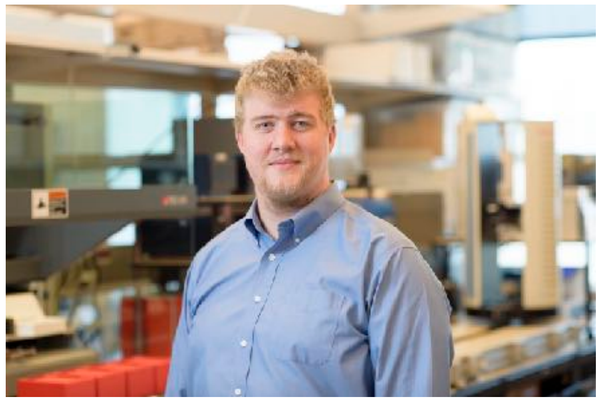
BioExcel Webinar Series #47

COVID-19 Molecular Structure and Therapeutics Hub

Presenter: Levi Naden (Molecular Sciences Software Institute (MolSSI))

Host: Rossen Apostolov (KTH Royal Institute of Technology)

Today's Presenter



Levi Naden

Software Scientist at MolSSI
lnaden@vt.edu

Levi Naden is a Software Scientist at MolSSI and lead for COVID-19 Molecular Structure and Therapeutics Hub project. Levi has a background in computational free energy algorithms, scientific software development and deployment, and software interoperability. He has a Bachelors in Chemical Engineering from Kansas State University and a PhD in Chemical Engineering from University of Virginia.

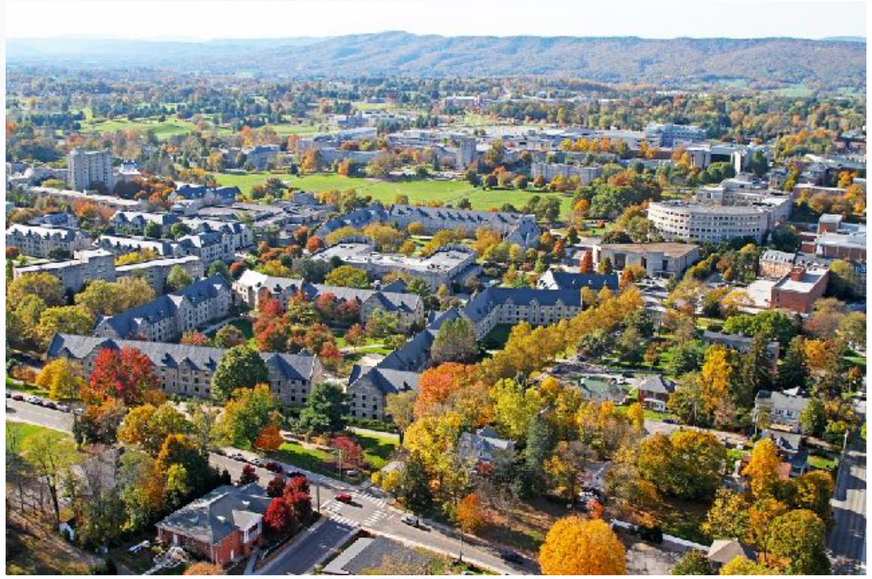
The Molecular Sciences Software Institute: https://www.molssi.org/

What is MolSSI?

Designed to serve and enhance the software development efforts of the broad field of computational molecular science (CMS) – a domain that includes quantum chemistry, computational materials science, and biomolecular simulation.

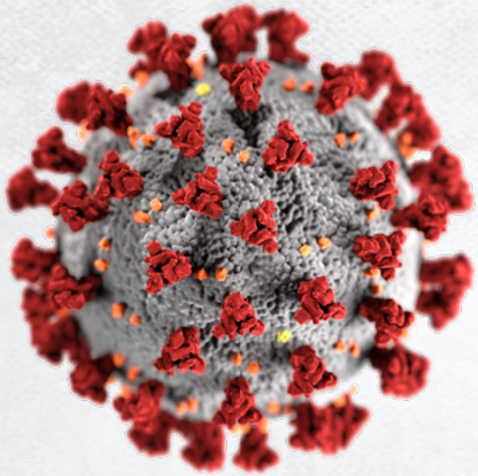
Goals:

1. Provide software expertise and infrastructure
2. Provide education and training
3. Provide community engagement and leadership



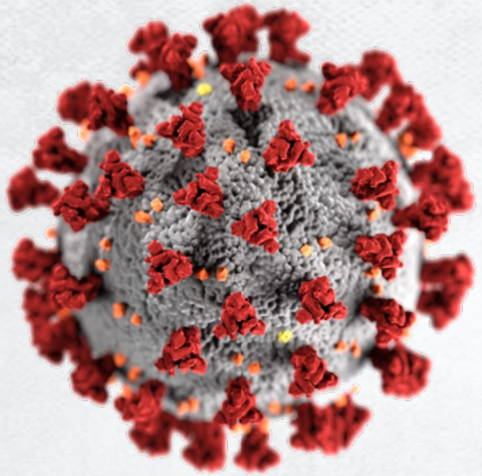
There are ups and downs to the current state of research

The Enemy



There are ups and downs to the current state of research

The Enemy



The Good

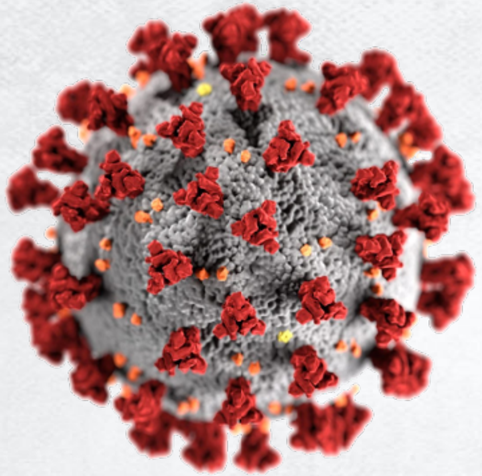
Covid-19 Changed How the World Does Science, Together

Never before, scientists say, have so many of the world's researchers focused so urgently on a single topic. Nearly all other research has ground to a halt.

Source: New York Times

There are ups and downs to the current state of research

The Enemy



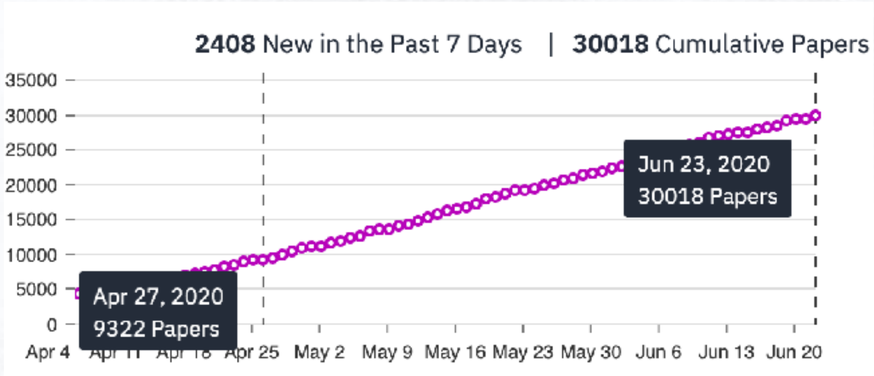
The Good

Covid-19 Changed How the World Does Science, Together

Never before, scientists say, have so many of the world's researchers focused so urgently on a single topic. Nearly all other research has ground to a halt.

Source: New York Times

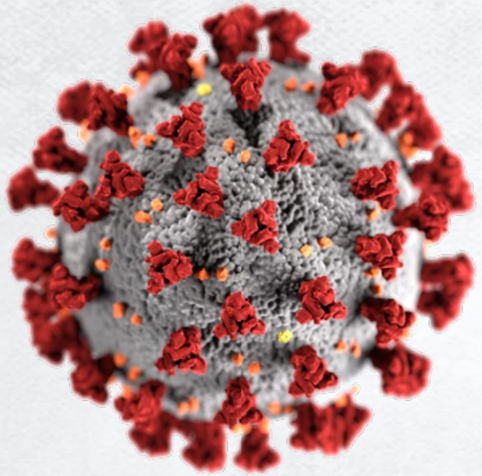
The Problems



Total Volume of COVID-19 Research Papers
Source: covid19primer by Primer.ai.

