPR2).

Categories / Classification
Research Areas: Genetics & Heredity
Web of Science Category: Genetics & Heredity
In addition, thermophiles, originally isolated for their outstanding growth properties—short generation time, fast milk acidification, casein proteolysis, exopolysaccharide production, natural competence, and robustness to bacteriophage infection—S. thermophilus DGCC7710 (a.k.a. RD534) has become a model organism for the study of multiple CRISPR-Cas systems. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) loci are usually flanked by a heterogenous set of cas (CRISPR-associated) genes that together provide acquired immunity against invasive nucleic acids such as viruses and plasmids. While three types (I, II, III) of CRISPR-Cas systems have been defined throughout the bacterial and archaeal domains, DGCC7710 is a rare strain which has been shown to contain at least one system of each type: type I-E (CRISPR4), type II-A (CRISPR1 and CRISPR3), and type III-A (CRISPR2).

**Categories / Classification**

**Research Areas:** Genetics & Heredity

**Web of Science Category:** Genetics & Heredity

**Taxonomic Data:**

<table>
<thead>
<tr>
<th>SUPER TAXA</th>
<th>TAXA NOTES</th>
<th>Organism Classifier</th>
<th>Organism Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microorganisms, Bacteria, Eubacteria</td>
<td>Bacteria, Eubacteria, Microorganisms</td>
<td>Gram-Positive Cocci</td>
<td>Streptococcus thermophilis</td>
<td>DGCC7710</td>
</tr>
</tbody>
</table>

**Document Information**

**Language:** English

**Accession Number:** DRCI:DATA201608007249642

**Other Information**

**Method:** Scope=single isolate; Material=DNA; Selection=genome

**Cited References in Data Citation Index:** 0

**See fewer data fields**

---

**Cited References**

**How to cite this Resource**

**Most recently cited by:**

- Barrangou, Rodolphe; Coute-Monvoisin, Anne-Claire; Stahl, Buffy; et al. Genomic impact of CRISPR immunization against bacteriophages. BIOCHEMICAL SOCIETY TRANSACTIONS (2013)
- Karvelis, Toutsoudis; Gasinat, Giedrius; Miksys, Algirdas; et al. crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. RNA BIOLOGY (2013)

**Use in Web of Science**

**Web of Science Usage Count**

0 0

**Last 180 Days**  Since 2013

**Learn more**

**This record is from:** Data Citation Index

**Suggest a correction**

If you would like to improve the quality of the data in this record, please suggest a correction.
### Citing articles for a Data Citation Index record

#### Web of Science

<table>
<thead>
<tr>
<th>Citation</th>
<th>Date</th>
<th>Times Cited</th>
<th>Usage Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Genomic impact of CRISPR immunization against bacteriophages</td>
<td>2013-01-01</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>2. miRNA and tracrRNA guide Cas9-mediated DNA interference in <em>Streptococcus thermophilus</em></td>
<td>2013-01-01</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>3. The Population and Evolutionary Dynamics of Bacteria and Viruses with CRISPR-Mediated Immunity</td>
<td>2013-01-01</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>4. In vitro reconstitution of cascade-mediated CRISPR immunity in <em>Streptococcus thermophilus</em></td>
<td>2013-01-01</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5. Phage mutations in response to CRISPR diversification in a bacterial population</td>
<td>2013-01-01</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Refine Results

- **Publication Years**
  - 2013
  - 2012
  - 2011
  - 2010

#### Web of Science Categories

- BIOSCIENCE REVIEW
- MICROBIOLOGY
- MULTIDISCIPLINARY SCIENCES
- MICROBIOLOGY APPLIED
- MICROBIOLOGY
- IMMUNOLOGY
- MICROBIOLOGY

#### Filter results by:
- Highly Cited (1)
- Open Access (3)
- Associated Data (1)

#### Web of Science

**Trust the difference**
1. **Genomic impact of CRISPR immunization against bacteriophages**
   By: Barrangou, Rodolphe; Coute-Monvoisin, Anne-Claire; Stahl, Tiffy; et al.
   *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* Volume: 411 Pages: 1383-1391
   Published: DEC 2013
   [Free Full Text from Publisher](https://www.sciencedirect.com/science/article/pii/S0006291X13005482)

2. **crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus**
   By: Karvelis, Tautvydas; Gasiunas, Giedrius; Miksys, Algirdas; et al.
   *RNA BIOLOGY* Volume: 10 Issue: 5 Special Issue: 51 Pages: 841-851
   Published: MAY 1 2013
   [Free Full Text from Publisher](https://www.sciencedirect.com/science/article/pii/S1547625411001327)
Best practices for data citation
**Standardized citation formats are important for data**
see FORCE11 Joint Declaration of data citation principles — [https://www.force11.org/datacitationprinciples](https://www.force11.org/datacitationprinciples)

<table>
<thead>
<tr>
<th>Data Citation Examples: Recommended</th>
<th>Data Citation Examples: Not Recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. PANGAEA. <a href="http://dx.doi.org/10.1594/PANGAEA.726855">http://dx.doi.org/10.1594/PANGAEA.726855</a></td>
<td>Irino &amp; Tada (2009). Chemical and mineral compositions of sediments from ODP Site 127-797. Published by PANGAEA [<a href="http://www.pangaea.de">www.pangaea.de</a>]</td>
</tr>
</tbody>
</table>
Elements of a data citation

Required elements follow the basic, discipline-agnostic data citation guidelines proposed by DataCite ([https://www.datacite.org/services/cite-your-data.html](https://www.datacite.org/services/cite-your-data.html)). Metadata elements needed for data citation include:

