

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?



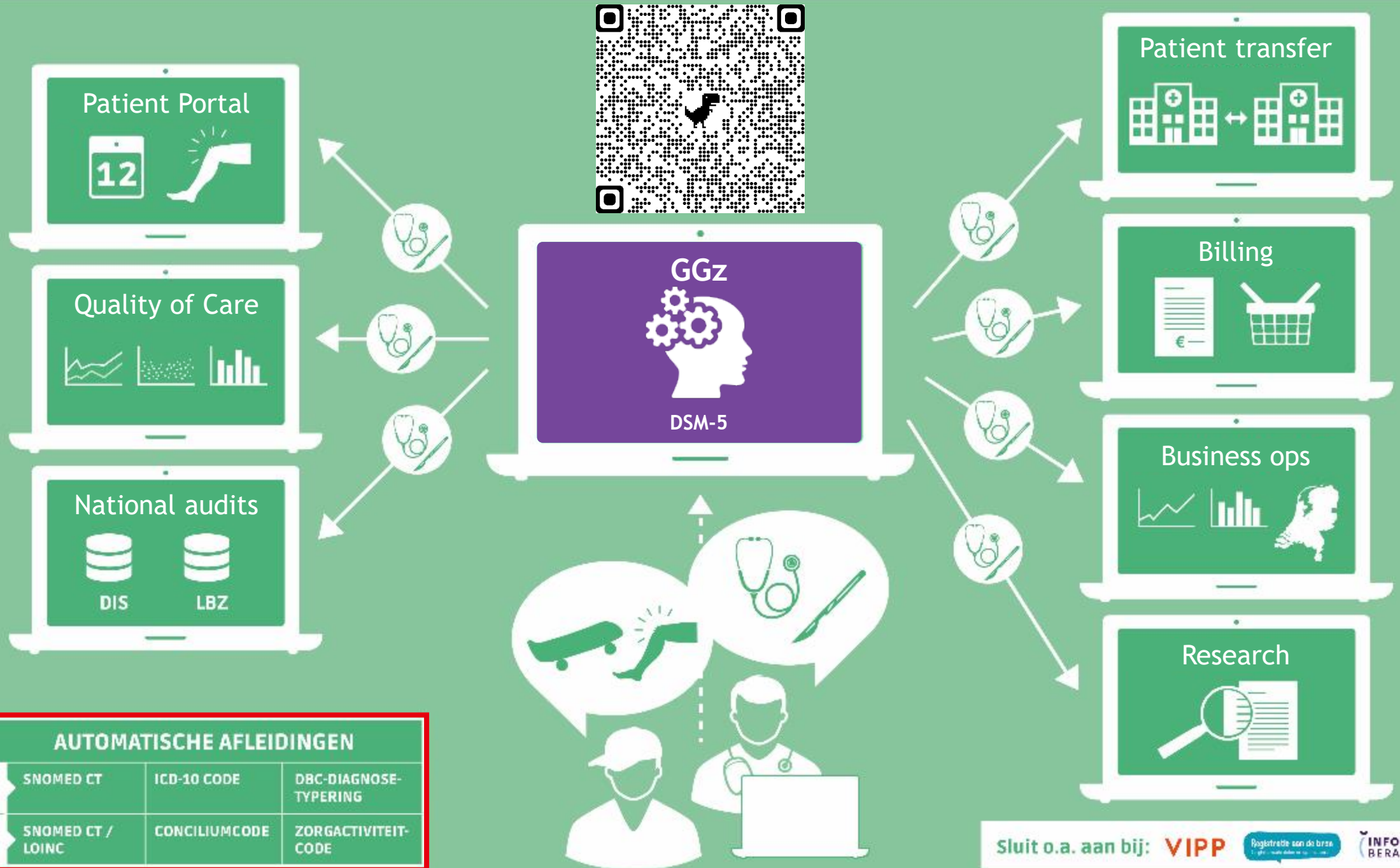
KNOWLEDGE ORGANIZATION
RESEARCH OBSERVATORY





Use and reuse of healthcare data





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 ↙

↘ Yes

Blood cultures and lumbar puncture STAT

Blood cultures STAT

Dexamethasone^b + empirical antimicrobial therapy^{c,e}

Dexamethasone^b + empirical antimicrobial therapy^c

CSF findings c/w bacterial meningitis

Negative CT scan of the head

Positive CSF Gram stain

Perform lumbar puncture

No ↙

↘ Yes

Dexamethasone^b + empirical antimicrobial therapy^c

Dexamethasone^b + targeted antimicrobial therapy^d

Criterion

Comment

Immunocompromised state

HIV infection or AIDS, receiving immunosuppressive therapy, or after transplantation

