

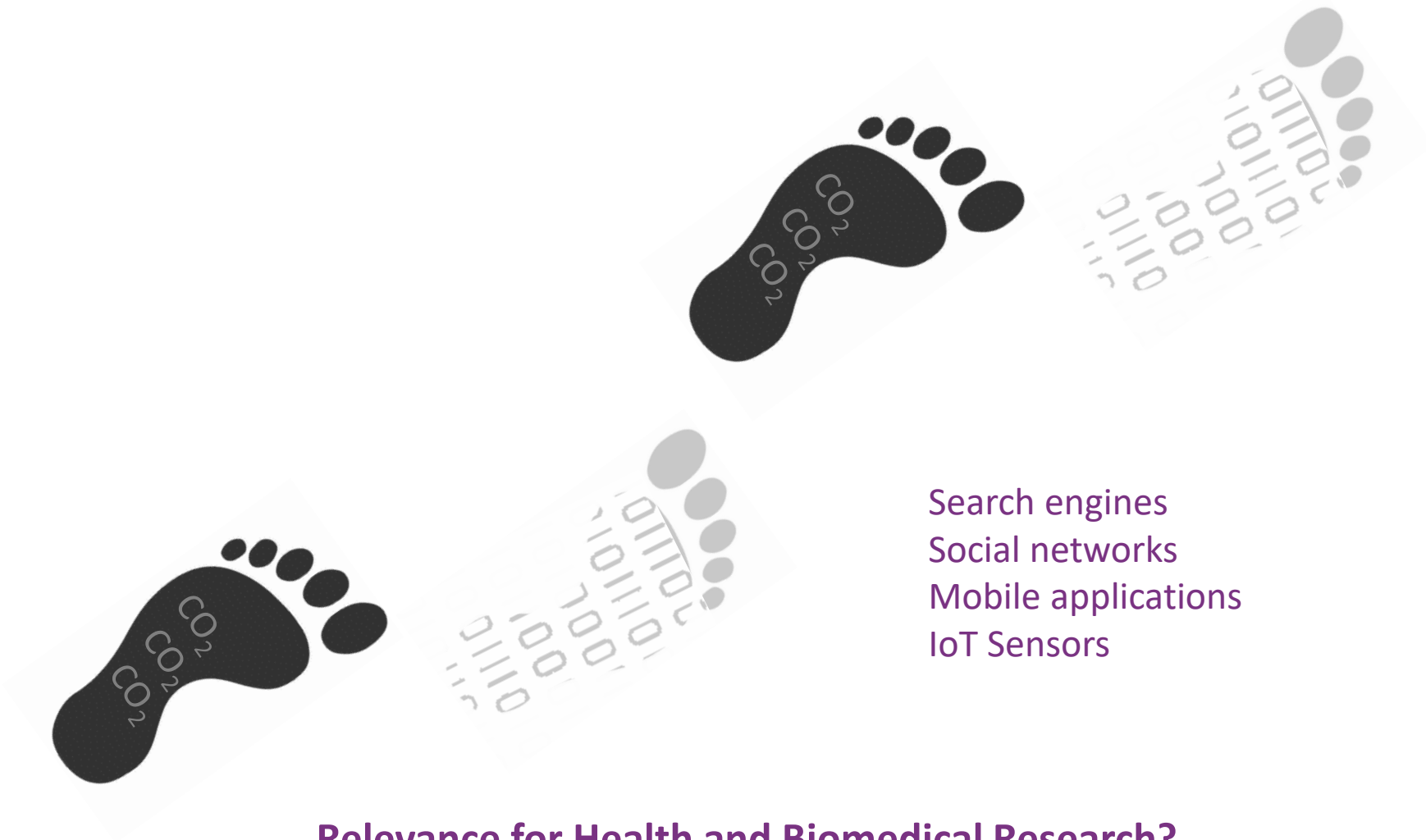


● ● ● BIOMARKER RESEARCH

Digital Biomarkers – Patient data mining & precision medicine

Stefan Schulz, Medical University of Graz
Donausymposium Vienna, March 14, 2018

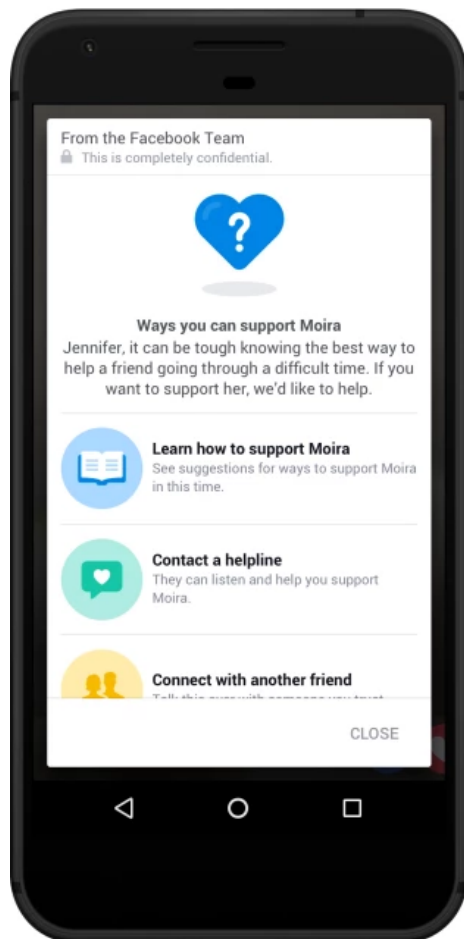
Digital Footprints



Relevance for Health and Biomedical Research?



New from the *Weekly Spark*



Can Facebook's Machine-Learning Algorithms Accurately Predict Suicide?

March 10, 2017

News Type: Weekly Spark, Weekly Spark News

Scientific American

Facebook has just expanded the array of tools it provides to reach users at risk for suicide and connect them with mental health resources. The menu of options that allows Facebook users to report posts with content indicating potential thoughts of suicide or self-harm will now be available for Facebook live streams as well. The social media company is also piloting a pattern recognition algorithm that it hopes will automatically identify posts of concern even if they have not yet been reported by users. According to Facebook spokesperson William Nevius, the algorithm will use words or phrases related to suicide or self-harm in a user's post, and in comments added by friends, to determine if the person may be at risk. The system will automatically alert Facebook's Community Operations team about posts of concern so that the team can quickly review them. If the team determines that support is warranted, they will ensure that information about helping resources will appear in the user's news feed.

Spark Extra! Check out a [community guide](#) for Facebook users.

Planning and Implementing: New and Social Media

Digital Biomarkers?



Mining "Digital Biomarkers" from Electronic Health Records

Mining "Digital Biomarkers" from Electronic Health Records

Michael Huber
03.10.1951 (65 Jahre)

ALLERGIEN UND RISIKEN
