Making healthcare data FAIR data
the ontologies-data models-instances conundrum

Ronald Cornet
Ronald Cornet, PhD

• Associate professor and Principal Investigator at Amsterdam UMC, Amsterdam Public Health Research Institute, department of Medical Informatics
• Research and teaching on knowledge representation; ontology auditing; SNOMED CT; reusable healthcare data; FAIR data
(Conflict of) Interests

- Chair of the interim Executive Board of GO-FAIR
- Member of SNOMED Modeling Advisory Group
- Member of the Dutch Norm Committee
Funding

• FAIR4Health.eu Horizon2020 #824666

• EJPrarediseases.org Horizon2020 #825575

• Capable-project.eu Horizon2020 #875052
Outline

• Use and reuse of healthcare data
• FAIR data Principles
• How to make healthcare data FAIR
• What’s next?
Use and reuse of healthcare data

New Yorker June 26th, 1995, drawing by J.P. Rini
Practice Guidelines for the Management of Bacterial Meningitis

Suspicion for bacterial meningitis

- Yes

  Immunocompromise, history of CNS disease, new onset seizure, papilledema, altered consciousness, or focal neurologic deficit\(^a\) or delay in performance of diagnostic lumbar puncture

  - No
  
  Blood cultures and lumbar puncture STAT

  - Yes

    Dexamethasone\(^b\) + empirical antimicrobial therapy\(^c\)

    - Yes

      CSF findings c/w bacterial meningitis

      - Yes

        Positive CSF Gram stain

        - No

        Dexamethasone\(^b\) + empirical antimicrobial therapy\(^c\)

      - Yes

        Blood cultures STAT

        - Yes

          Dexamethasone\(^b\) + empirical antimicrobial therapy\(^c\)

          - Yes

            Negative CT scan of the head

            - Yes

              Perform lumbar puncture

              - No

                Dexamethasone\(^b\) + targeted antimicrobial therapy\(^d\)

\(^a\) or delay in performance of diagnostic lumbar puncture

<table>
<thead>
<tr>
<th>Data should be</th>
<th>FAIR Guiding Principles</th>
<th><a href="https://go-fair.org/">https://go-fair.org/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Findable</strong></td>
<td>F1. (meta)data are assigned a <strong>globally unique and persistent identifier (DOI)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2. data are described with rich metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F3. metadata clearly and explicitly include the identifier of the data it describes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F4. (meta)data are registered or indexed in a searchable resource</td>
<td></td>
</tr>
<tr>
<td><strong>Accessible</strong></td>
<td>A1. (meta)data are retrievable by their identifier using a standardized communications protocol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1.1 the protocol is open, free, and universally implementable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1.2 the protocol allows for an authentication and authorization procedure, where necessary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2. metadata are accessible, even when the data are no longer available</td>
<td></td>
</tr>
<tr>
<td><strong>Interoperable</strong></td>
<td>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I2. (meta)data use vocabularies that follow FAIR principles</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I3. (meta)data include qualified references to other (meta)data</td>
<td></td>
</tr>
<tr>
<td><strong>Reusable</strong></td>
<td>R1. <strong>meta</strong>(data) are richly described with a plurality of accurate and relevant attributes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1.1. (meta)data are released with a clear and accessible data usage license</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1.2. (meta)data are associated with detailed provenance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1.3. (meta)data meet domain-relevant community standards</td>
<td></td>
</tr>
</tbody>
</table>
FAIR Principles - concise

• Findable
  • Metadata and data should be easy to find for both humans and computers

• Accessible
  • The user needs to know how data can be accessed, possibly including authentication and authorization

• Interoperable
  • Data need to be integrated with other data and interoperate with applications for analysis, storage, and processing

• Reusable
  • (Licensing & provenance) metadata and data should be well-described so that they can be replicated and/or combined in different settings
FAIR Principles = “What”, not “how”

- Globally unique and persistent identifiers
  - https://orcid.org/0000-0002-1704-5980
  - https://www.linkedin.com/in/ronaldcornet/
  - ...
- Freedom of format
Examples of (more or less) FAIR repositories

- https://home.fairdatapoint.org/  ─ Links to FAIR data points
- https://fairsharing.org/
- https://www.openaire.eu/
- https://www.ohdsi.org/  ─ “Human” entry to harmonized data
FAIR healthcare data - focus on “I”

• Use “knowledge organization systems”: ontologies, vocabularies, terminologies
  1. Identification
  2. Characterization
  3. Organization
 Identification

- 350,000 concepts
- Versions available in multiple languages (English, Spanish, Dutch, Swedish, ...)
- Licensed to over 40 countries
Characterization
FAIR healthcare data - focus on “I”

• Use “knowledge organization systems”: ontologies, vocabularies, terminologies
  1. Identification
  2. Characterization
  3. Organization
• For (categorial) values (diseases, procedures, etc.)
• For data elements (“diagnosis”, “blood pressure”, etc.)
**Identification Characterization**

<table>
<thead>
<tr>
<th>LOINC CODE</th>
<th>LONG COMMON NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>8480-6</td>
<td>Systolic blood pressure</td>
</tr>
</tbody>
</table>

**Fully-Specified Name**

<table>
<thead>
<tr>
<th>Component</th>
<th>Intravascular systolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property</td>
<td>Pres</td>
</tr>
<tr>
<td>Time</td>
<td>Pt</td>
</tr>
<tr>
<td>System</td>
<td>Arterial system</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn</td>
</tr>
<tr>
<td>Method</td>
<td></td>
</tr>
</tbody>
</table>

**Additional Names**

<table>
<thead>
<tr>
<th>Short Name</th>
<th>BP sys</th>
</tr>
</thead>
</table>
LOINC, SNOMED CT & Many Others

- BioPortal includes about 900 biomedical ontologies, over 13M concepts/classes

<table>
<thead>
<tr>
<th>BioPortal Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontologies</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Properties</td>
</tr>
<tr>
<td>Mappings</td>
</tr>
</tbody>
</table>
Next steps - Semantic Web Standards

- OWL - ontologies
- ShEx - (clinical) data models
- RDF - instances

Open license → structured → open format → URI-based → linked
Semantic Web Standards

• Representation to enable **Reasoning**
• Important to detect “iso-semantic expressions”, i.e., different ways to convey the same meaning, e.g.,

• Measurement:
  • type = 8459-0 | Systolic blood pressure--sitting
• Measurement:
  • type = 8480-6 | Systolic blood pressure
  • 704326004 | Precondition = 33586001 | Sitting position (finding)
Semantic Web Services

• Ontology alignment services
  • Static, e.g., UMLS, Athena
  • Dynamic, e.g., AML, FCA-Map, LogMap

• Instance alignment services (e.g., https://www.sameas.cc/)
Conclusion

- Ontologies such as SNOMED CT play an increasing role in clinical data capture
- Information models benefit from use of systems such as LOINC, but this is not yet common practice
- Aligning data represented using different ontologies and different information models is challenging
- Semantic web technology can contribute to overcoming this
- Uniquely identifying data elements and values using broadly accepted standards contributes to healthcare data becoming FAIR
Ronald Cornet, PhD | Associate professor ; Principal Investigator Medical Informatics
Amsterdam Public Health Research Institute
Location AMC | J1b-115 | Meibergdreef 15, 1105 AZ Amsterdam
T: +31 20-566 51 88 | E:
r.cornet@amsterdamumc.nl | kik.amc.nl/home/rcornet
www.amsterdamumc.nl | www.amc.nl