There are ups and downs to the current state of research

The Enemy



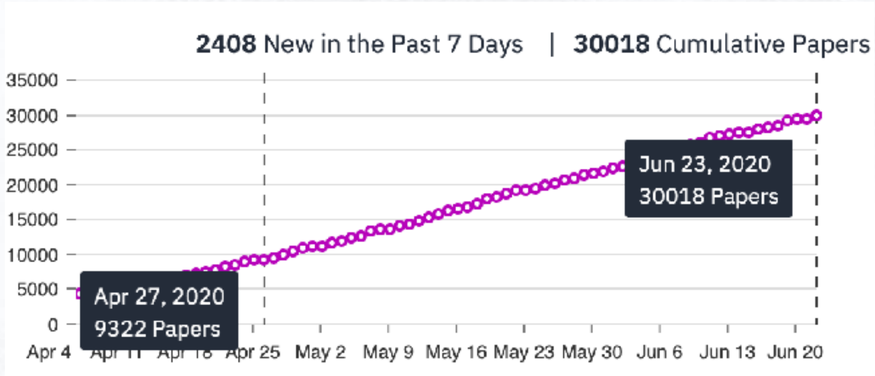
The Good

Covid-19 Changed How the World Does Science, Together

Never before, scientists say, have so many of the world's researchers focused so urgently on a single topic. Nearly all other research has ground to a halt.

Source: New York Times

The Problems



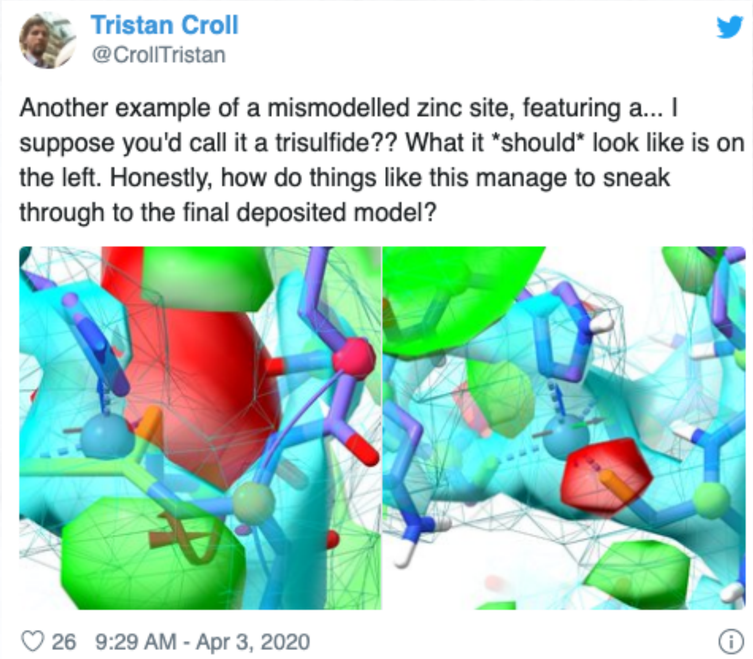
Total Volume of COVID-19 Research Papers
Source: covid19primer by Primer.ai.

Lots of scientists stuck at home right now looking to help

Source: First google search for "antsy person"

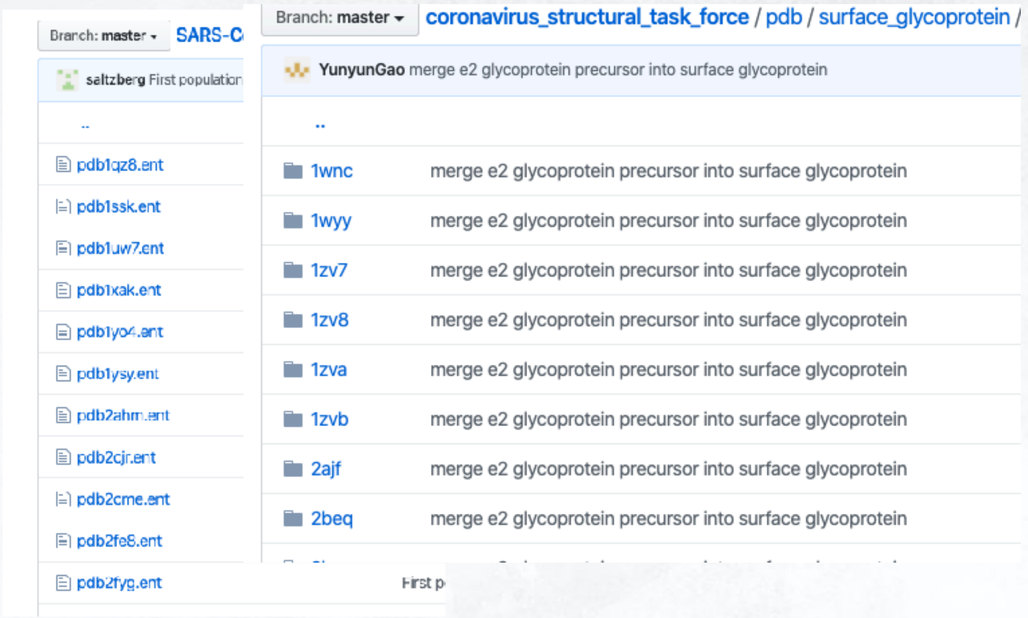
Every institute and lab under the sun is doing something with SARS-CoV-2

- Overall a good thing, but has downsides
  - Editors: “We are being overwhelmed with papers”
  - Quality: Mixed bag
  - Collaborators: We have data, people want data, how do I get them data?



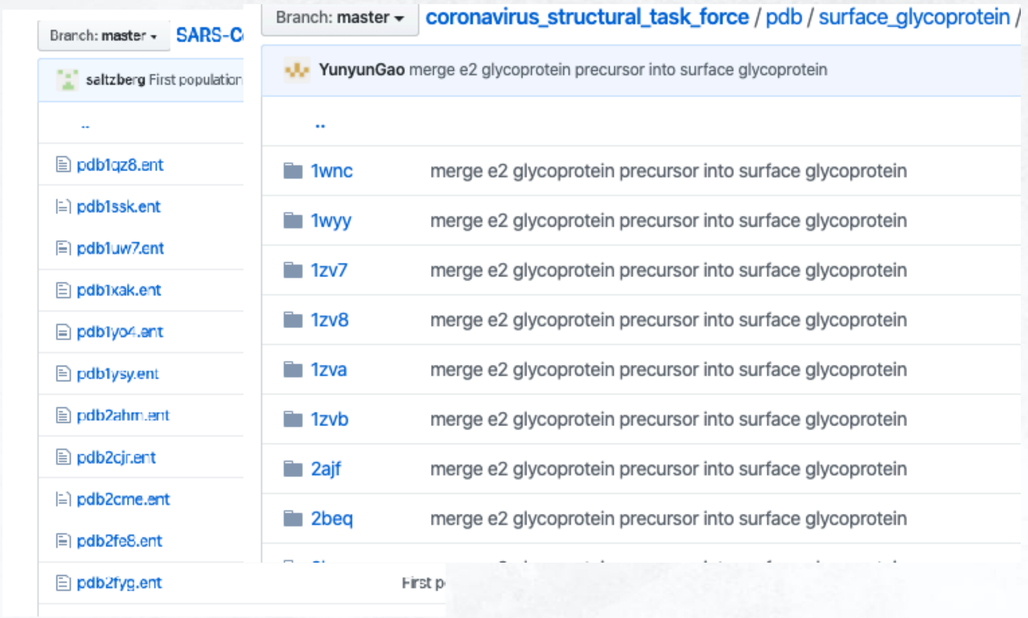
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!