Penicillin, Kontrastmittel,
Blutverdünner, Herzschrittmacher,
(1) weitere

LABOR WERTE
Virologie 28.10.16
Aspirin 22.10.16

DIAGNOSEN (281) **ZEITLEISTE** **MEDIKATION (11)** **BEFUNDE (4)** **VOR-OPERATIONEN (5)** **STAMMDATEN** **RISIKOFAKTOREN (5)**

☒ Labor ☒ Radiologie ☒ Befunde ☒ Diagnosen ☐ Chemotherapie ☒ Operationen

Lab data
Leukozyten: 8 Tsd/µl
CRP-Wert: 152 mg/l

Biosignals, Images
RADIOLOGIE
MRT-Scan

ADT patterns
281 Diagnosen

Extracts from Clinical narratives
Surgical Pathology Report

Diagnoses
Subkutaner Tumor ...
Lumbago
Bandscheibenprotrusion LWS
Akute Cholezystitis

Redefining "Digital Biomarker"

Redefining "Digital Biomarker"

- **Person-related data**
- **Heterogeneous in format, quality, correctness, completeness, degree of standardisation**
- **Not primarily acquired for prediction of conditions / events**
- **Often implicit contexts**
- **Different levels of complexity**
 - Simple: single data points
 - Complex: series of data points
 - Algorithmic: data + multivariate prediction models
- **Predictive value: digital biomarker allows prediction of conditions or events to a relevant degree**

Examples for digital biomarkers

Digital Biomarker	Condition / Event	Specificity	Sensitivity
*GAITRite® signals	Bradykinesia	+	+
*Wii Balance Board signals	Postural instability	+	+
**HITEx algorithm	Current Smoker	+++	++
Mention of "Metformin" in the EHR	Diabetes mellitus type 2	++	-
Administrative ICD codes I10 or I11 or I12 or I13 or I15	Hypertensive disease	++	+/-
substring "malign" in pathology report	malignancy	--	+
*** Regular expression pattern matching	Gleason score, Clark level, Breslow depth	++	++

*Godinho C, Domingos J, Cunha G, et al. A systematic review of the characteristics and validity of monitoring technologies to assess Parkinson's disease. Journal of NeuroEngineering and Rehabilitation. 2016;13:24.

** Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Medical Informatics and Decision Making. 2006;6:30

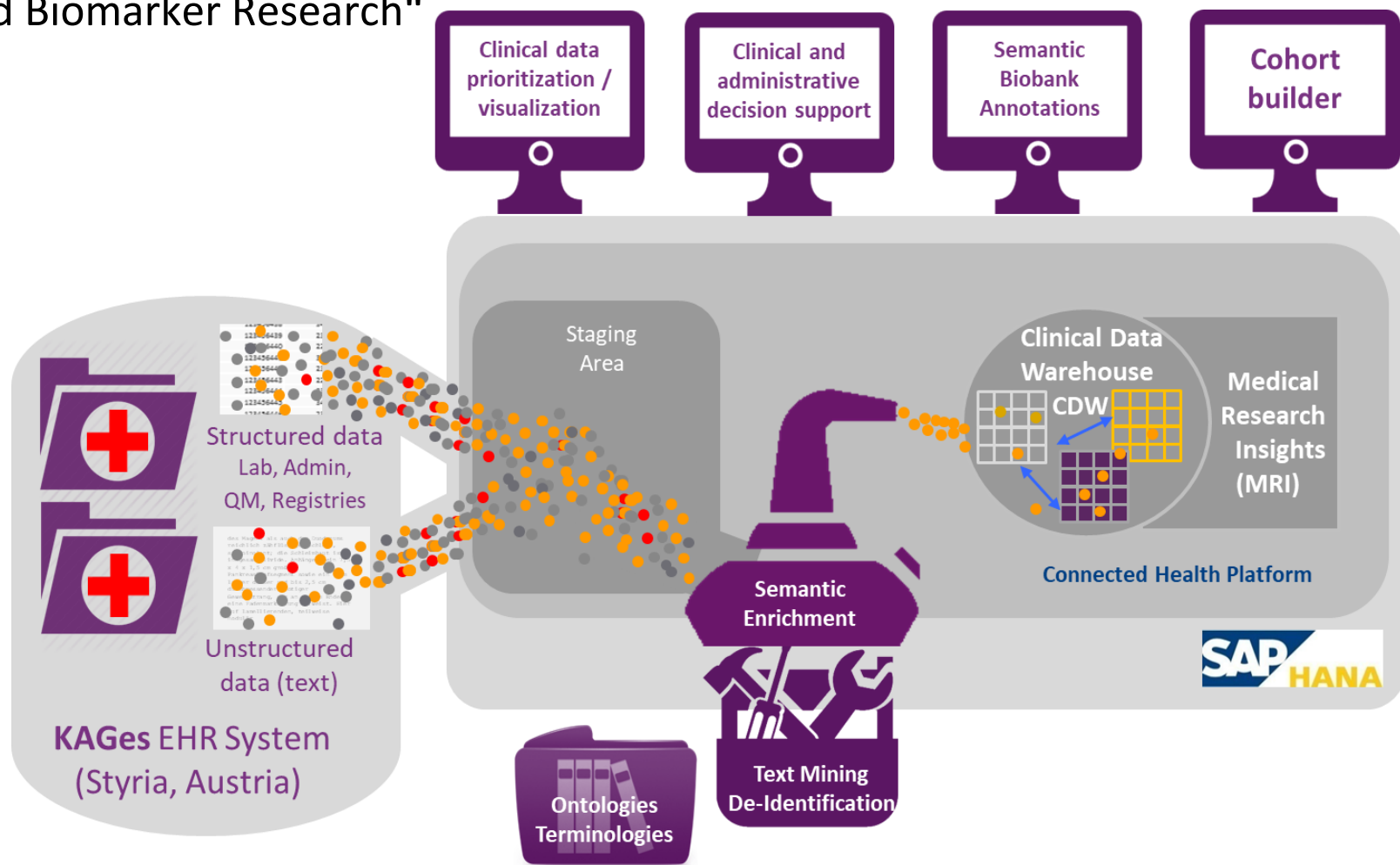
*** Napolitano G, Fox C, Middleton R, Connolly D. Pattern-based information extraction from pathology reports for cancer registration. Causes Control. 2010 Nov;21(11):1887-94.