History of CNS disease

Mass lesion, stroke, or focal infection

New onset seizure

Within 1 week of presentation; some authorities would not perform a lumbar puncture on patients with prolonged seizures or would delay lumbar puncture for 30 min in patients with short, convulsive seizures

Papilledema

Presence of venous pulsations suggests absence of increased intracranial pressure

Abnormal level of consciousness

...

Focal neurologic deficit

Including dilated nonreactive pupil, abnormalities of ocular motility, abnormal visual fields, gaze palsy, arm or leg drift

FAIR Guiding Principles



<https://go-fair.org/>



Data should be Findable

- F1. (meta)data are assigned a globally unique and persistent identifier (DOI)
- F2. data are described with rich metadata
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource



Data should be Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available



Data should be Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data



Data should be Reusable

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



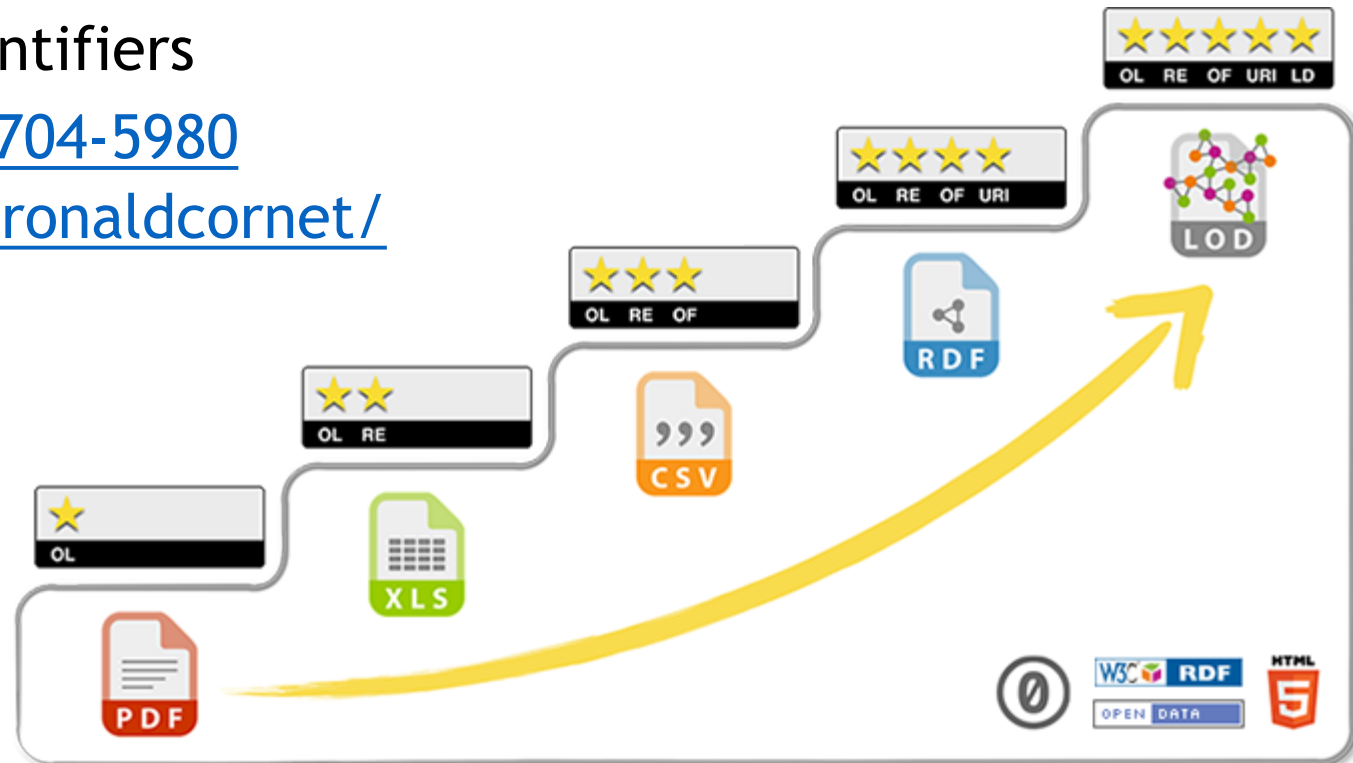
FAIR Principles - concise

- 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/ronalddcornet/>
 - ...
- Freedom of format



Open license → structured → open format → URI-based → linked



Examples of (more or less) FAIR repositories

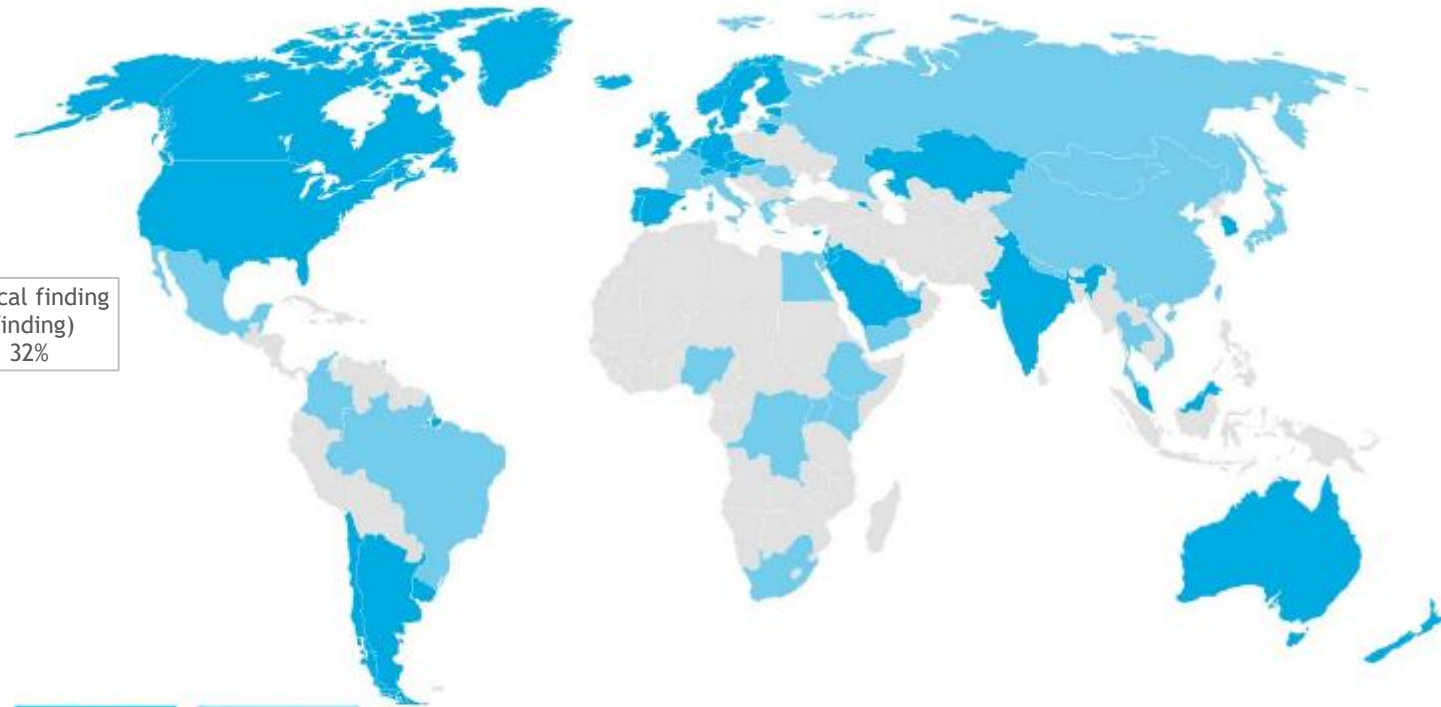
- <https://home.fairdatapoint.org/> ← Links to FAIR data points
- <https://fairsharing.org/>
- <https://www.openaire.eu/>
- <https://www.ohdsi.org/> ← “Human” entry to harmonized data