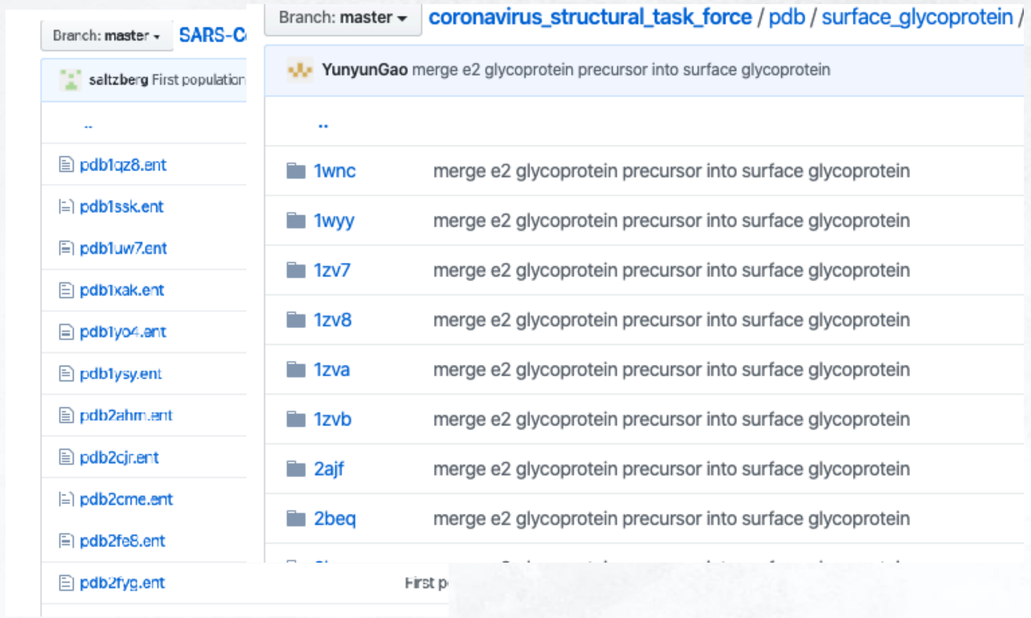
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways



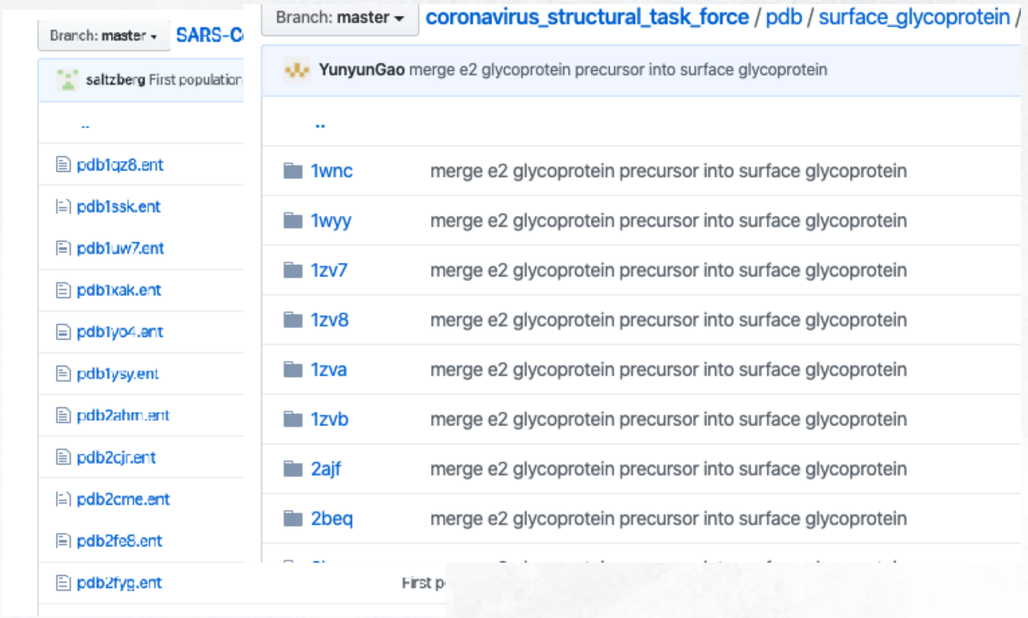
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret



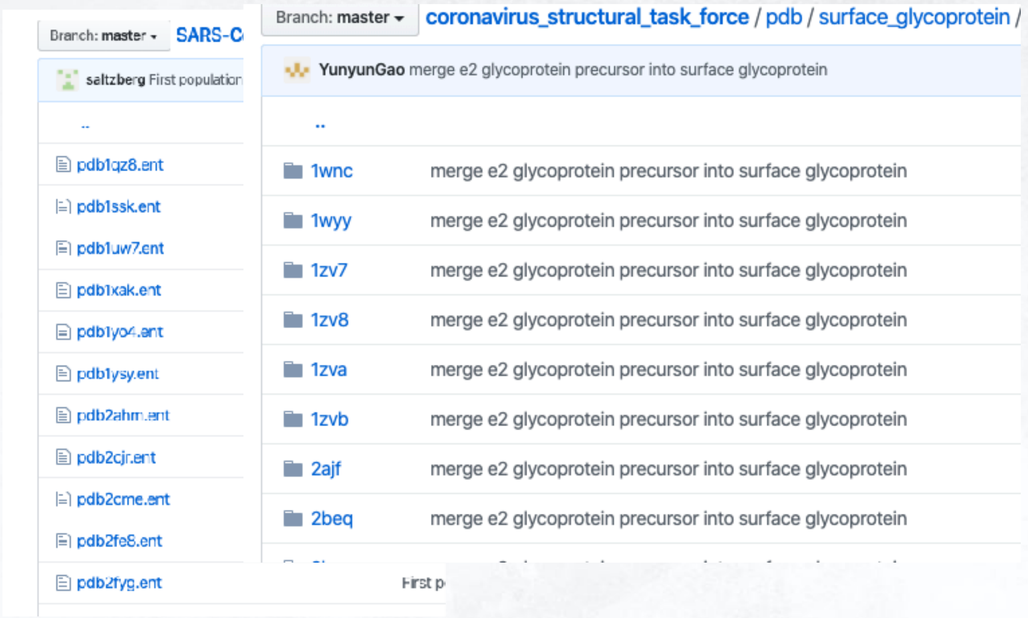
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?



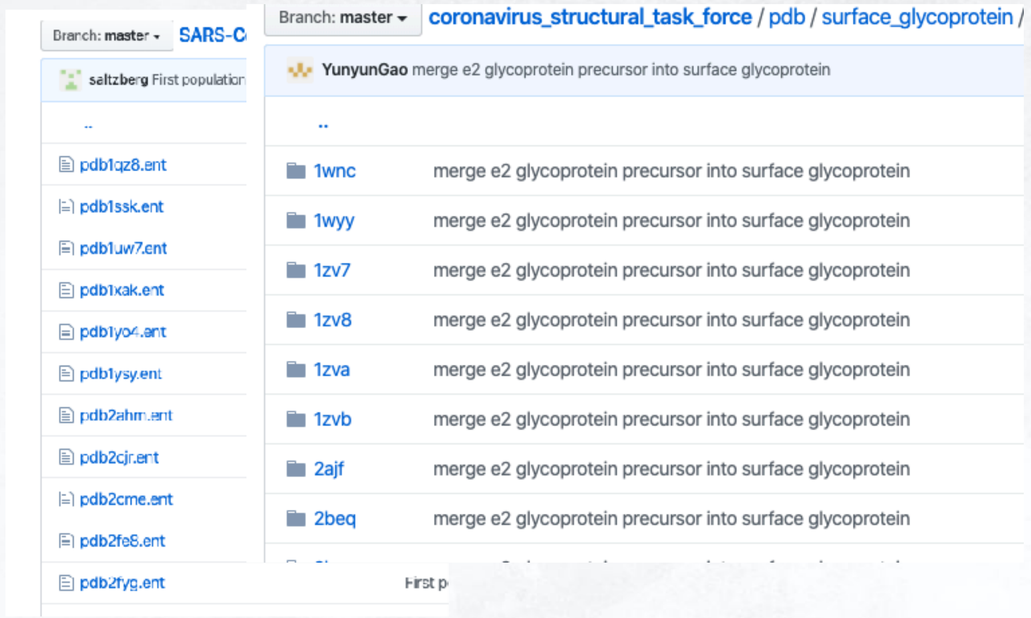
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?



