Domain-specific Design of Patient Classification in Cancer-related Cachexia Research

Tiziana Margaria
University of Limerick and Lero
Limerick, Ireland

Alexander Wickert
University of Potsdam
Potsdam, Germany

Anna-Lena Lamprecht
Utrecht University
Utrecht, The Netherlands
Motivation

Cachexia is a complex wasting syndrome associated with a marked detrimental effect upon life quality and survival in patients with cancer, chronic obstructive pulmonary disease (COPD), chronic heart failure, AIDS, and chronic kidney disease, among other conditions. Its prevalence is of around 5 to 15% in cardiac patients at end stage, rising up to 30%, in COPD and chronic kidney disease patients, and to 80% in patients with advanced cancer. Cachexia symptoms include pronounced weight loss, due to both lean and fat mass wasting: anorexia, malabsorption, nausea, asthenia, neuroendocrine changes, immune system function impairment, and disruption of energy metabolism.

Despite its unquestionable relevance to the poorer outcome of treatment in disease and its high prevalence among patients, the syndrome is still underdiagnosed and seldom treated. Part of the difficulty in treating cachexia relies on the fact that, in the clinical setting, the syndrome is recognised solely in its most advanced stages, when therapy available to the present day is not able to fully reverse its symptoms.
Motivation

Therefore, scientists and clinicians should focus on identifying early changes, as to intervene in a precocious manner. Taken together, the issue provides insights on the importance of detecting early signs of inflammatory changes in patients and examines the mechanisms that act in concert, inducing cachexia symptoms.
## External Groups

Table 1. Remote associated groups and their geographical location and affiliation.

<table>
<thead>
<tr>
<th>Group</th>
<th>Institute/Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alessandro Laviano,</td>
<td>Department of Clinical Medicine</td>
</tr>
<tr>
<td>Maurizio Muscaritoli</td>
<td>Sapienza University of Rome, Rome, Italy.</td>
</tr>
<tr>
<td>Giorgio Trinchieri,</td>
<td>Center for Cancer Research</td>
</tr>
<tr>
<td>Romina Goldzmid</td>
<td>National Cancer Institute, Bethesda, Maryland USA</td>
</tr>
<tr>
<td>Josep M. Argilès,</td>
<td>Cancer Research Group, Institut de Bioomedicina, Univ. Barcelona</td>
</tr>
<tr>
<td>Silvia Busquets</td>
<td>Barcelona, Spain</td>
</tr>
<tr>
<td>Nicolaas Deutz</td>
<td>Department of Health &amp; Kinesiology</td>
</tr>
<tr>
<td></td>
<td>Texas A&amp;M University, Bryan, Texas, USA</td>
</tr>
<tr>
<td>Stephen Farmer</td>
<td>Department of Biochemistry</td>
</tr>
<tr>
<td></td>
<td>Boston University School of Medicine</td>
</tr>
<tr>
<td></td>
<td>Boston, MA, USA</td>
</tr>
<tr>
<td>Gerhard Paul Püschel</td>
<td>Institute of Nutritional Science</td>
</tr>
<tr>
<td></td>
<td>University of Potsdam, Potsdam, Germany</td>
</tr>
<tr>
<td>Tiziana Margaria</td>
<td>Institute of Informatics</td>
</tr>
<tr>
<td></td>
<td>University of Potsdam, Potsdam, Germany</td>
</tr>
<tr>
<td>Barry D. Floyd</td>
<td>California State Polytechnic University</td>
</tr>
<tr>
<td></td>
<td>San Luis Obispo, CA, USA</td>
</tr>
</tbody>
</table>

Application Experts

Technical Experts
Main Challenge

- Life sciences researchers are seldom IT professionals
  - they need to work together efficiently
  - with the data definition and management techniques that complex and evolving experimental settings require
Definition of Cachexia

Cachexia: A new definition


**CACHEXIA DIAGNOSIS**

- Decreased muscle strength
- Fatigue
- Anorexia
- Low fat-free mass index
- Abnormal biochemistry:
  - Increased inflammatory markers (CRP, IL-6)
  - Anemia (Hb < 12 g/dL)
  - Low serum albumin (<3.2 g/dL)

Weight loss of at least 5% in 12 months or less
(or BMI <20 kg/m²)

3 of 5
What **data** do we have?

- Questionnaire scores
- Anthropometric data
- Body composition images
- Brain images
- Histological images
- Immunodetection images
- Biochemical parameters in plasma
- Endocrine parameters in plasma
- Inflammatory parameters in plasma
- Biochemical parameters in tissues/organs
- Inflammatory parameters in tissues/organs
- Cell sorting and phenotyping spectra
- Chromatographic spectra
- Molecular parameters in cells
- Microarray analysis
- Gene sequencing
- Physiological parameters associated with the effect of exercise
How to work with all these information?