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author/Creator</td>
<td>Individuals or organizations that created or contributed to the data set; this metadata element is vital to guarantee attribution and credit for data contributor, and to provide metrics for their nontraditional scholarly output</td>
</tr>
<tr>
<td>Year</td>
<td>The year of “publication” of the data; when it is made publicly available, such as through deposition in a repository</td>
</tr>
<tr>
<td>Title</td>
<td>The title of the data object, which may differ from the title of the parent research paper/project</td>
</tr>
<tr>
<td>Publisher</td>
<td>The data repository that houses the data and/or the governing organization responsible for publishing, (i.e., making available) the data</td>
</tr>
<tr>
<td>Version</td>
<td>Dynamic data sets or those where new editions may be issued (such as with error corrections or new values) must employ proper version control to guarantee accuracy and uniqueness in data citation</td>
</tr>
<tr>
<td>Permanent Identifier</td>
<td>A unique and persistent identifier should be assigned; for example, a Digital Object Identifier (DOI); in Data Citation Index citations, this bibliographic element may take the form of a unique URL, databank accession number, or other permanent identifier such as Handle (hdl) (<a href="http://www.handle.net/">http://www.handle.net/</a>)</td>
</tr>
<tr>
<td>Permanent Identifier</td>
<td>A unique and persistent identifier should be assigned; for example, a Digital Object Identifier (DOI); in Data Citation Index citations, this bibliographic element may take the form of a unique URL, databank accession number, or other permanent identifier such as Handle (hdl) (<a href="http://www.handle.net/">http://www.handle.net/</a>)</td>
</tr>
</tbody>
</table>

**A unique identifier is required — permanency is better**

only about 30–40% of data objects currently have a DOI or other permanent identifier — the remainder have URLs
Streptococcus thermophilus DGCC710 Genome sequencing

From Depository: European Nucleotide Archive

Group Author(s): DelPont Nutrition and Health

European Nucleotide Archive

Source URL: http://www.ebi.ac.uk/ena/data/view/PRJN42907

Viewed Date: 10 Jun 2015

Published: 2015

Document Type: Data study

Data Type: nucleotide sequencing information

Abstract

Streptococcus thermophilus is a key grade lactic acid bacterium widely used in industrial fermentations for the production of dairy products such as yogurt and various cheeses (cheddar, mozzarella, environal, pizza cheese...). Originally selected for its outstanding functional properties—short generation time, fast acidification, casein-proteolysis, exopolysaccharide production, natural competence, and robustness to bacteriophage infection—it has become a model organism for the study of multiple CRISPR-Cas systems. DGCC710 (Customarily regularly expressed Short Palindromic Repeat) loci are usually found by a heterogeneity test of our CRISPR-associated genes that together provide acquired immunity against invasive nucleic acids such as plasmids and phages. While three loci (K, J, I of CRISPR) CRISPR-Cas systems have been defined throughout the bacterial and archaeal domains, DGCC710 is a core system that has been chosen to contain at least one system of each type I (CRISPR/Cpf1), type II (CRISPR/Cas9), and type III (CRISPR) CRISPR/Cas systems.

Categories / Classification

Research Areas: Genetics & Mendelism

Web of Science Category: Genetics & Mendelism

Taxonomic Data

<table>
<thead>
<tr>
<th>SUPERKINGDOM</th>
<th>PHYLUM</th>
<th>CLASS</th>
<th>ORDER</th>
<th>FAMILY</th>
<th>GENUS</th>
<th>SPECIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microorganisms</td>
<td>Bacteria, Eubacteria</td>
<td>Bacilli, Eubacteria</td>
<td>Lactobacillaceae</td>
<td>Streptococcus</td>
<td>Thermophilus</td>
<td>DGCC710</td>
</tr>
</tbody>
</table>

Document Information

Language: English

Accession Number: 10010000057274H42

Other Information

Method: Sequencing single isolates (manual Sanger, Sequenase) Dna A, dideoxy-nucleotide chain termination

Cited References in Data Citation Index: 0

See fewer data fields
A4. Document has been shown that short antisense nucleic acids such as viruses and plasmids. While there are few edited genomes, such as yogurt and cheese, the piz genes (CRISPR-associated) have been shown to be involved in the production of yogurt and cheeses. There are 16 new products in the CRISPR-associated system, which are described as: type I-E (CRISPR4), type II-A (CRISPR1), and type III (CRISPR3). These systems consist of regular and a rare CRISPR-Cas systems, which are described as: Indi. (2013) CRISPR-Cas systems. European Nucleotide Archive. http://www.ebi.ac.uk/ena/data/view/PRJNA213367

Clarivate Analytics recommends citing this resource as:
curate and validate metadata for completeness, accuracy, and consistency
issue permanent IDs for data objects
provide unique landing pages for data objects
practice versioning and maintain detailed update information
indicate data resource type in metadata
ensure clear attribution for data objects
do document the aim of the repository and its inclusion policies
create and publish guidelines for data deposition
develop and promote formal data citation policies
state and enforce requirements for data sharing and citation
establish metadata criteria to allow persistent and unique identification of data in citations
cite papers that describe the data in addition to the data itself
cite in the bibliography or specific acknowledgement, not as footnotes
formal citation lets us link and track data citations in the Web of Science
cite in a recognized style (e.g. DataCite or Clarivate)
be specific, cite the exact version of any dataset – provide any criteria (e.g. timeframes) applied
cite data at the finest level of granularity available – provide more information in the text if necessary
always include identifiers where available (DOI, repository-assigned ID)
consider data as a primary record of research – cite and be cited
The Data Citation Index in numbers
Subject categories

- Life sciences: 39%
- Physical sciences: 22%
- Social sciences: 22%
- Multidisciplinary: 13%
- Arts & humanities: 3%
- Technology: 2%

Global coverage

- North America: 50%
- Europe: 38%
- Asia: 4%
- Australia: 2%
- Africa: 2%
- Worldwide: 1%

Note that geographic coverage does not represent geographic origin of the data objects — this the location of the repository.