Even good data are nebulous

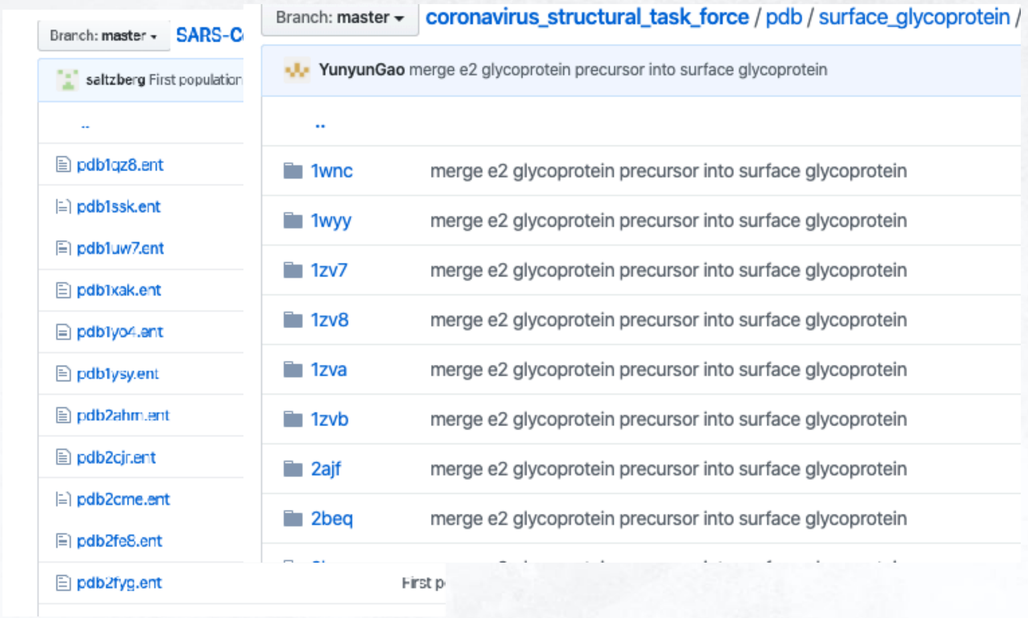
- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?



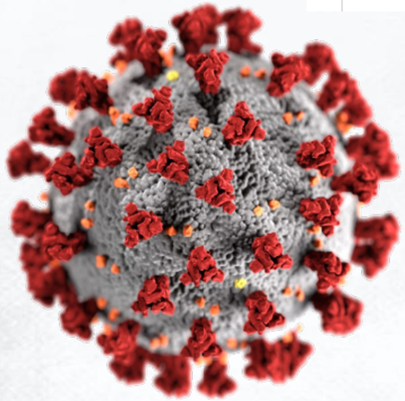
- Most data are only small parts of the large picture

Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?

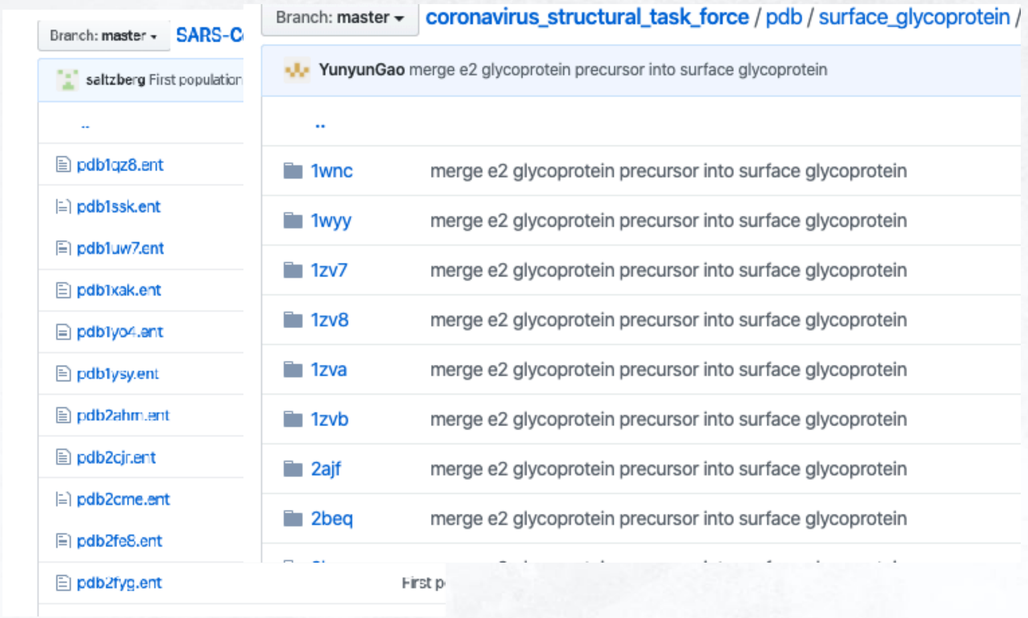


- Most data are only small parts of the large picture

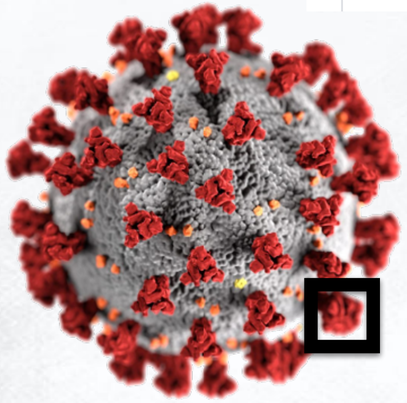


Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?

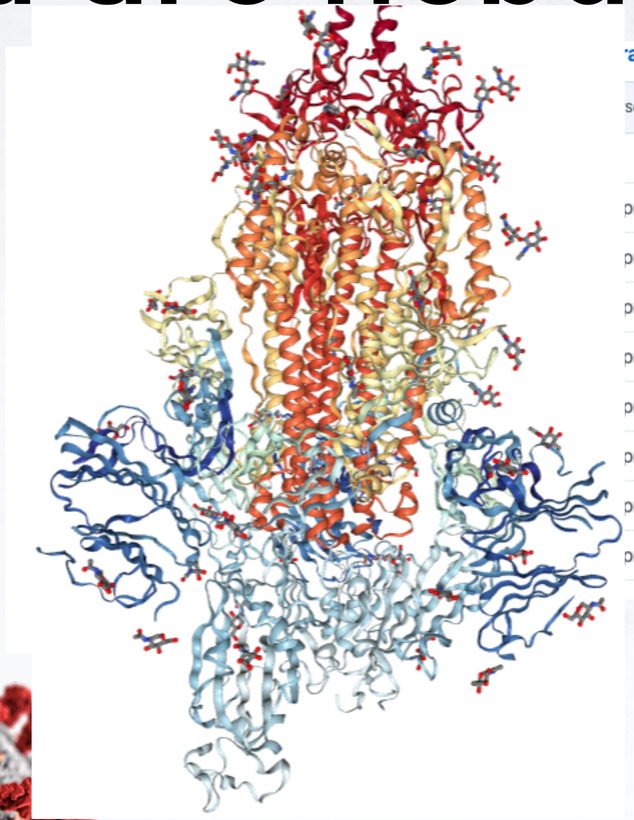


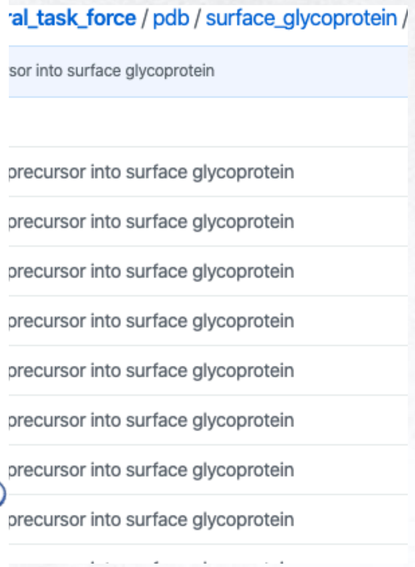
- Most data are only small parts of the large picture



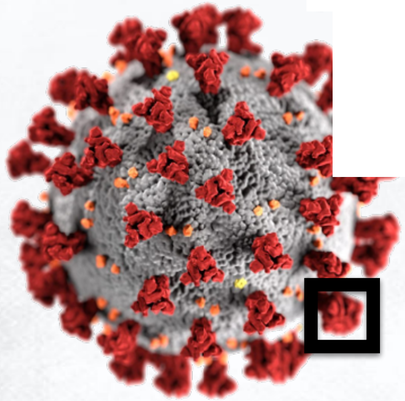
Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?