- PDF Files
- Excel spreadsheets
- Everything is done manually

Interdisciplinary:
Bio–chem lab
Nutrition science
Immunology
Sports/rehab
Oncology

Surgery
Gastroenterology
Psychology
(Computer Science)
(ED)

Managing Processes

By Rodolfo G. Camargo, USP
Managing Processes

Cachexia – the Big Picture
(DAAD, FAPESP, CAPES)

Interdisciplinary:
Bio–chem lab
Nutrition science
Immunology
Sports/rehab
Oncology
Surgery
Gastroenterology
Psychology
(Computer Science)
(ED)
Stakeholder Requirements:

- semiotically intuitive, graphical approach

- a framework that:
  - is able to manage complexity and change
  - helps to efficiently produce more reliable results
  - appears simple and intuitive to its users
## The Patient Classification Spreadsheet

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td><strong>B</strong></td>
<td><strong>C</strong></td>
<td><strong>D</strong></td>
<td><strong>E</strong></td>
<td><strong>F</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td><strong>PATIENT'S INFORMATION</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>FIRST CRITERION - WEIGHT LOSS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>Gender</td>
<td>Age (Years)</td>
<td></td>
<td>Weight variation</td>
<td>BMI (kg/m²)</td>
<td>Result</td>
</tr>
<tr>
<td>165A</td>
<td>Male</td>
<td>51</td>
<td></td>
<td>-10%</td>
<td>28.88</td>
<td>IN</td>
</tr>
<tr>
<td><strong>SECOND CRITERION - WEIGHT STRENGTH</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>THIRD CRITERION - FATIGUE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Score</td>
<td>Result</td>
<td></td>
<td>Method</td>
<td>Score</td>
<td>Result</td>
</tr>
<tr>
<td>Questionnaire (QLC-C30)</td>
<td>53,33333333</td>
<td>OUT</td>
<td></td>
<td>Questionnaire (QLC-C30)</td>
<td>33.33</td>
<td>IN</td>
</tr>
<tr>
<td>Answer 1</td>
<td>1</td>
<td>OUT</td>
<td></td>
<td>Answer 10</td>
<td>4</td>
<td>IN</td>
</tr>
<tr>
<td>Answer 2</td>
<td>4</td>
<td>OUT</td>
<td></td>
<td>Answer 12</td>
<td>3</td>
<td>IN</td>
</tr>
<tr>
<td>Answer 3</td>
<td>2</td>
<td>OUT</td>
<td></td>
<td>Answer 18</td>
<td>2</td>
<td>IN</td>
</tr>
<tr>
<td><strong>FOURTH CRITERION - ANOREXIA</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>FIFTH CRITERION - FAT FREE MASS INDEX</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Score</td>
<td>Result</td>
<td></td>
<td>Method</td>
<td>Score</td>
<td>Result</td>
</tr>
<tr>
<td>Questionnaire (QLC-C30)</td>
<td>100.00</td>
<td>OUT</td>
<td></td>
<td>DEXA Scan</td>
<td>6.09</td>
<td>IN</td>
</tr>
<tr>
<td>Answer 13</td>
<td>1</td>
<td>OUT</td>
<td></td>
<td>Lean mass (kg)</td>
<td>15.4</td>
<td>IN</td>
</tr>
<tr>
<td><strong>SIXTH CRITERION - BIOCHEMICAL PARAMETERS</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>GROUP CLASSIFICATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameters</td>
<td>Concentration</td>
<td>Result</td>
<td></td>
<td><strong>BARCODE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Reactive protein (mg/l)</td>
<td>6,10</td>
<td>IN</td>
<td></td>
<td>CACHEXIA WITHOUT CANCER</td>
<td></td>
<td>NONE</td>
</tr>
<tr>
<td>IL-6 (pg/ml)</td>
<td>5,34</td>
<td>IN</td>
<td></td>
<td>LEVEL OF EXCLUSION CRITERIA</td>
<td></td>
<td>NONE</td>
</tr>
<tr>
<td>Anemia - Hb (g/dl)</td>
<td>12,30</td>
<td>IN</td>
<td></td>
<td><strong>0</strong></td>
<td></td>
<td>NONE</td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
<td>4,89</td>
<td>IN</td>
<td></td>
<td>165A</td>
<td></td>
<td>NONE</td>
</tr>
</tbody>
</table>

Adapted from Evans, 2008
Example: Excel Formula (Cell B9)