95% open access
Artificial Intelligence in Open Access publishing

Mattia Albergante, Lead Product Manager
Frontiers’ Mission

“to build an Open Science platform where everybody has equal opportunity to seek, share and generate knowledge, and that empowers researchers in their daily work.”
Frontiers’ Mission

“to build an Open Science platform where everybody has equal opportunity to seek, share and generate knowledge, and that empowers researchers in their daily work.”
Frontiers’ Mission

“to build an Open Science platform where everybody has equal opportunity to seek, share and generate knowledge and that empowers researchers in their daily work.”
Frontiers’ Mission

“to build an Open Science platform where everybody has equal opportunity to seek, share and generate knowledge, and that empowers researchers in their daily work.”
13 Tasks require your attention

12 Invite an Associate Editor
1 Evaluate recommendation for rejection

75 Manuscripts in review

- Editorial Assignment: 20 on time, 15 delayed
- In Independent Review: 20 on time, 16 delayed
- In Interactive Review: 31 on time, 13 delayed
- Review Finalized: 4 on time, 3 delayed
- Final Validation: 0 on time, 0 delayed
YOUR TASKS

13 Tasks require your attention

12 Invite an Associate Editor

1 Evaluate recommendation for rejection
Frontiers’ Mission

“to build an Open Science platform where everybody has equal opportunity to seek, share and generate knowledge, and that empowers researchers in their daily work.”
The Frontiers peer-review
The Frontiers peer-review

Submission
Independent Review
Interactive Review
Decision
Final Validation
Published

Reviewer selection
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar, Sangeetha Rajendran, Shakti Lalshram, Shalini Anandan, Veeraraghavan Babji* and Indranil Biswas**

*Department of Clinical Microbiology, Christian Medical College, Vellore, India. **Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

Acinetobacter baumannii is a nosocomial pathogen involved in various infections ranging from minor soft-tissue infections to more severe infections such as ventilator-associated pneumonia and bacteremia. The severity and the type of infections depend on the genetic and phenotypic variations of the strains. In this study, we compared the extent of biofilm formation and motility displayed by 60 multidrug-resistant A. baumannii clinical strains isolated from blood and sputum samples from patients from Southern India. Our results showed that isolates from the sputum samples formed significantly more robust biofilm compared to the blood isolates. On the other hand, we observed that the blood isolates were more motile than the sputum isolates. To the best of our knowledge, this is the first study that systematically evaluated the correlation between these two phenotypic traits and the nature of the isolates.

Keywords: Acinetobacter baumannii, biofilm, motility, multidrug-resistance, India

INTRODUCTION

Acinetobacter baumannii is a nosocomial pathogen that can cause a wide array of infections ranging from minor skin and soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP typifies serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii induced VAP is between 40 and 70% (1, 2). The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections (3). The pathogen primarily enters into the bloodstream through lower respiratory tract infections and intravascular devices (4-7). Wound and urinary tract infections also lead to bloodstream infections (8). Like VAP, the risk factors for bloodstream infections include among others immunosuppression, colonization with A. baumannii, and invasive procedures (6-8). The mortality rates associated with the A. baumannii bloodstream infections ranges between 28 and 48%; however, the issue is highly debatable (3, 4, 9-11).

The pathogen’s ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for health-care-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii is an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including aminoglycosides,
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar1, Sangeetha Rajendran1, Shaldi Lalishram1, Shalini Anandar, Veeranraghavan Balaji2* and Indrani Biswas2**

1Department of Clinical Microbiology, Christian Medical College, Vellore, India, 2Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

Acinetobacter baumannii is a nosocomial pathogen involved in various infections ranging from minor soft-tissue infections to more severe infections such as ventilator-associated pneumonia and bacteremia. The severity and the type of infections depend on the genetic and phenotypic variations of the strains. In this study, we compared the extent of biofilm formation and motility displayed by 60 multidrug-resistant A. baumannii clinical strains isolated from blood and sputum samples from patients from Southern India. Our results showed that isolates from the sputum samples formed significantly more robust biofilm compared to the blood isolates. On the other hand, we observed that the blood isolates were more motile than the sputum isolates. To the best of our knowledge, this is the first study that systematically evaluated the correlation between these two phenotypic traits and the nature of the isolates.

Keywords: Acinetobacter baumannii, biofilm, motility, multidrug-resistance, India

INTRODUCTION

Acinetobacter baumannii is a nosocomial pathogen that can cause a wide array of infections ranging from minor skin and soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP typifies serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii induced VAP is between 40 and 70% (1, 2). The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections (3). The pathogen primarily enters into the bloodstream through lower respiratory tract infections and intravascular devices (4–7). Wound and urinary tract infections also lead to bloodstream infections (8). Like VAI, the risk factors for bloodstream infections include among others immunosuppression, colonization with A. baumannii, and invasive procedures (5–8). The mortality rates associated with the A. baumannii bloodstream infections range between 28 and 48%; however, the issue is highly dreadful (3, 4, 9–11).