- Most data are only small parts of the large picture

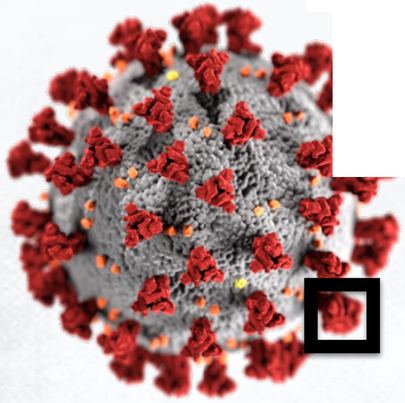


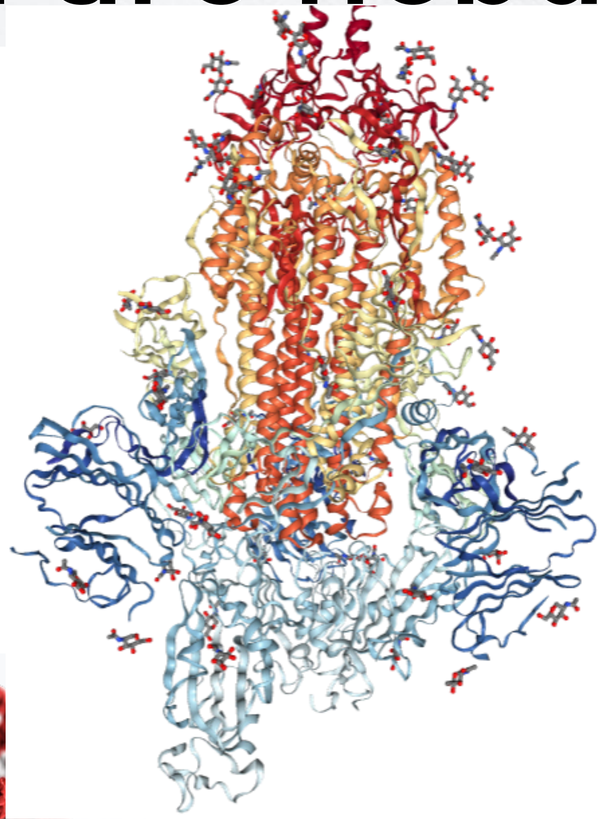
Spike

Even good data are nebulous

- Everyone does their own work
  - And is great in their own rights!
- The data are all organized in domain specific ways
- Only domain experts can interpret
- Whats good? Whats not?

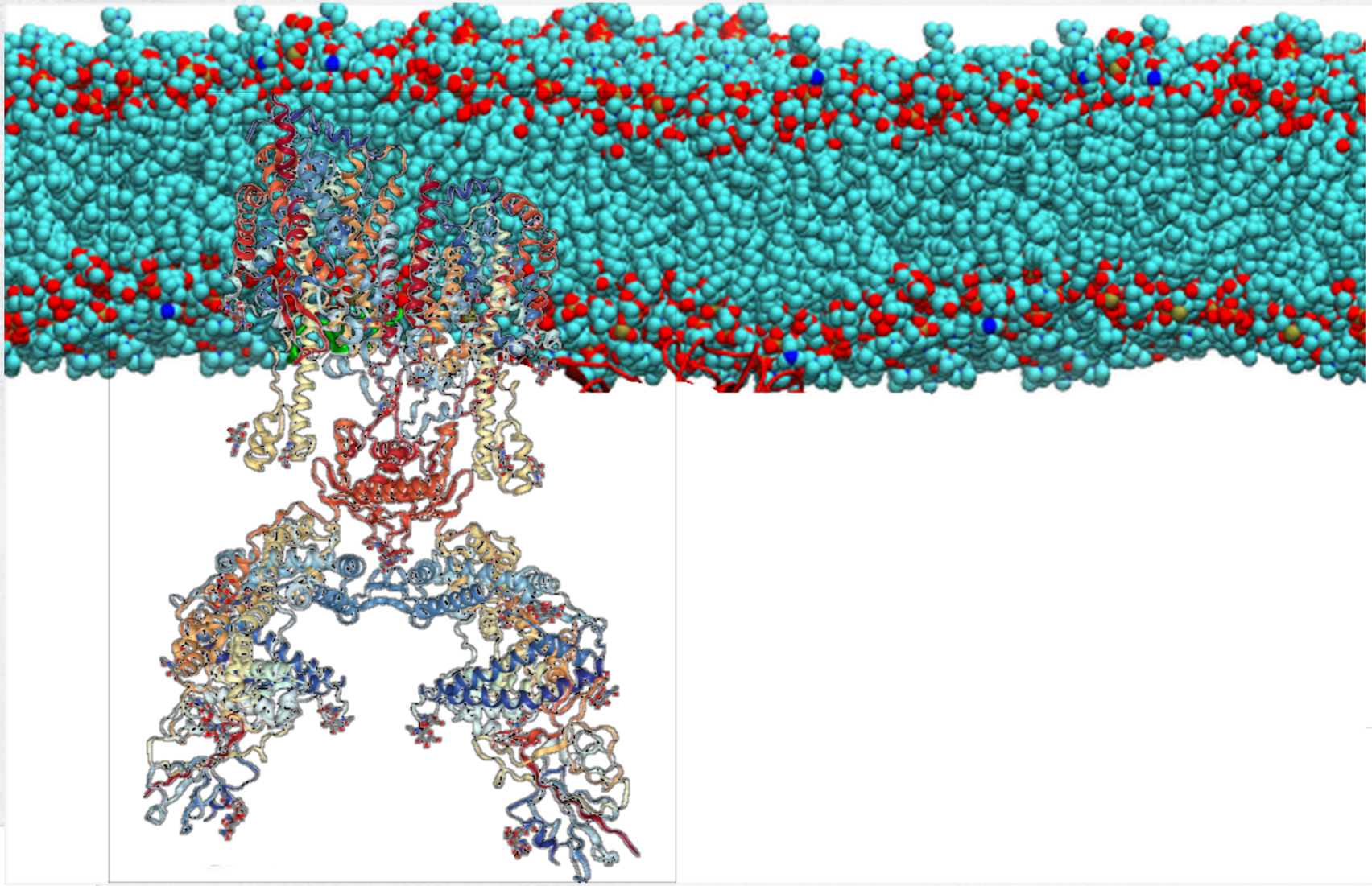
- Most data are only small parts of the large picture

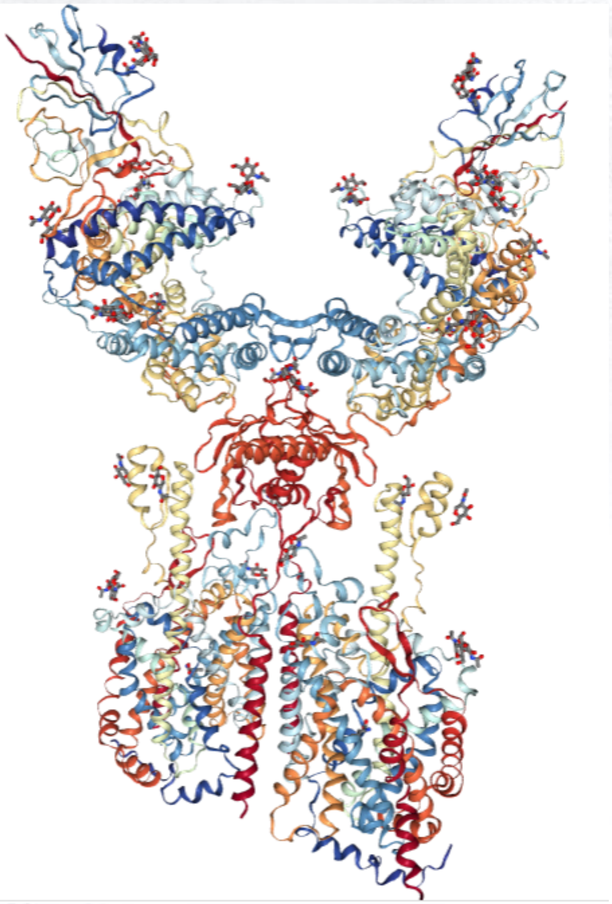




Spike

ACE2





ACE2

Modified from: Human Brain Project

We needed a hub for easily conveyed science

Simple descriptors
Community peer reviewed
Centralized or pointers to data
Accessible to experts in all domains

Two PI's sparked what is now a global network of cooperation

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Adrian Mulholland

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Teresa Head-Gordon

Adrian Mulholland

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Teresa Head-Gordon

Adrian Mulholland

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Cecilia Clementi

Teresa Head-Gordon

Adrian Mulholland

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Cecilia Clementi

Teresa Head-Gordon

Adrian Mulholland

Prototype (3 months ago)

Two PI's sparked what is now a global network of cooperation

Rommie Amaro

Cecilia Clementi

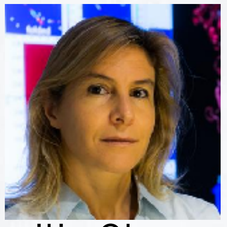
Teresa Head-Gordon

Adrian Mulholland

Prototype (3 months ago)

Two PI's sparked what is now a global network of cooperation

Rommie Amaro



Cecilia Clementi

Teresa Head-Gordon

Adrian Mulholland

MOLSSI

Prototype (3 months ago)

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

D E Shaw Research

RIKEN BDR

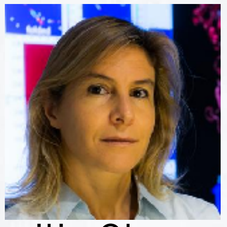
FOLDING @HOME

MOLSSI

bioexcel

Two PI's sparked what is now a global network of cooperation

Rommie Amaro



Cecilia Clementi

Teresa Head-Gordon

Adrian Mulholland

Prototype (3 months ago)

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

D E Shaw Research

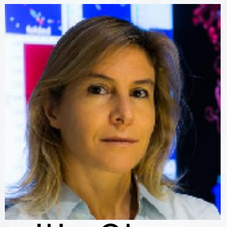
RIKEN BDR

FOLDING @HOME

One month a whole bunch of this later...