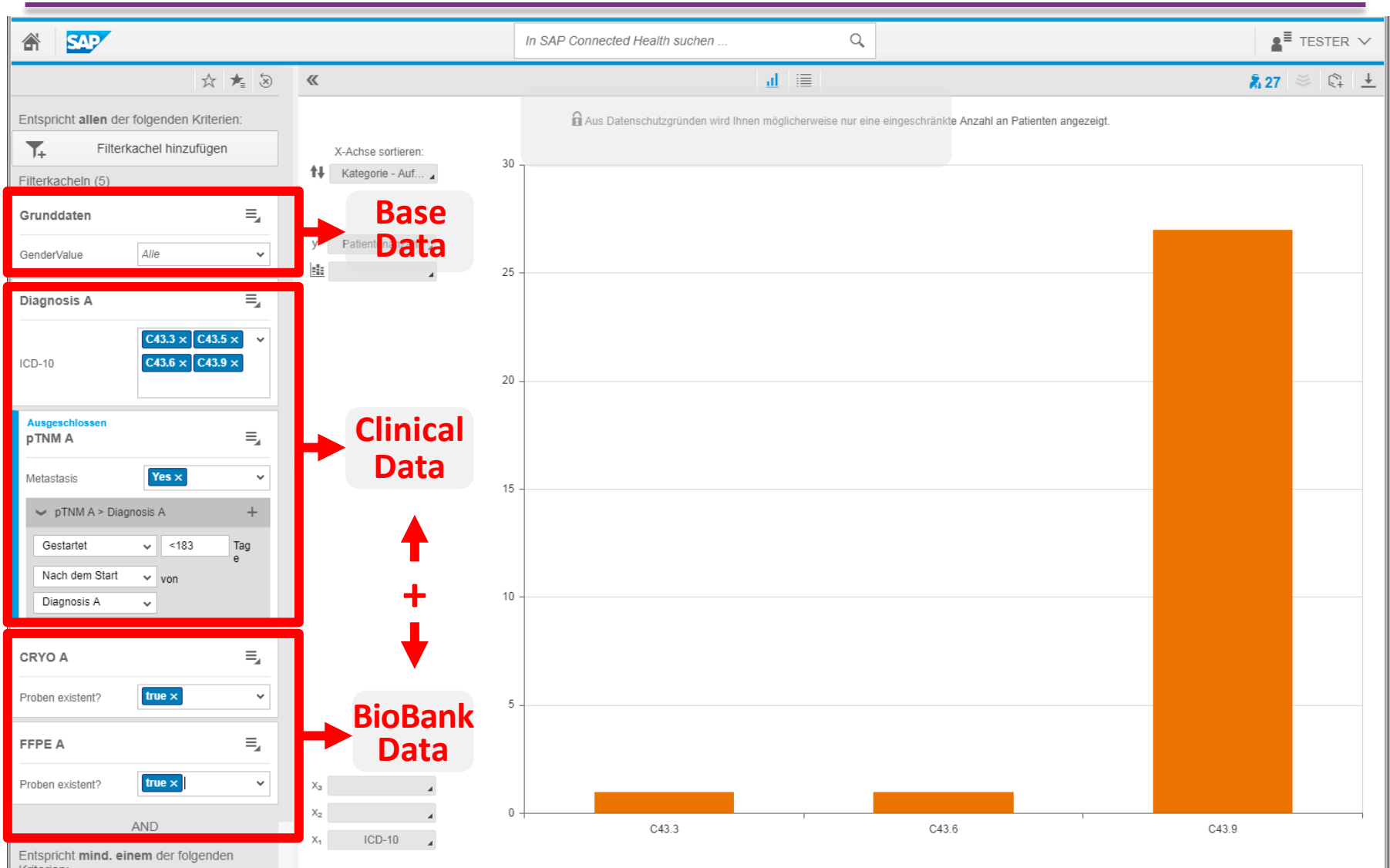
Predictive value of digital biomarkers

- **A priori prevalence (a priori) of condition**
- **Fulfilment of FAIR criteria (findability, accessibility, interoperability, reusability of data)**
 - Semantic standards used in source
 - Natural (sub-)language used in source (if information to be extracted free text)
 - Completeness of information
 - Correctness of information
 - Sophistication of information extraction routines
- **Validation of digital biomarkers necessary**
 - need of reference standard
 - similar to the validation of other biomarkers

Project 1.2 "DBM4PM Digital biomarkers for precision medicine"

- Continuation of "IICCAB – Innovative Use of Information for Clinical Care and Biomarker Research"





Example: prediction of delirium

[Stud Health Technol Inform.](#) 2017;236:32-39.

Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients.

[Kramer D](#)¹, [Veeranki S](#)², [Hayn D](#)², [Quehenberger F](#)³, [Leodolter W](#)¹, [Jagsch C](#)¹, [Schreier G](#)².

Author information

Abstract

Delirium is an acute confusion condition, which is common in elderly and often misdiagnosed in hospitalized patients. Early identification and prevention of delirium could reduce morbidity and mortality rates in those affected and reduce hospitalization costs. We have developed and validated a multivariate prediction model that predicts delirium and gives an early warning to physicians. A large set of patient electronic medical records have been used in developing the models. Classical learning algorithms have been used to develop the models and compared the results. Excellent results were obtained with the feature set and parameter settings attaining accuracy of 84%.

KEYWORDS: Delirium; electronic medical record; hospitalized patients; predictive model

PMID: 28508776

- Important amount of "digital" footprint in routine documentation (EHRs): heterogeneous, incomplete, biased, low structured
- Justifies extending the notion of "digital biomarker" to pieces of information to be extracted from EHRs and which have predictive value
- Information extraction and machine learning approaches, dependent on natural language resources play important role
- Validation of digital biomarkers essential