Why data management? Use cases and user insights

Piotr Zadora
Division „Systems Biology of Signal Transduction“
German Cancer Research Center (DKFZ), Heidelberg
Why data management is so important?
General information on the project

Goal:
Improvement of anemia treatment of lung cancer patients

Means:
Mathematical model predicting treatment options of the patient

Available data:
- Lung cancer patients stage 4
- Longitudinal measurements of hemoglobin (Hb) and inflammation (CRP) values
- Indication of additional perturbations
  - Blood transfusion
  - Chemotherapy
  - Treatment with erythropoiesis-stimulating agents (ESAs)
Exemplary dataset

**Patient**

- **Hb [mmol/l]**
  - Values range from approximately 8 to 11 mmol/l.

- **CRP [mg/dl]**
  - Values range from almost 0 to 20 mg/dl.

**Time [days]**

- Periods indicated by shaded areas (0-100, 300-400 days).

Graphs showing changes in Hb and CRP over time.
Pitfalls of data collection and availability

<table>
<thead>
<tr>
<th>patient_num</th>
<th>concept_cd</th>
<th>name_char</th>
<th>modifier_cd</th>
<th>start_date</th>
<th>instance_number</th>
<th>val_char</th>
<th>rval_num</th>
<th>units_cd</th>
<th>end_date</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>ADIT-0</td>
<td>No adjuvant therapy</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>CARBOPEM</td>
<td>CARBOPEM</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>CARBOPEM</td>
<td>@</td>
<td>01.08.20 00:00</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>CHEIGHT-XM</td>
<td>Height</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>0</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>CLENTH-XN</td>
<td>Length</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>3.3</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>CSIZE-35</td>
<td></td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td></td>
<td>@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>CWIDTH-XM</td>
<td>Width</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>0</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>DB00642</td>
<td>Pemetrexed</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>Pemetrexed</td>
<td>@</td>
<td>01.08.20 00:00</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>DB00958</td>
<td>Carboplatin</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>Carboplatin</td>
<td>@</td>
<td>01.08.20 00:00</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>GRAD:UN</td>
<td>Uncertain</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td></td>
<td>@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>ICD10:C34</td>
<td>Lung Cancer Histology</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>ICD10:C34</td>
<td>@</td>
<td>01.08.20 00:00</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>ICD-O-3:814</td>
<td>Adenocarcinoma n.o.s.</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td></td>
<td>@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>K:ICD10:i10:00</td>
<td>Benign Essential Hypertension : No indication of a hypertensive crisis</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>K:ICD10:i10.00</td>
<td>@</td>
<td>01.08.20 00:00</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>KNM:J</td>
<td>Bone metastasis</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td></td>
<td>@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>0.24</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>0.25</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>0.36</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>2.98</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>2.63</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>4.68</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>5.14</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>8.63</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>5.75</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>3.43</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>2.67</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:1988-5</td>
<td>CRP [Mass,Ävolu] in Serum or Plasma</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>1.43</td>
<td>mg/dL</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:3141-9</td>
<td>Body weight measured</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>77</td>
<td>kg</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:718-7</td>
<td>Hemoglobin in Blood</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>9.0045</td>
<td>mmol/l</td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>L:718-7</td>
<td>Hemoglobin in Blood</td>
<td>@</td>
<td>01.07.20 00:00</td>
<td>1</td>
<td>E</td>
<td>9.0065</td>
<td>mmol/l</td>
<td></td>
</tr>
</tbody>
</table>
Pitfalls of data collection

- Data stored at different „areas“ within the clinic or at different clinics
- Missing indications of medication
Pitfalls of data collection

- Data stored at different „areas“ within the clinic or at different clinics
  - Missing indications of medication
  - Duplication of therapies

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISVIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>VINORELBIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>CISPLATIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>DOCETAXEL</td>
<td>7200</td>
<td>7920</td>
</tr>
<tr>
<td>DOC</td>
<td>7200</td>
<td>7920</td>
</tr>
</tbody>
</table>
Pitfalls of data collection

- Data stored at different „areas“ within the clinic or at different clinics
  - Missing indications of medication
  - Duplication of therapies

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISVIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>VINORELBIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>CISPLATIN</td>
<td>432</td>
<td>1104</td>
</tr>
<tr>
<td>DOCETAXEL</td>
<td>7200</td>
<td>7920</td>
</tr>
<tr>
<td>DOC</td>
<td>7200</td>
<td>7920</td>
</tr>
</tbody>
</table>

- Not possible for theoreticians to understand and distinguish an artefact from a „true“ treatment
- Lack of Standardization!
Data collection using OpenBis

- **Code:** 00_P_INFO_LC_8951_
- **Experiment:** PATIENT_DATABASE_LUNGCANCER
- **Space:** PATIENT_DATABASE_LUNGCANCER

**Parents:**
List of Sample (codes or identifiers) separated by commas (","), or one Sample per line.

**Tags:**
List of tag names separated by commas (","), or one tag name per line. If a tag does not exist, it will be created.

**General information**
- **Patient ID:** Real number value
- **Gender:** Choose...
- **YOB:** Real number value
- **YOD:** Real number value
Data collection using OpenBis

General information
Patient ID: *
Gender: *
YOB: *
YOD:

Visit date: *
YYYY-MM-DD HH:MM:SS ZZZZ

Hb values
HB value: * Integer value
Hb Unit: *

CRP values
CRP value: * Integer value
CRP Unit: *
Data collection using OpenBis

- Code: *
- Experiment: Code: *
- Space: *
- Parents: Code: *

Tags: List of Sample (codes or identifiers) separated by commas (",") or one Sample per line.

Add Parent...

Tags: List of tag names separated by commas (",") or one tag name per line. If a tag does not exist, it will be created.

Add Tag...

General information
- Patient ID: *
- Gender: *
- YOB: *
- YOD: *

General
- Visit date: *
- Date of ESA administration: * YYYY-MM-DD HH:MM:SS ZZZZ
- Used ESA: *
- Used ESA conc.: *
- ESA Unit: *

Hb values
- Hb value: *

CRP values
- CRP value: *
- CRP Unit: *

Integer value
Data collection using OpenBis

List of Sample (codes or identifiers) separated by commas (",") or one Sample per line.

List of tag names separated by commas (",") or one tag name per line. If a tag does not exist, it will be created.
Distinguished spaces for patient information and data

- Patient Info
- Patient visit
- Patient chemotherapy
- Patient transfusion
- Patient ESA treatment
Structure of OpenBis data management in B200/DKFZ

Initial phase
Structure of OpenBis data management in B200/DKFZ

Initial phase

Patient data storage
Structure of OpenBis data management in B200/DKFZ

Initial phase

Patient data storage

Biological validation/question
Structure of OpenBis data management in B200/DKFZ

![Diagram showing the structure of OpenBis data management in B200/DKFZ with phases and roles]

- **Initial phase**
- **Patient data storage**
- **Biological validation/question**