Two PI's sparked what is now a global network of cooperation

Rommie Amaro



Cecilia Clementi

Teresa Head-Gordon

Adrian Mulholland

MOLSSI

Prototype (3 months ago)

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

D E Shaw Research

RIKEN BDR

FOLDING @HOME

One month a whole bunch of this later...

Dozens of individual contributors

Community in US, EU, Switzerland, Korea, Japan

Upwards of a dozen Orgs./Institutes aiding or in talks

covid.molssi.org

MOLSSI

bioexcel

covid.molssi.org

Let's take a look at what is there!

github.com/MolSSI/covid

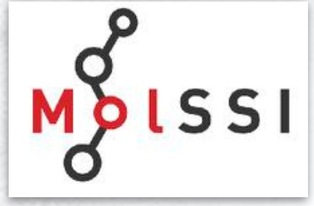
The Hub is bringing in companies to better aid researchers

- DataCite for issuing DOIs on demand
  - Make data citable in papers!

The Hub is bringing in companies to better aid researchers

- DataCite for issuing DOIs on demand
  - Make data citable in papers!

- AWS Open Data Program for storage
  - 100+ TB of storage
  - Explore options of storing the data we generate correctly



The Hub is bringing in companies to better aid researchers

- DataCite for issuing DOIs on demand
  - Make data citable in papers!

- AWS Open Data Program for storage
  - 100+ TB of storage
  - Explore options of storing the data we generate correctly

- Zenodo storage for everyone and a group to link data

The Hub is bringing in companies to better aid researchers

- DataCite for issuing DOIs on demand
  - Make data citable in papers!

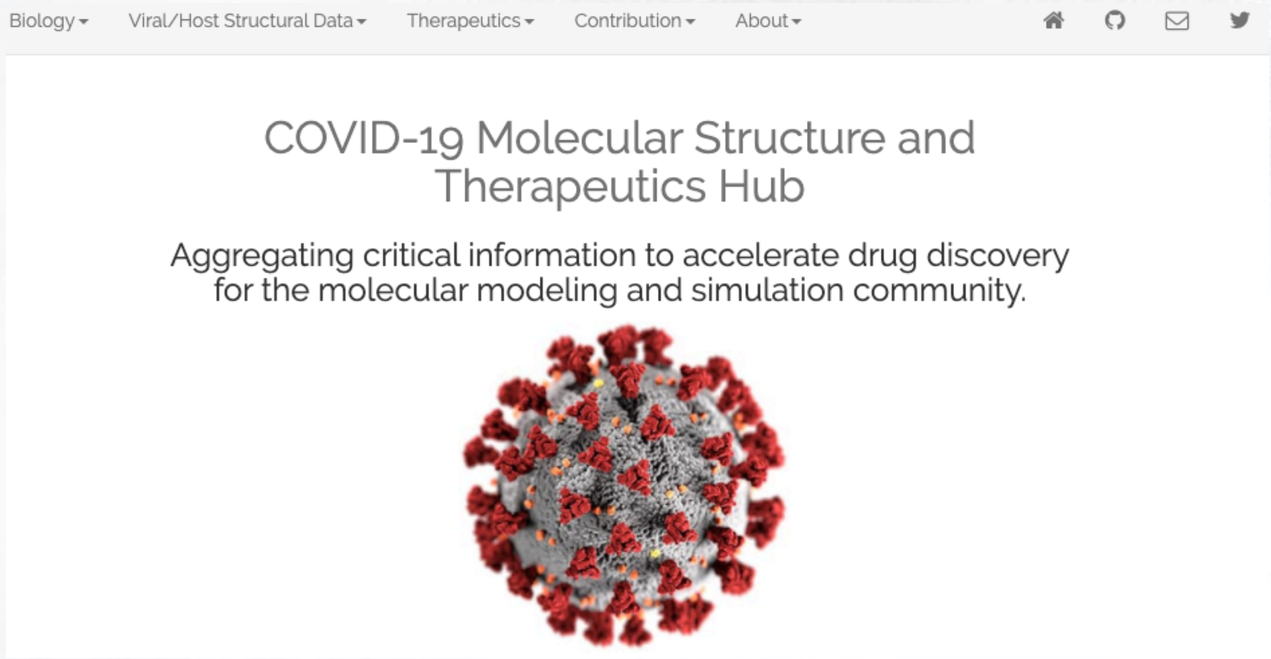
- AWS Open Data Program for storage
  - 100+ TB of storage
  - Explore options of storing the data we generate correctly

- Zenodo storage for everyone and a group to link data

- Potential Improvements:
  - Knowledge Graphs for data
  - Data licensing schemes

The Hub showed unexpected social insights

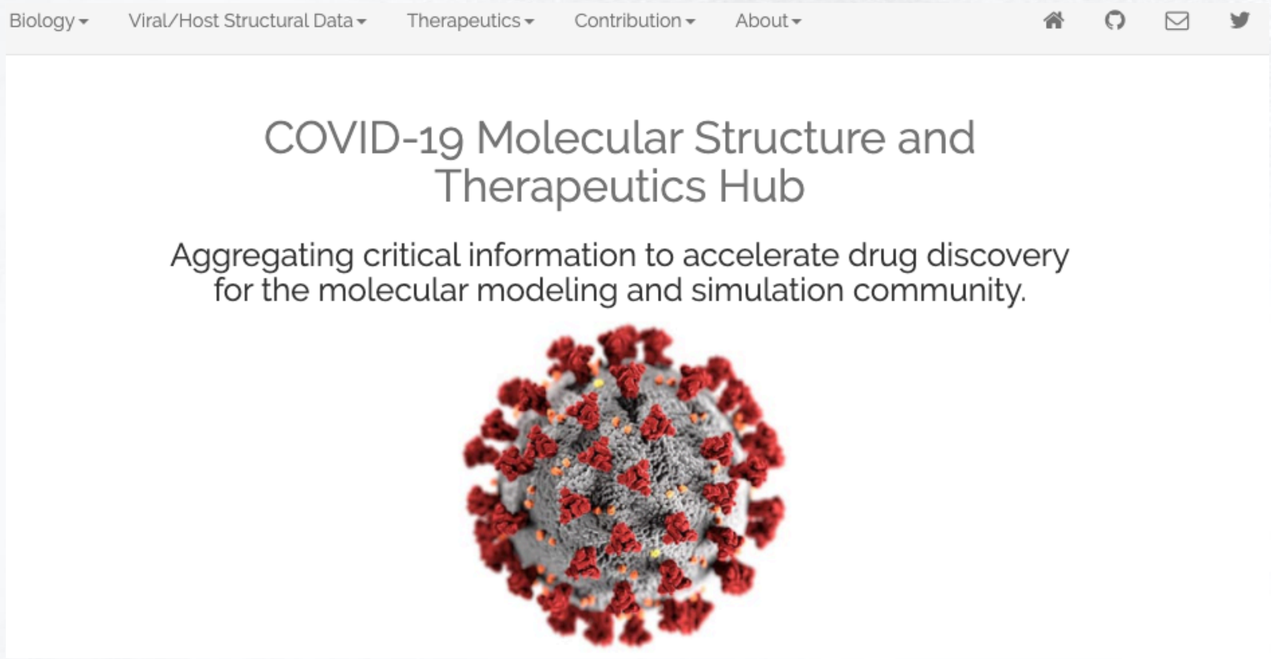
- Simple descriptors:
  - Positive response, effective.
  - Iteration and Expansion