The pathogen’s ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for healthcare-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii is an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including aminoglycosides,
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar¹, Sangeetha Rajendran², Shakti Lalshnam³, Shalini Anandar⁴, Veeranagghan Balaji⁵* and Indranil Biswas⁵*

¹Department of Clinical Microbiology, Christian Medical College, Vellore, India, ²Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

OPEN ACCESS

Edited by:
Rausen Arnbak,
Technical University of Denmark, Denmark

Revised by:
Lorenzo Grasso,
Harvard Medical School, USA
Damarvis Mahan,
Carnegie Mellon,
Harvard Medical School, USA

Correspondence:
Veeranagghan Balaji
vbalaji@cmc.ac.in
Indranil Biswas
biswas@ckmc.edu

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Public Health

Received: 23 January 2016
Accepted: 06 May 2016
Published: 24 May 2016

Keywords: Acinetobacter baumannii, biofilm, motility, multidrug resistance, India

INTRODUCTION

Acinetobacter baumannii is a nosocomial pathogen that can cause a wide array of infections ranging from minor skin and soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP typifies serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii induced VAP is between 40 and 70% (1,2). The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections (3). The pathogen primarily enters into the bloodstream through lower respiratory tract infections and intravascular devices (4, 5). Wound and urinary tract infections also lead to bloodstream infections (6). Like VAP, the risk factors for bloodstream infections include among others immunosuppression, colonization with A. baumannii, and invasive procedures (6, 7). The mortality rates associated with the A. baumannii bloodstream infections ranges between 28 and 47%; however, the issue is heavily debated (3, 4, 6–11).

The pathogen's ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for health-care-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii is an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including aminoglycosides,
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayan Kumar, Sangeetha Rajendran, Shakti Lalshah, Shalini Anandar, Veeranagavan Balaji**, and Indrani Biswas**

*Department of Clinical Microbiology, Christian Medical College, Vellore, India; **Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

Acinetobacter baumannii is a nosocomial pathogen involved in various infections ranging from minor soft-tissue infections to more severe infections such as ventilator-associated pneumonia and bacteremia. The severity and the type of infections depend on the genetic and phenotypic variations of the strains. In this study, we compared the extent of biofilm formation and motility displayed by 60 multidrug-resistant A. baumannii clinical strains isolated from blood and sputum samples from patients from Southern India. Our results showed that isolates from the sputum samples formed significantly more robust biofilm compared to the blood isolates. On the other hand, we observed that the blood isolates were more motile than the sputum isolates. To the best of our knowledge, this is the first study that systematically evaluated the correlation between these two phenotypic traits and the nature of the isolates.

Keywords: Acinetobacter baumannii, biofilm, motility, multidrug resistance, India

INTRODUCTION

Acinetobacter baumannii is a nosocomial pathogen that can cause a wide array of infections ranging from minor skin and soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP, typical serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii-induced VAP is between 40 and 70% (1, 3). The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections (3). The pathogen primarily enters into the bloodstream through the lower respiratory tract infections and intravascular devices (4–7). Wound and urinary tract infections also lead to bloodstream infections (3). Like VAP, the risk factors for bloodstream infections include among others, immunosuppression, colonization with A. baumannii, and invasive procedures (5–7). The mortality rates associated with the A. baumannii bloodstream infections ranges between 20 and 43%; however, the issue is highly debatable (3, 4, 5–7).

The pathogen’s ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for healthcare-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii are an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including amoxicillin,
Why Open Access
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar1, Sangeetha Rajendran1, Shakti Lalsham1, Shalini Anandani2, Veeranagahan Balaji2* and Indrand Balasubramaniyan2*

1Department of Clinical Microbiology, Christian Medical College, Vellore, India. 2Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

Introduction

Acinetobacter baumannii is a nosocomial pathogen involved in various infections ranging from minor soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP typifies serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii-induced VAP is between 40 and 70% (1–3). The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections (3). The pathogen primarily enters into the bloodstream through lower respiratory tract infections and intravascular devices (4–7). Wound and urinary tract infections also lead to bloodstream infections (8). Like VAI, the risk factors for bloodstream infections include among others immunosuppression, colonization with A. baumannii, and invasive procedures (5–8). The mortality rates associated with the A. baumannii bloodstream infections range between 28 and 45%; however, the issue is highly debatable (3, 4, 9–11).

The pathogen’s ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for healthcare-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii is an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including aminoglycosides,
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar1, Sangeetha Rajendran1, Shakti Lalitharam1, Shalini Anandaran1, Veeranaghravan Balaji* and Irindri Blassaw1*
1Department of Clinical Microbiology, Christian Medical College, Vellore, India, 2Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS, USA

**Correspondence:**
Veeranaghravan Balaji
vbalaji@cmc-vellore.ac.in
Irindri Blassaw
blassaw@kumc.edu

**SPECIALTY SECTION:**
Infectious Diseases; a sector of the journal
Frontiers in Public Health

**EDITORIAL:**
This article was submitted to Infectious Diseases, a sector of the journal
Frontiers in Public Health

**RECEIVED:**
21 January 2016

**ACCEPTED:**
8 May 2016

**PUBLISHED:**
24 May 2016

**CITATION:**

**KEYWORDS:** Acinetobacter baumannii, biofilm, motility, multidrug resistance, India

Introduction

Acinetobacter baumannii is a nosocomial pathogen that can cause a wide array of infections ranging from minor skin and soft-tissue infections to more severe invasive diseases, such as bacteremia, meningitis, and ventilator-associated pneumonia (VAP). VAP typifies serious hospital-acquired infections due to colonization of A. baumannii in the airway via environmental exposure. The mortality rate associated with A. baumannii induced VAP is between 40 and 70% [1, 3]. The patients with the highest mortality tend to be older, immunocompromised, have prolonged intubation, and are at a greater risk of infection by other pathogens. In the intensive care setting, A. baumannii also causes serious bloodstream infections [3]. The pathogen primarily enters into the bloodstream through lower respiratory tract infections and intravascular devices (4–7). Wound and urinary tract infections also lead to bloodstream infections (8). Like VAP, the risk factors for bloodstream infections include among others immunosuppression, colonization with A. baumannii, and invasive procedures (6–8). The mortality rates associated with the A. baumannii bloodstream infections range between 28 and 48%; however, the issue is a highly debatable (3, 4, 6–11).