FAIR healthcare data - focus on “I”

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



Member

Affiliate Licensee

AMERICAS

- Argentina
- Canada
- Chile
- United States
- Uruguay

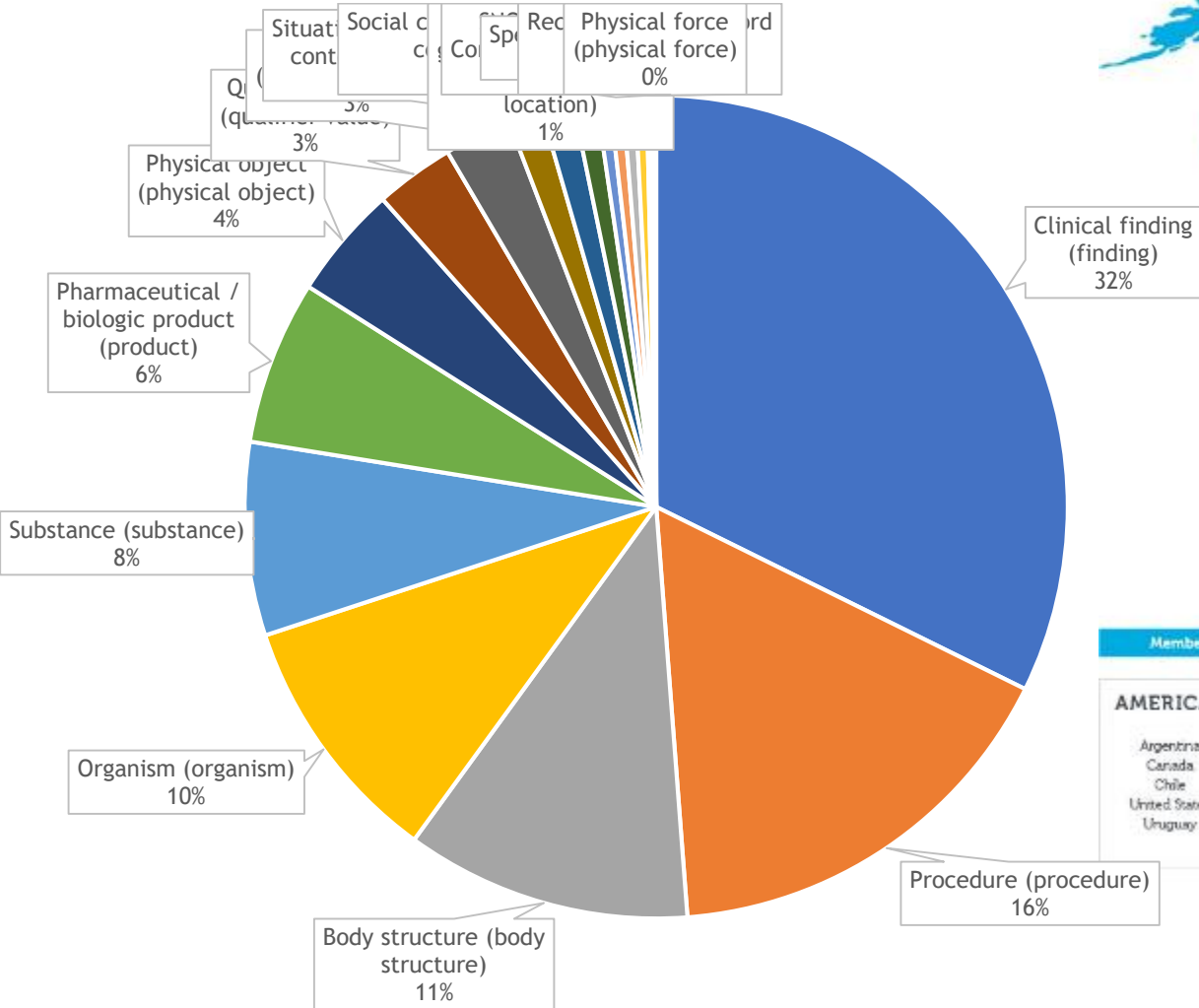
EUROPE, MIDDLE EAST & AFRICA

- | | | | | |
|----------------|---------|------------|----------------------|-----------------|
| Austria | Estonia | Israel | Netherlands | Slovak Republic |
| Belgium | Finland | Jordan | Norway | Spain |
| Cyprus | Germany | Lithuania | Portugal | Sweden |
| Czech Republic | Iceland | Luxembourg | Republic of Slovenia | Switzerland |
| Denmark | Ireland | Malta | Saudi Arabia | United Kingdom |