=WENN(UND(A9="Questionnaire (QLC-C30)";B10<>"";B11<>"";B12<>"";B13<>"";B14<>""):(1-(MITTELWERT(B10:B14)-1)/3)*100;
WENN(UND (A9="HandgripTest"; B3="Male";B10<44;B10<>"")) ; "POOR";
WENN (UND (A9="HandgripTest"; B3="Male";B10>44;B10<>"")) ; "GOOD";
WENN (UND (A9="HandgripTest"; B3="Female";B10<23;B10<>"")) ; "POOR";
WENN (UND (A9="HandgripTest"; B3="Female";B10>22;B10<>"")) ; "GOOD"; "---")

for regular use in large-scale research projects it has significant drawbacks:

– for every patient the data has to be entered manually in a single spreadsheet → also the result has to be transferred manually
– Excel formulas are complex and not easily understandable
– maintenance, modifications, and extensions are difficult and error-prone
– no statistics on the data possible
– no unified database
The Patient Classification Spreadsheet

- complex **patient classification determines** which individuals belong to the different patient and control groups

- it is a **multifaceted evaluation** of a number of interdisciplinary criteria

We

- organized and simplified the patient classification process,

- made it **easily accessible** and **shareable** worldwide via a web application

- the integrated framework collects the data of patients in a **central repository** (easy to access and back up)
Approach

IDE for **co-design** and **co-evolution** of data and process models

- Data: Dynamic Web Application (**DyWA**),
- Behaviour: Java Application Building Center 4 (**jABC4**) modeling framework

**Result:**

easily executable domain-specific processes
Co-Development with DyWA and jABC4

1. Domain modeling
   - Web Interface
   - API
   - Domain Model
   - Processes
   - Persistence Layer

2. Generate
   - Domain-specific SIBs
   - Model Library

3. Process modeling
   - Domain Processes
   - Business Processes

4. Deploy

Patient samples

Domain expert

Process manager

Workflows
Prototype-Driven Development  (DevOps style)

- start development cycle
- model the domain
- Data types
- generate „domain actions“
- step-by-step migration
- Behaviours
- deploy processes
- generate executable processes
- adapt process models

Margaria, Wickert, Lamprecht
DyWA: Data Type Definition

- DSL design approach = start by fixing the vocabulary of the domain
- definition of a set of domain-relevant “things” with their respective types
- DyWA provides domain-independent Java types (e.g. String, Integer) as initial type collection
- any self-modeled domain specific type (e.g. Weight) becomes directly available and can be used as a field or attribute of complex data types
- CRUD operations = Create, Read, Update, Delete are automatically generated for every defined type and field by the DyWA
DyWA: Data Type Definition
(Margaret Hamilton: who you are)

Types

Available Types

- Answer_enum_values
- Gender_enum_values
- Height_enum_values
- Height_type
- Identification
- Length_enum_values
- Length_type
- Patient_information
- Patient_Questionary

Patient_information

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>In Use As</th>
</tr>
</thead>
<tbody>
<tr>
<td>identification</td>
<td>Identification</td>
<td>true identification</td>
</tr>
<tr>
<td>gender</td>
<td>Gender_enum_values</td>
<td>true gender</td>
</tr>
<tr>
<td>age</td>
<td>integer number</td>
<td>true age</td>
</tr>
<tr>
<td>height</td>
<td>Height_type</td>
<td>true height</td>
</tr>
<tr>
<td>weight_previous</td>
<td>Weight_type</td>
<td>true weight_previous</td>
</tr>
</tbody>
</table>

(model the domain)
DyWA: Data Type Definition

```
Edit type: Patient_Information

Meta
Name: Patient_Information

Fields of type
1. Identification: identification
2. Gender_enum_values: gender
3. Integer number: age
4. Height_type: height
5. Weight_type: weight_previous
6. Weight_type: weight_current
    ...  
15. Floating-point number: albumin_in_g_per_dl

Text: 
```

UML Class Diagram of the Domain Model
Taxonomy of SIBs: the „Microservices“
(Margaret Hamilton: what you do)

CRUD SIBs automatically generated
“emerging” DSL for healthcare
Process Modeling with jABC4

1. Start development cycle
2. Model the domain
3. Generate domain actions
4. Adapt process models
Let \( AP \) be a set of atomic propositions. A **Kripke Transition System** over \( AP \) is a 4-tuple \( K=(S,\text{Act},\text{Trans},I) \) with:

- \( S \) a set of states
- \( \text{Act} \) a set of actions
- \( \text{Trans} \subseteq S \times \text{Act} \times S \) a transition relation
- \( I : S \rightarrow 2^{AP} \) an interpretation function
Process Model: TransformWeight2kg

start

success → read success

read weight_raw_value

read success

read weight_enum_type

read success

read factor2kg from weight_enum_type

read success

weight_raw_value * factor2kg = weight_in_kg

success
Path through Graph Level Hierarchy
Executable Processes deployed to DyWA

Available Processes
- Filter
- Barcode
- Calculate_Ali_Handgrip_DEXA
- Calculate_Ali_Handgrip_MUAMA
- Calculate_Ali_QLC_DEXA
- Calculate_Ali_QLC_MUAMA
- Criterion1_Calculate_BMI
- Criterion1_Calculate_Result
- Criterion1_Calculate_WeightVariation
- Criterion2_Calculate_HandgripTest_Result
- Criterion2_Calculate_HandgripTest_Score
- Criterion2_Calculate_QLC_C30_Result

Criterion1_Calculate_Result

- Description
- Process Canvas

- Start development cycle
- Model the domain
- Generate domain actions
- Adapt process models

- Generate executable processes
- Deploy processes
Patient Training
(treadmill, 6 weeks)
Blood sampling

1. Plasma
   - centrifugation
   - fill upper phase into Eppendorfs, label with number and "P"
   - keep on ice
   - store in -80°C freezer
   - analyse blood profile

2. Serum
   - centrifugation
   - fill upper phase into Eppendorfs, label with number and "S"
   - keep on ice
   - store in -80°C freezer
   - analyse blood profile

3. Lactate
   - centrifugation
   - fill upper phase into 1 Eppendorf, label with number and "L"
   - keep on ice
   - store in -80°C freezer
   - send to RDO for lactate analysis

- nurse (Raquel): take blood
  (2 tubes for untrained patients, 3 tubes for trained patients)
Outcome:
Incremental Modeling of Data and Processes

<table>
<thead>
<tr>
<th>Group</th>
<th>Institute/Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alessandro Laviano, Maurizio Muscaritoli</td>
<td>Department of Clinical Medicine, Sapienza University of Rome, Rome, Italy.</td>
</tr>
<tr>
<td>Giorgio Trinchieri, Romina Goldzimid</td>
<td>Center for Cancer Research, National Cancer Institute, Bethesda, Maryland USA</td>
</tr>
<tr>
<td>Josep M. Argilés, Silvia Busquets</td>
<td>Cancer Research Group, Institut de Biomedicina, Univ. Barcelona, Barcelona, Spain</td>
</tr>
<tr>
<td>Nicolaas Deutz</td>
<td>Department of Health &amp; Kinesiology, Texas A&amp;M University, Bryan, Texas, USA</td>
</tr>
<tr>
<td>Stephen Farmer</td>
<td>Department of Biochemistry, Boston University School of Medicine, Boston, MA, USA</td>
</tr>
<tr>
<td>Gerhard Paul Piischel</td>
<td>Institute of Nutritional Science, University of Potsdam, Potsdam, Germany</td>
</tr>
<tr>
<td>Tiziana Margaria</td>
<td>Institute of Informatics, University of Potsdam, Potsdam, Germany</td>
</tr>
<tr>
<td>Barry D. Floyd</td>
<td>California State Polytechnic University, San Luis Obispo, CA, USA</td>
</tr>
</tbody>
</table>

Application Experts

Technical Experts
Outcome: Incremental Modeling of Data and Processes

- Patient classification process is now fully automated as a workflow
  - modeled with jABC4, and
  - integrated into a database provided by the DyWA

- jABC4 = model-driven environment for designing the processes

- DyWA = meta-schema based data definition and management tool with standard relational database

- Their interplay provides an integrated environment for data and process modeling along the XMDD paradigm

- Supports a Service-oriented Continuous Engineering approach to the formalization and definition of a domain-specific language and process landscape
eXtreme Model Driven Design

http://cinco.scce.info/

http://hope.scce.info

http://dime.scce.info/
Thank you!
Questions?

Contact: tiziana.margaria@lero.ie
Conclusion

- integrated data and process modeling environment:
  - data collection,
  - data transformation,
  - automation,
  - reproducibility of results

- many processes have the potential to be reused by other health care applications, or even in other domains
Conclusion

- **adaptations** to changing experimental setups are still possible

- **processes** are **immediately executable** and remain customizable

- this environment provides a significant step towards the large-scale applicability of a **formal** model-based and methods-supported, model-driven, generative IDE for scientists

- the IDE ensures that the modeling of domain-specific **data types and processes** using these components happens in one coherent system at a user-accessible level

- **result**: immediate availability, consistency and reproducibility of the outcomes, and the coherence and evolvability of the entire collection of data schema and processes
Ongoing Work

- **access to data and processes** in the web application should be based on a proper **roles and rights management** (e.g. DIME)

- **provenance tracking and auditing** of all the data collected and accessed
  - to know **who did what when** with **which permissions**, and
  - to maintain truly **complete records** of experimental results