The pathogen's ability to survive and to persist for extended periods of time on surfaces makes it a frequent cause for health-care-associated infections. Moreover, emergence and spread of multiple drug resistance (MDR) A. baumannii is an area of great clinical concern. A. baumannii is becoming resistant to most of the commonly used antibiotics, including aminoglycosides,

30x information richness [1]

...Biofilm Formation and Antibiotic Tolerance...

promotes biofilm formation and increases antibiotic tolerance to...

Bacterial programmed cell death and quorum sensing are direct examples of prokaryote...

Laurence Rahme
Harvard Medical School
Knowledge graph of 100 authors, approx. 5'000 publications
NF-kBp50 and HDAC1 Interaction Is Implicated in the Host Tolerance to Infection Mediated by the Bacterial Quorum Sensing Signal 2-Aminoacetophenone

Arunava Bandypadhyaya • Amy Tsurumi • Laurence G. Rahme

Frontiers in Microbiology
Biofilm Formation and Motility Depend on the Nature of the Acinetobacter baumannii Clinical Isolates

Saranya Vijayakumar · Sangeetha Rajenderan · Shakti Laishram · Shalini Anandan · Veeraraghavan Balaji · Indranil Biswas

Frontiers in Public Health
Published on 24 May 2016

3,486 views  View Impact
Mattia Albergante

ORCID ID
https://orcid.org/0000-0002-8406-3986

Work (18 of 18)

Sparse representation of signals: From astrophysics to real-time data analysis for fusion plasmas and system optimization analysis for ITER and TCV
Plasma Physics and Controlled Fusion
2016 | journal-article
DOI: 10.1088/0741-3335/58/12/123001
EID: 2-s2.0-84997848466
Source: Scopus - Elsevier

Neutron spectroscopy measurements of tritium beam transport at JET
Nuclear Fusion
2014 | journal-article
Gaetano Santulli

Columbia University - Present
University of Naples Federico II

Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy

Frontiers

BIO
Gaetano Santulli, MD, PhD

RESEARCH FIELDS
MEDICAL AND HEALTH SCIENCES

EDITORIAL BOARD MEMBERSHIPS

CURRENT MEMBERSHIPS

Frontiers in Endocrinology
European Medical Journal

Journal of Aging and Gerontology
Cellular and Molecular Life Sciences
Initiative for Open Citations

The Initiative for Open Citations (I4OC) is a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data.

About

Citations are the links that knit together our scientific and cultural knowledge. They are primary data that provide both provenance and an explanation for how we know facts. They allow us to attribute and credit scientific contributions, and they enable the evaluation of research and its impacts. In sum, citations are the most important
We use artificial intelligence, machine learning and data science to bring different types of data together to set research in context and enable serendipity.

Learn more about Dimensions
Open Science
SESSION 2

Dr. Lucas Anastasiou
Project Officer, Knowledge Media Institute (KMI)
The Open University, UK
A bit about me

• Research Software Engineer (RSE)
• Part of research team at Knowledge Media institute, The Open University
• Develop, build, maintain and operate CORE (https://core.ac.uk)
• Participation in TM related projects (OpenMinTed, FosterPlus, FIT4RRI, Jisc dedup, Jisc analytics dashboard)
What is CORE

- [https://core.ac.uk](https://core.ac.uk)
- An Open Access scientific literature aggregator
- Collects (aggregates) scholarly data from >3500 data sources (IRs, journals, subject repositories)
- At the moment: 125 million metadata records
- 11m full text records
- World’s largest full text OA dataset
- Content Enrichments
- Set of services on top

(“Google scholar for Open Access“)
Mission of CORE

Aggregate all open access content distributed across different systems worldwide, enrich this content and provide access to it through a set of services ...
Value of aggregation

- Reduce the time to access information
- Standardise and harmonise content from many providers
- Enrich content with new information
- Provide harmonised access to users
- Enable the discovery of new information

[Diagram]

Typically little or no support
Aggregating OA content

- Up to **90%** of a typical text-miner’s time is spent on gathering and harmonising the data. This is wasteful!
- Large scale aggregations are expensive, so can hardly be maintained professionally by individual text-miners => professional maintenance and wide sharing of the aggregated data needed.
- Reduce costs and enable the quick start of TDM projects.
- Enable the deployment of TDM solutions as services in real-world applications.
Why Text Mining on scientific literature

- Literature Based Discovery
- Support exploratory access to research literature
- Summarisation of research findings
- Question answering and semantic search
- Monitor research trends
- Understand how to direct research funding
- Plagiarism detection
- Automate peer review

“Research papers are the most complete representation of human knowledge.”
TDM scientist workflow

CORE services

• API
  • Fully RESTful API for interactive, transactional access to the corpus
• Datasets
  • Processed, ready to TM dataset
• FastSync
  • Fast enterprise incremental access to CORE’s data
• Publisher Connector
  • Seamless access to OA articles from non-standard publishers systems
• Recommender
  • Provision of relevant/similar publications
• Dashboard
  • Allows IR managers to control their CORE content
OpenMinTed