ASIA PACIFIC

- | | |
|------------|---------------------|
| Australia | Malaysia |
| Bruni | New Zealand |
| Hong Kong | Republic of Armenia |
| China | Republic of Korea |
| India | Singapore |
| Kazakhstan | |

Updated 01/01/2021





Characterization

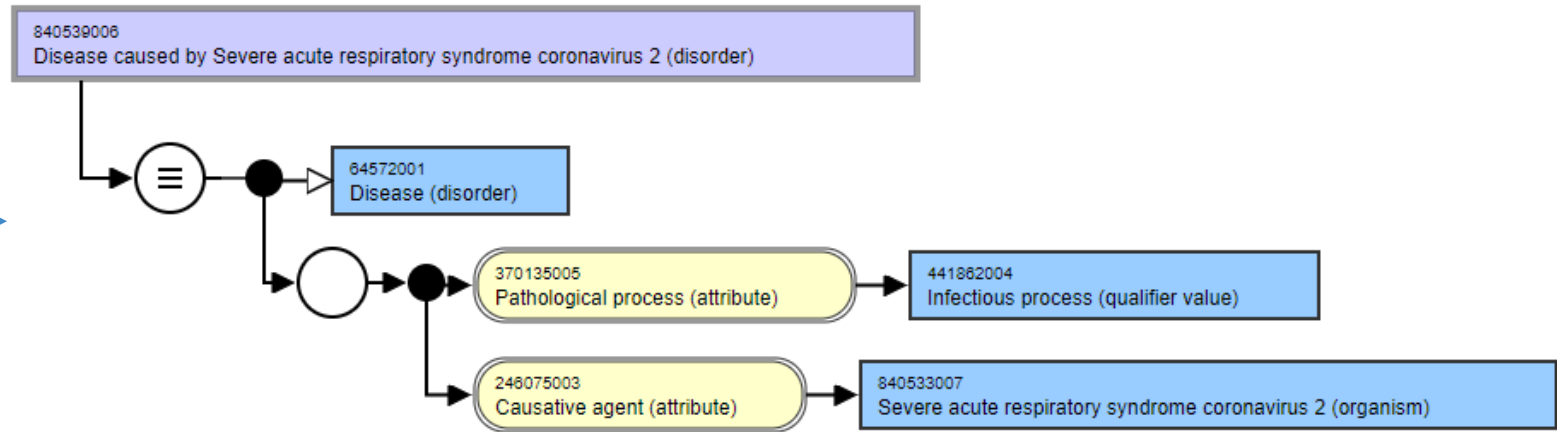
SNOMED International
Leading healthcare terminology, worldwide

INTERNATIONAL PASSPORT

☰ Disease caused by Severe acute respiratory syndrome coronavirus 2 (disorder) ☆ ↗
SCTID: 840539006

840539006 | Disease caused by Severe acute respiratory syndrome coronavirus 2 (disorder) |

- en COVID-19
- en Disease caused by Severe acute respiratory syndrome coronavirus 2 (disorder)
- en Disease caused by Severe acute respiratory syndrome coronavirus 2
- en Disease caused by 2019-nCoV
- en Disease caused by 2019 novel coronavirus





Organization

INTERNATIONAL PASSPORT

● Severe acute respiratory syndrome coronavirus 2 (organism) ☆

SCTID: 840533007

840533007 | Severe acute respiratory syndrome coronavirus 2 (organism) |

- en Severe acute respiratory syndrome coronavirus 2 (organism)
- en 2019-nCoV
- en Severe acute respiratory syndrome coronavirus 2
- en SARS-CoV-2
- en 2019 novel coronavirus

