Access to COVID-19 research data
The European COVID-19 Data Platform

Niklas Blomberg, Director, ELIXIR
OECD GSF / RDA workshop
23 April 2021
Open data is extensively reused, also in pandemics

CORONAVIRUS

Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody

Rappazzo et al, Science 19 Feb 2021, 10.1126/science.abf4830
What is SARS-CoV-2 genome?

Showing 15 of 161,552 in Viral sequences

<table>
<thead>
<tr>
<th>Accession</th>
<th>Collection date</th>
<th>Country</th>
<th>Host</th>
<th>Strain</th>
<th>Isolate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN908947</td>
<td>Dec, 2019</td>
<td>China</td>
<td>Homo sapiens</td>
<td>Wuhan-Hu-1</td>
<td></td>
</tr>
<tr>
<td>LR991698</td>
<td>Sep 21, 2020</td>
<td>United Kingdom</td>
<td>Homo sapiens</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data types

- All (1,213,180)
- Sequences (161,592)
- Reference sequences (2)
- Raw reads (518,656)
- Sequenced samples (519,885)
- Studies (331)
- Genes (32)

powered by ENA
Open data identifiers are a cornerstone of modern bioeconomy

Global, open and unique identifiers from open research data are used in patents, software, and data management systems to unambiguously reference biological entities.
European COVID-19 Data Platform

- Open and rapid access to data, tools and workflows
- Built upon open standards
- Global data coverage and global access
- Enabling diverse research to fight COVID-19

https://www.covid19dataportal.org/
COVID-19 Data Portal

- Over 750,000 records across molecular platforms and literature
- Web, API and download
- 126,000 users, 4 million requests

Viral Sequences
- All (621,031)
- Sequences (59,050)
- Reference sequences (1)
- Raw reads (288,428)
- Sequenced samples (260,578)
- Studies (260)
- Genes (22)
- Browser (1)
- Variants (12,691)

Host Sequences
- All (10,377)
- Human studies (controlled access) (7)
- Human reads (consented for full access) (7,759)
- Other species reads (2,573)
- Association studies (38)

Expression
- All (87)
- Gene expression (4)
- Gene expression experiments (23)
- Single cell expression (4)
- Single cell expression experiments (19)
- Protein expression experiments (37)

Proteins
- All (1,045)
- Protein sequences (92)
- Protein families (315)
- Protein structures - Knowledge Base (7)
- Protein structures (260)
- Electron microscopy density maps (355)
- Electron microscopy public image archive (16)

Biochemistry
- All (2,710)
- Pathways (16)
- Interactions (2,202)
- Complexes (29)
- Compound document (8)
- Drug targets (453)
- Metabolomics experiments (2)

Imaging
- All (18)
- Images (2)
- Electron microscopy public image archive (16)

Literature
- All (297,147)
- Coronaviruses (150,188)
- Diseases (138,267)
- Related viruses and diseases (2,299)
- Genes, receptors and antibodies (6,388)
- Supplementary material (5)

https://www.covid19dataportal.org/
• What are the main challenges for access to research data in this field in relation to COVID-19?
  • “Mobilisation” – marshal the data, make it ready for reuse
• How are these challenges being addressed?
• What is needed to be better prepared for future crises?
  • Move pioneering efforts and
• What are the implications for science policy (taking note of the OECD Research Data recommendation and RDA work)?
Data challenges and future implications

“The normative value of open data in the knowledge economy”
available to all – openly referenceable – can be integrated

Mobilise data for reuse – across domains and countries

Build on the pioneering research community efforts – Connected infrastructure of future pandemics

FAIR Research Data – identify, reference and mix data has large societal value