• A platform to foster TDM over scientific literature
• Brings together data sources otherwise scattered all around
• Address the lack of interoperability of TDM tools
• Portal for TDM experts but supports non-expert users to run prepared workflows
• Ready workflows to use
• Launching on 24th May – join us!
OpenMinTed platform

Components and Applications Providers

Maven Central

docker

Repos/Sites Web Services

Corpus Building Process

Scholarly Content

Ontology

ML Model

Term Lexicon

Typesystem

Component 1

Component 2

... Component N

Processed Output

Community Repos

OpenAIRE

CORE

Knowledge Resources Providers

Maven Central
OpenMinTeD workflow editor
One number to take away

(up to) 90% time on a TM task is spent to acquire content and “plumbing” around

90% Of content is closed (does it imply that 90% of possible apps is also closed?)

90 Tb (as of 2018) to store all of OA literature (and that is just the OA part!)
Thank you!

Lucas Anastasiou
Project Officer - Open Access Publishing
Knowledge Media Institute,
The Open University

http://core.ac.uk/about
twitter.com/oacore
Panel discussion: Cutting-edge technologies and services for data-intensive health research

Moderator: Dr. Frederick Fenter – Frontiers

Dr. Samuel Kerrien – Blue Brain Project/EPFL
Dr. Ian Potter – Clarivate Analytics
Dr. Lucas Anastasiou – The Open University
Dr. Christine Durinx – Swiss Institute of Bioinformatics
Dr. Mattia Albergante – Frontiers
Achieving open research data in health – overcoming the EU policy, institutional, and regulatory challenges

*Moderator:* Dr. Monica Dietl – Senior Advisor Science|Business
SESSION 3 – OPENING KEYNOTE

Dr. Cornelius Schmaltz
Head of Unit for Strategy, European Commission
DG Research and Innovation, Brussels, Belgium
‘Open Research Data to Support Sustainable Health Initiatives’

Session 3: Policy, regulatory and institutional challenges & actions at EU level to support open research data in health

Brussels, 24 April 2018

Ellas PAPADOPOULOU
Cornelius SCHMALTZ, MD (Head of Unit)
RTD.E.1 Health Research - "Strategy"
European Commission
April 24th, 2018, Brussels
Why Open Science?

Open Science = Systemic transition of science system which affects the way:

- Research is performed
- Knowledge is shared/diffused/preserved
- Research projects/results are evaluated
- Research is funded
- Researchers are rewarded
- Future researchers are trained
Why Open Science?

Open access to research data: right to access and re-use research data

Opening up research data has the potential to:
• improve the quality of scientific results
• avoid unnecessary duplications
• involve societal actors
• and significantly contribute to economic growth (through open innovation)
Why Open Science?

May 2016, Competitiveness Council conclusions:
‘The transition towards an Open Science system’

“Open Science has the potential to increase the quality, impact and benefits of science and to accelerate advancement of knowledge by making it more reliable, more efficient and accurate, better understandable by society and responsive to societal challenges, and has the potential to enable growth and innovation through reuse of scientific results by all stakeholders at all levels of society, and ultimately contribute to growth and competitiveness of Europe”.
"Open Science is a new paradigm that emphasizes the sharing of research early in the process. Why is this important? Because data is the fuel of Open Science and making data FAIR (findable, accessible, interoperable and re-useable) is crucial to making the engine of Open Science work."

Source: BLOG POST By Carlos Moedas 29 November 2017
Open Science and Health Research Data

- The integration of fragmented information systems into the clinical life cycle => discovery of relevant associations
- Clinical trials/interoperable health records => easier to find suitable participants and to design and assess the feasibility of new studies
- A more systematic identification of drug safety signals => personalized medicine analyses via appropriate patient and/or population stratification methodologies.
- Benefits for translational research into health and well-being => improve models of common disease to better understand the progression of rare diseases
Open Science and Health Research Data

- Render health research data FAIR, Findable, Accessible, Interoperable and Re-usable and avoid the duplication of efforts => gain time for patients, avoid waste of precious patients’ efforts (clinical trials)
- Facilitate and enhance the validation/verification of health research results
- Increase collaborative methods and technological progress in health research applications
- Opportunities for enhanced personalized prevention, prognosis, diagnosis and treatment
- Efficiency and effectiveness of prediction and prevention strategies or of medical interventions, health services, and health policies
From Challenges to Policy...

Holistic Policy Agenda – Scope & Ambitions

... 4 policy ambitions with regard to the use & management of research results and data

✓ **Open Data**: FAIR data sharing is the default for funding scientific research

✓ **Science cloud**: All EU researchers are able to deposit, access and analyse European scientific data through the open science cloud, without leaving their desk

✓ **Altmetrics**: Alternative metrics (next generation metrics) to complement conventional indicators for research quality and impact (e.g. Journal Impact Factors and citations)

✓ **Future of scholarly communication**: All peer reviewed scientific publications are freely accessible
From Challenges to Policy...

Holistic Policy Agenda – Scope & Ambitions

4 policy ambitions with regard to **relations with research actors (researchers, institutions and funders)**

- **Rewards**: The European research career evaluation system fully acknowledges Open Science activities
- **Research Integrity**: All publicly funded research in the EU adheres to commonly agreed Open Science Standards of Research Integrity
- **Education and skills**: All young scientists in Europe have the necessary skills and support to apply Open Science research routines and practices
- **Citizen Science**: CS significantly contribute and are recognised as valid knowledge producers of European science
From Policy to Action...