The Hub showed unexpected social insights

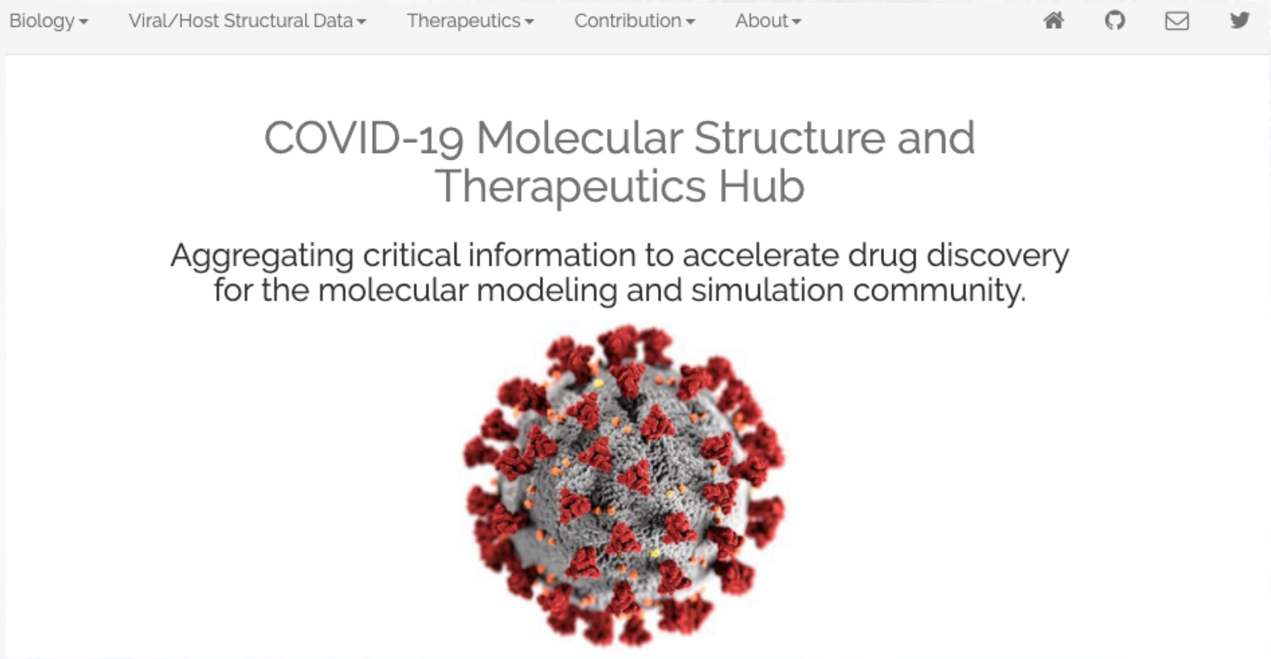
- Simple descriptors:
  - Positive response, effective.
  - Iteration and Expansion

- Community peer reviewed
  - Original: Data ASAP - Review after
  - Actual: Reluctant* data additions

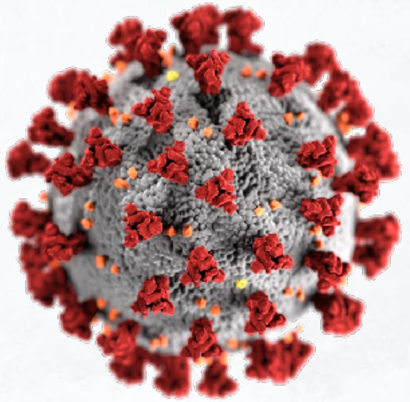


The Hub showed unexpected social insights

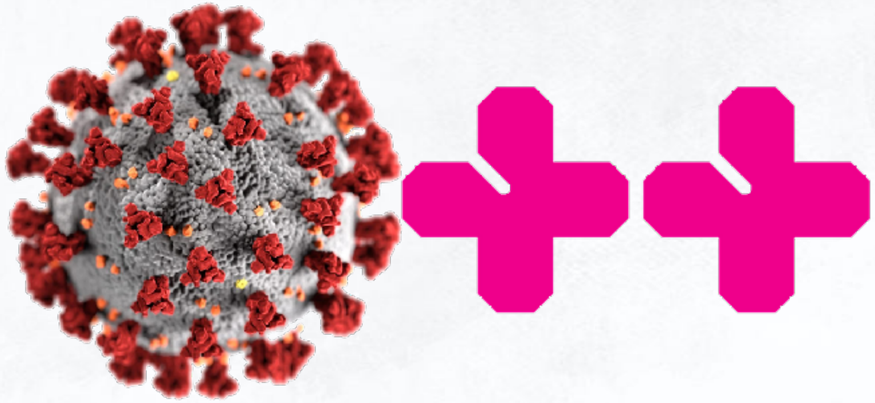
- Simple descriptors:
  - Positive response, effective.
  - Iteration and Expansion
- Community peer reviewed
  - Original: Data ASAP - Review after
  - Actual: Reluctant* data additions
- Centralized or pointers to data
  - Can we link to it? We'll share it.
  - Can we quantify access? Even better.



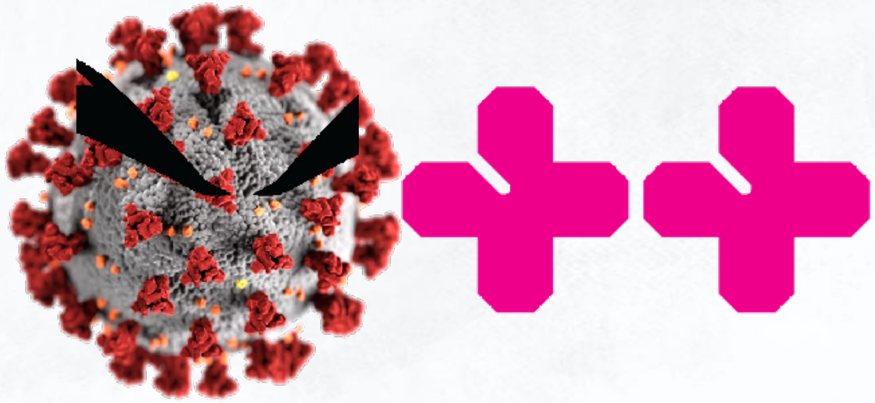
What can we learn for the next research cohesion?



What can we learn for the next research cohesion?

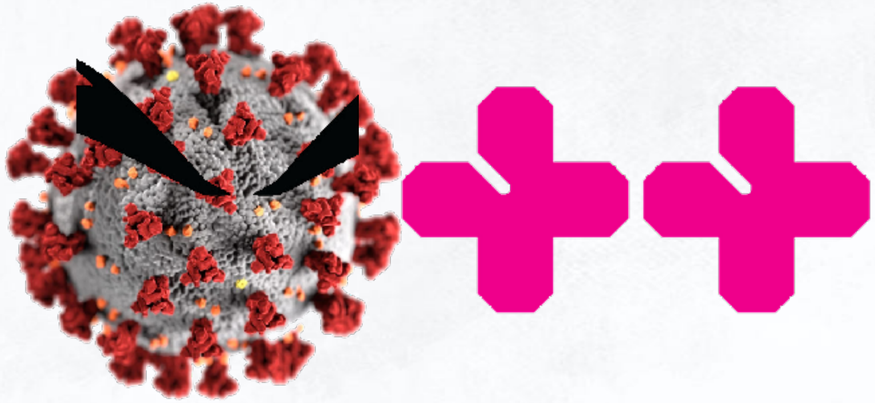


What can we learn for the next research cohesion?



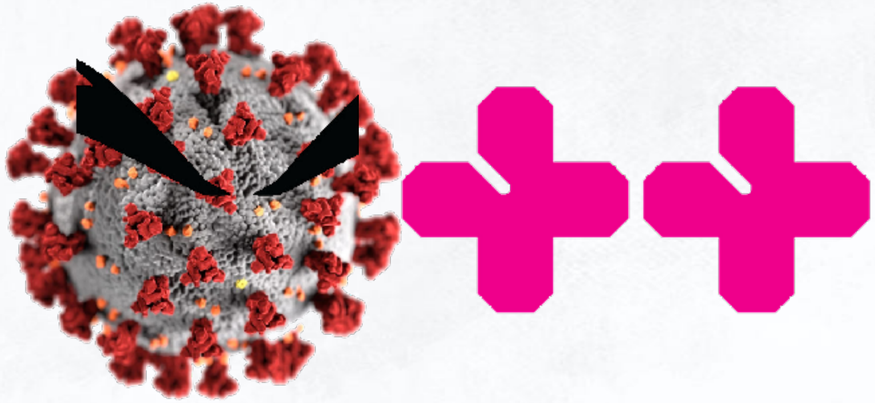
What can we learn for the next research cohesion?