- ● Subfamily Orthocoronavirinae (organism)
- ▲ ● Human coronavirus (organism)
- ● Subfamily Orthocoronavirinae (organism)
- ▲ ● Genus Betacoronavirus (organism)
- ▲ ● Subgenus Sarbecovirus (organism)



FAIR healthcare data - focus on “I”

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



LOINC, SNOMED CT & Many Others

- [BioPortal](#) includes about 900 biomedical ontologies, over 13M concepts/classes

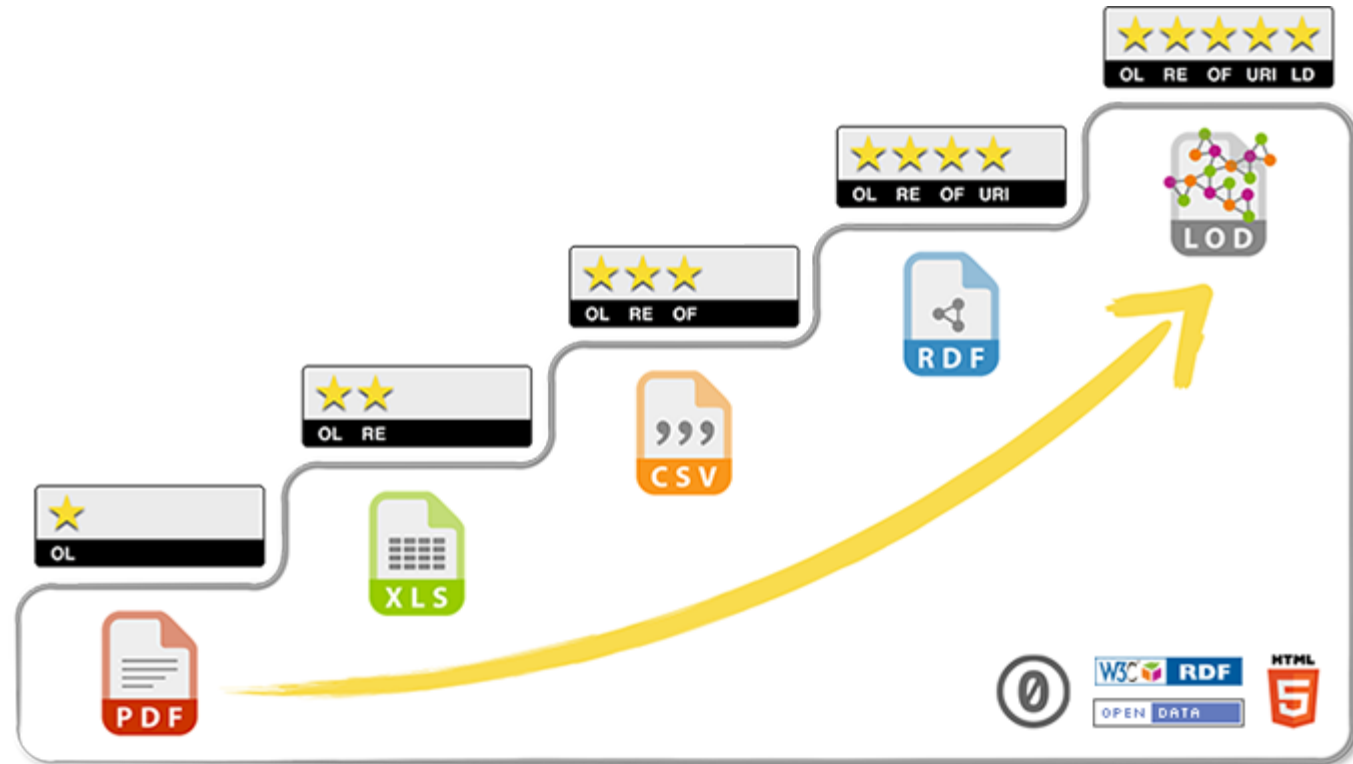
BioPortal Statistics	
Ontologies	891
Classes	13,519,971
Properties	36,286
Mappings	55,648,584



Next steps - Semantic Web Standards

Data Metadata

- 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