- The **H2020 Open Research Data Pilot (ORDP)** aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects.

- The ORDP takes into account the **need to balance openness and protection of scientific information**, commercialisation and IPR, privacy concerns, security as well as data management and preservation questions.

- As of **01 January 2017** the ORDP is applicable by default to Health research H2020 funded projects.
From Policy to Action…

Art. 29.3 of H2020 MGA – extract

[OPTION 1a for actions participating in the Open Research Data Pilot: Regarding the digital research data generated in the action (‘data’), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible;

[...]

(iii) other data, including associated metadata, as specified and within the deadlines laid down in the ‘data management plan’ (see Annex 1);

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves)
Participation in the ORD pilot is not part of the evaluation of proposals.

"[...] However, good research data management as such should be addressed under the impact criterion, as relevant to the project."

A Data Management Plan (DMP) should include information on:
- the handling of research data during and after the end of the project;
- what data will be collected, processed and/or generated which methodology and standards will be applied;
- whether data will be shared/made open access and;
- how data will be curated and preserved (including after the end of the project).
From Policy to Action...

**DMP template**: provided in Annex I [of the H2020 FAIR Data Guidelines]

**DMP - a 'living' document** "[...] Since participation in the ORD pilot is not an evaluation criterion, the proposal is not expected to contain a fully developed DMP"

**Costs related to ORDP** are eligible for reimbursement (Article 6 and Article 6.2.D.3, but also other articles relevant for the cost category chosen)

**Guidelines on FAIR Data Management in Horizon 2020**


**Annotations specific for health research projects** are available in the H2020 Participant Portal

From Policy to Action...

The following data is concerned:

- **Data underlying scientific publications**, appropriately anonymised and complying with data protection rules
- **Additional data defined and agreed by the consortium in the data management plan (DMP)** (avoiding potential IP and confidentiality infringements)

The following data is not concerned:

- **Any type of data prior to publication** (unless otherwise agreed upon by the consortium)
- **Raw/individual patient data (IPD)** unless appropriately anonymised and compliant with data protection rules
Art. 29.3: "**As an exception**, the beneficiaries do not have to ensure open access to specific parts of their research data under Point (a)(i) and (iii), if the achievement of the action's main objective (as described in Annex 1) would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access."

**Opt-out possibilities**
- during the application phase
- during the grant agreement preparation (GAP) phase and
- after the signature of the grant agreement

**Reasons for opting out**
- Obligation to protect results
- Security obligations
- Protection of personal data
- Project does not collect/generate data
- Other legitimate reasons
From Policy to Action...

Opting out is possible for concrete and well-justified reasons.

But:

- It concerns data underlying publications => results need to be protected before publication => generally no conflict with obligation to protect results.
- It does not concern non-anonymised patient data => generally no conflict with data protection rules.

Therefore:

- for SC1 projects generally no reason to opt out.
- If there is a reason to opt out this must be specifically explained and not just refer to the obligation to protect personal data.

First Results of ORDP in SC1

- Intense information campaign...
- ≥ 49 grant agreements signed in Jan2018 - only three projects opted out

Successful outcome: demonstrates understanding of scientists that the ORDP does not threaten intellectual property protection and/or protection of personal data.
From Policy to Action...

The case of Public Health Emergencies (PHEs)

Recent change to Article 29.3 of H2020 MGA (18/10/2017): 'Open access to research data' to provide for third party access to research data in health actions in cases of public health emergencies

[...] (ii) [OPTION A for health actions that participate in the Open Research Data Pilot, if foreseen in the work programme: data which is relevant for addressing a public health emergency, if specifically requested by the [Commission][Agency] and within the deadline specified in the request][OPTION B: not applicable];

Rationale: acquire and share relevant research data as rapidly as possible for the most adequate and timely public health response.
Examples of international initiatives/organisations

- **International Human Epigenome Consortium (IHEC):** "A long term objective of IHEC is to understand the extent to which the epigenome has shaped human populations over generations and in response to the environment"
  - [http://epigenomesportal.ca/ihec/](http://epigenomesportal.ca/ihec/) - IHEC data portal

- **International Cancer Genome Consortium (ICGC):** "To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe"
  - [https://dcc.icgc.org/](https://dcc.icgc.org/) - ICGC data portal

- **International Human Mircobiome Consortium (IHMC):** "...The Consortium’s efforts are focused on generating a shared comprehensive data resource that will enable investigators to characterize the relationship between the composition of the human microbiome (or of parts of the human microbiome) and human health and disease."
From Policy to Action...

Examples of EU-funded projects/initiatives

**RD-Connect** (Topic: HEALTH.2012.2.1.1-1-C - Databases, biobanks and 'clinical bio-informatics' hub for rare diseases FP7-HEALTH-2012-INNOVATION-1): an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. To help researchers study rare diseases, RD-Connect links different data types - omics (e.g. genomics), clinical information, patient registries and biobanks - into a common resource. RD-Connect enables scientists and clinicians around the world to analyse genomics data and share them with other researchers. By making data accessible beyond the usual institutional and national boundaries, RD-Connect speeds up research, diagnosis and therapy development to improve the lives of patients with rare diseases - [http://rd-connect.eu/](http://rd-connect.eu/)

**European Network for Cancer Research in Children and Adolescents** (Topic: HEALTH.2010.2.4.1-3 - Structuring clinical research in paediatric and adolescent oncology in Europe. FP7-HEALTH-2010-single-stage): developed a cloud-based solution architecture of the Advanced Biomedical Collaboration Domain for Europe (ABCD-4-E), allowing integration of pseudonymised data pooled from different sources and linking clinical trials data with biological data from biobanks in the context of the development of European Virtual Institute for translational and clinical research in paediatric oncology. A 'Survivorship Passport', a single document summarising the patients’ cancer treatments and providing long-term recommendations based on individual cancer treatment related risks, developed jointly by ENCCA and PanCareSurFUP (PanCare Childhood and Adolescent Cancer Survivor Care and Follow-Up Studies) project is being further developed to become a cloud application available across borders. - [https://cordis.europa.eu/project/rcn/98250_en.html](https://cordis.europa.eu/project/rcn/98250_en.html)

From Policy to Action...