- Encourage easy contribution



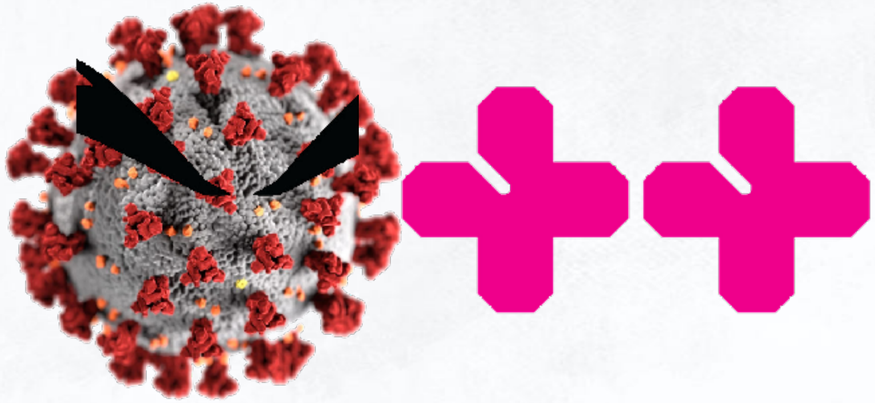
What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations



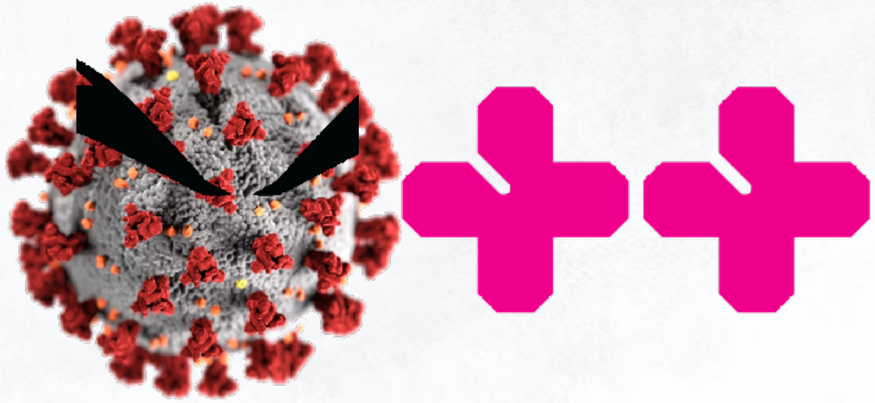
What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations
- Highlight and promote research gaps



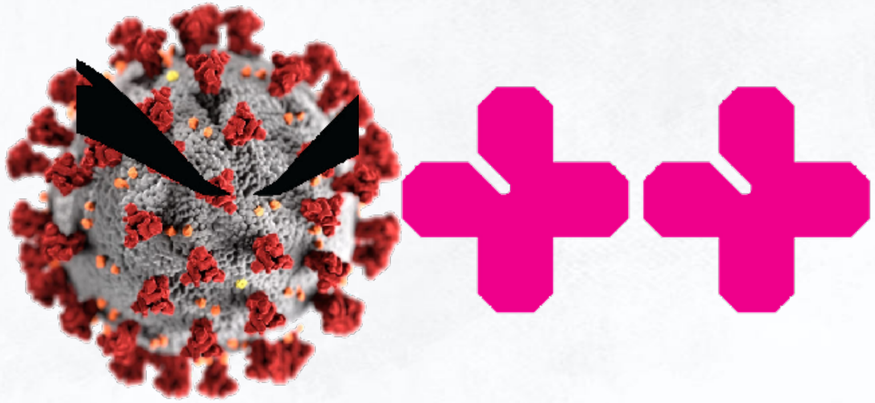
What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations
- Highlight and promote research gaps
- Network data then connect people



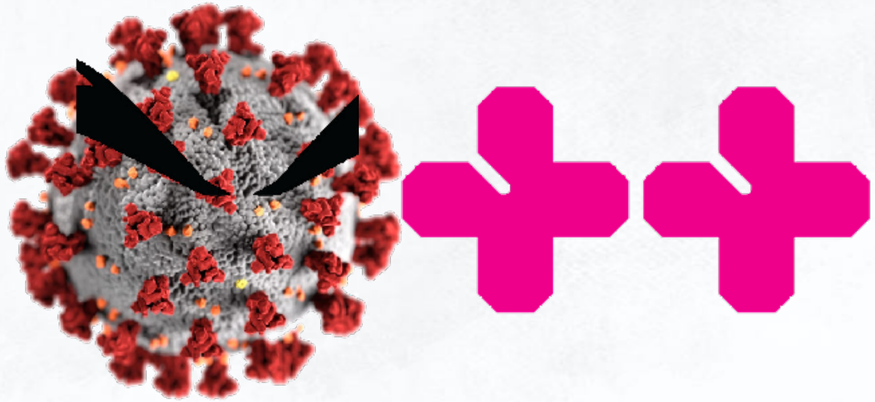
What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations
- Highlight and promote research gaps
- Network data then connect people
- Replace review teams with outreach teams



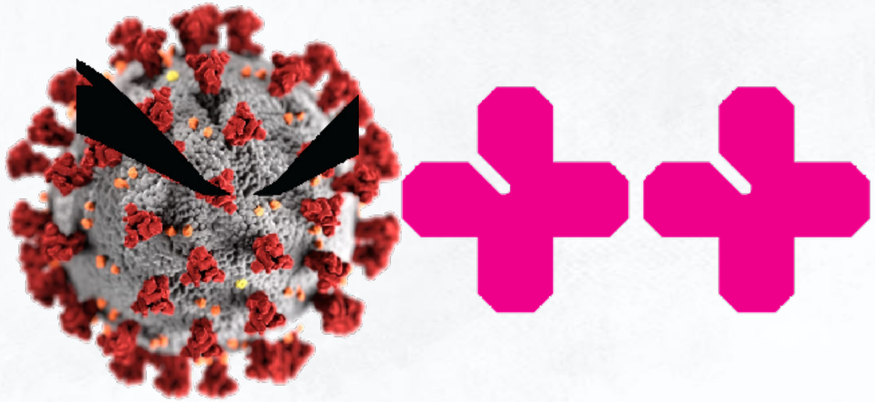
What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations
- Highlight and promote research gaps
- Network data then connect people
- Replace review teams with outreach teams
- Secure large storage early with download metrics



What can we learn for the next research cohesion?

- Encourage easy contribution
- Keep metadata simple with small iterations
- Highlight and promote research gaps
- Network data then connect people
- Replace review teams with outreach teams
- Secure large storage early with download metrics
- Bring in UX expertise



Sam Ellis

Andrew Abi-Mansour

Additional Thanks

Levi Naden	Andrew Abi-Mansour	Sam Ellis
Ben Pritchard	Dosa Allarawy	Jessica Nash
Daniel Crawford	Teresa Head-Gordon	Cecilia Clementi
Rommie Amaro	Adrian Mulholland	John Chodera
Erica Goldberger	Rafal Wiewiora	Alex Payne
Binisha Karki	Hanna Bruce Macdonald	Chiara Mancinelli
Melissa Boby	Jaime Rodriguez-Guerra	Andrea Thorn
Yunyun Gao	Michael Feig	Lim Heo
Richard Friesner	Peter Tieleman	Syma Khalid
Shantanu Jha	Rossen Apostolov	Bert de Groot
Erik Lindahl	Dmitry Morozov	Gerrit Groenhof
Modesto Orozco	Thomas M. Hermans	Stacey Wetmore
Peter Elmer	Frank Wuerthwein	Tim Smith
Lukas Heinrich	Markus Elsing	Alessandra Forti
Mario Lassnig	Tibor Simko	Greg Bowman
Jean-Philip Piquemal	Callie Rodgers Chappell	Thomas Splettstößer
Charlie Brooks	William Jorgensen	D. E. Shaw Research
Tristan Croll	Vincent Voelz	Maxoto Taiji
Yuji Sugita	Tsuhisa Komatsu	Noriaki Okimoto
Stan Soiland-Reyes	Alexandre Bonvin	Yang Zhang
Chengxin Zhang	Voth Group	Schwede Group
David Huggins	Kurzman Group	Daniel McKay

D E Shaw Research

ACI-1547580

Thank you for listening!

@molssi_nsf

@bioexcel_coe

EU Horizon 2020 #823830, #675728