Embedding OS principles in H2020 SC1 (Health) WP 2018-2020 (examples)

• Exploiting research results and potential of the human microbiome for personalised prediction and prevention of disease - 50M€ - "[...] Proposals should build on data from existing microbiome projects and, as appropriate, on data from other international initiatives. Focussed production of new data should make subject coverage more comprehensive with the aim of delivering more valuable clinical tools [...]"

• Data integration and data-driven in-silico models for enabling personalised medicine - a European standardization framework - 2M€ "[...] The proposal should establish a forum for in-silico methodologies applied in translational and clinical research, where different transnational initiatives should meet and debate on their standardisation strategies. The project should evaluate the data integration and data-driven in-silico models strategies and identify best practices for integrating and modelling heterogeneous human disease data transnationally [...]"

• International flagship collaboration with Canada for human data storage, integration and sharing to enable personalised medicine approaches: 40M€ "[...] To build a collaboration of stakeholders in Europe and Canada in the domain of repositories storing and sharing human –omics data that will create a framework for long-term cooperation. In order to do so, this programme aims to enhance and standardise data deposition, curation and exchange procedures thus ensuring better data reuse and increased benefit to the scientific communities worldwide [...]"
Future framework programme and Open Science

"I am convinced that the core values of Horizon 2020 and its successor have to be:"

EXCELLENCE
OPENNESS
IMPACT
Health Research and Innovation Cloud (HRIC)

• Digital health platform to enable for data sharing and analysis for health research across the EU
• 13th of March, 2018: Workshop to explore the possibility and necessity to establish a Health Research and Innovation Cloud (HRIC) Participants: the user-community
• Objective: exchange ideas /concepts on a Health cloud for health research, the benefits for the end-users and the technical specifications for establishing a health cloud
• Next steps: explore options for implementation of a HRIC
Challenges ahead

- Need for data and metadata standards: health research-specific standards for data sharing but also rules how to apply those standards
- Need for common ontologies for better data harmonization per health research community/discipline
- Need for additional (specific and generic) searchable repositories
- Strong need for data protection and privacy guidelines and compliance with the GDPR provisions
Challenges ahead

- Digital tools for data collection and exchange and facilitating analytically-driven data integration (e.g. the integration of a wide variety of real-world data with clinical trials data)

- **How to share ‘rich’, connected datasets in a GDPR compliant manner?**
  - Definition of ‘personal data’ - re-identification of ‘anonymised’ data through linkages
  - Specific consent (or specific legal base) vs. the need for re-use, re-analysis
SESSION 3 – PANEL DISCUSSION

Saila Rinne
Programme Officer, Unit Data Policy and Innovation
DG CONNECT, Luxembourg
Unlocking the data for healthcare

Data-driven Research and Innovation

Public sector information
European Open Science Cloud
Access to research data
Big Data technologies
Free flow of non-personal data
GDPR
Privacy enhancing technologies
HPC
e-infrastructures
ePrivacy
SESSION 3 – PANEL DISCUSSION

Prof. Eva Mendez
Young European Research Universities (YERUN), OSPP
Madrid, Spain
Stephan Kuster
Secretary General, Science Europe
Brussels, Belgium
Research Data Management in times of Big Data and Open Science

**Research Data Management**
Science Europe / NWO Initiative

**What?** Voluntary international alignment of DMP Core Requirements and Criteria for Trusted Repositories

**Specific considerations on health data**
Can be highly confidential and sensitive
- subject to strict regulations

Importance of data sharing to advance health research
- Promote and facilitate data sharing

**Why?** Support scientists with data management in times of big data and open science

**What?** Framework supporting domains to develop DMP templates (Protocols)

**Framework for domain-specific RDM Protocols**

**Why?** Recognise necessity for domain-specific approach

When? Published January 2018

**How?** Collaboration SE Data WG with different communities

**How?** Collaboration SE Data WG with broad stakeholder community

When? Aim: policies ready by end of 2018, first organisations to implement in 2019
SESSION 3 – PANEL DISCUSSION

Tamás Bereczky
Coordinator EUPATI Germany, Patient Cluster lead IMI Big Data Project Harmony
Berlin, Germany
SESSION 3 – PANEL DISCUSSION

Jeanette Frey
Director BCU Lausanne, Vice-President
LIBER, Lausanne, Switzerland
Panel discussion:

Achieving open research data in health – overcoming the EU policy, institutional, and regulatory challenges

Moderator: Dr. Monica Dietl
Science|Business

Saila Rinne
DG Connect

Prof. Eva Mendez
YERUN, OSPP

Stephan Kuster
Science Europe

Tamás Bereczky
EUPATI Germany

Jeanette Frey
LIBER

Dr. Cornelius Schmalz
DG RTD

Organised by

In collaboration with
WELCOME

Dr. Frederick Fenter
Frontiers Executive Editor
Heartfelt thanks for your participation!

Please join us at the networking reception.