Open Data and its enemies

Dries van Oosten

Debye Institute for NanoMaterials Science

Center for Extreme Matter and Emergent Phenomena

/hy?

Why not?



L

Content

Open data

Why?

Why not?

How?

Open data

/hy?

Vhy not?



Universiteit Utrecht

Open data

vvny?

Why no

How?

- ► Findable
- Accessible
- Interoperable
- ► Reusable



Open data

Why?

Why not

How?



Open data

Why?

Why not?

▶ Funding agencies (will) demand it



Open data

Why?

Why not

- ▶ Funding agencies (will) demand it
- ▶ Less room for monkey business

Universiteit Utrecht

Open data

Why?

Why not

- ▶ Funding agencies (will) demand it
- ► Less room for monkey business
- Contributions can be made by groups that cannot obtain the data



Why not

Why do we fear opening our data?

Open data

Why?

Why not?

How?



Why not

Open data

Why?

Why not?

How?

Why do we fear opening our data?

▶ What if someone does better with our data?



Why not

Open data

Why?

Why not?

Why do we fear opening our data?

- ▶ What if someone does better with our data?
- What if someone does something horrible with our data?

Universiteit Utrecht

Why not

Open data

Why?

Why not?

How?

Why do we fear opening our data?

- ▶ What if someone does better with our data?
- ▶ What if someone does something horrible with our data?
- Preparing data to be FAIR is a lot of work

Why not

Open data

Why?

Why not?

Why do we fear opening our data?

- ▶ What if someone does better with our data?
- ▶ What if someone does something horrible with our data?
- Preparing data to be FAIR is a lot of work
- What does interoperability even mean?

Universiteit Utrecht

Why not

Open data

Why?

Why not?

What if someone does better with our data?



Why not

open data

Why?

Why not?

How?

What if someone does better with our data?

► Good for science, right?



Why not

What if someone does better with our data?

- ► Good for science, right?
- ▶ Bad for us, no glossy no grant!

Open data

Why?

Why not?

low?



Why not

\\/\...2

Why not?

What if someone does better with our data?

- ► Good for science, right?
- ▶ Bad for us, no glossy no grant!

This is why publishing of data should count!

Universiteit Utrecht

Why not

Open data

Why?

Why not?

low?

What if someone does something horrible with our data?



Why not

What if someone does something horrible with our data?

► Monkey business: people (sadly) photoshop data

Open data

Why?

Why not?

ow?



Why not

What if someone does something horrible with our data?

- Monkey business: people (sadly) photoshop data
- ▶ What if they do this with our data?

Open data

/Vhy?

Why not?

low?



Why not

What if someone does something horrible with our data?

- Monkey business: people (sadly) photoshop data
- What if they do this with our data?
- ▶ Will this reflect badly on us?

Open data

Why?

Why not?

low?



Why not

What if someone does something horrible with our data?

- Monkey business: people (sadly) photoshop data
- What if they do this with our data?
- Will this reflect badly on us?

We should be able to show this was their doing!

Open data

vvny?

Why not?

ow?



Why not

Preparing data to be FAIR is a lot of work

Open data

Why?

Why not?

low?



Why not

Preparing data to be FAIR is a lot of work

▶ Our data storage system might look like this

Open data

Why?

Why not?

ow?



Why not

Preparing data to be FAIR is a lot of work

Our data storage system might look like this





Open data

Why not?

low?



Why not

Preparing data to be FAIR is a lot of work

▶ Our data storage system might look like this





▶ The metadata are the handwritten notes of our students

Open data Why? Why not?



Why not

Preparing data to be FAIR is a lot of work

▶ Our data storage system might look like this





▶ The metadata are the handwritten notes of our students

Or in an electronic labbook that is abandonware

Open data Why? Why not?



Why not

Preparing data to be FAIR is a lot of work

Our data storage system might look like this





- ▶ The metadata are the handwritten notes of our students
- Or in an electronic labbook that is abandonware

All true, but we already know this is bad!

Open data Why? Why not?



Why not

Interoperability, what does it even mean?

Open data

Why?

Why not?

low?



Why not

Interoperability, what does it even mean?

"Ready to be combined with other datasets by humans as well as computer systems" Open data Why?

Why not? How?



Why not

Interoperability, what does it even mean?

- "Ready to be combined with other datasets by humans as well as computer systems"
- No vendor lock-in!

Open data Why? Why not?

ow?



Why not

Interoperability, what does it even mean?

- "Ready to be combined with other datasets by humans as well as computer systems"
- ► No vendor lock-in!
- No old pc in the corner to open old files



Open data
Why?
Why not?

J.....2



Why not

Interoperability, what does it even mean?

- "Ready to be combined with other datasets by humans as well as computer systems"
- No vendor lock-in!
- No old pc in the corner to open old files



Open data Why? Why not?

low?

Again, we already know we hate that old pc!



Why not

Interoperability, what does that mean in 10 years?

Open data

Why?

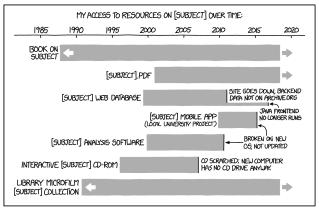
Why not?

low?



Why not

Interoperability, what does that mean in 10 years?



IT'S UNSETTLING TO REALIZE HOW QUICKLY DIGITAL RESOURCES CAN DISAPPEAR WITHOUT ONGOING WORK TO MAINTAIN THEM.

Open data
Why?
Why not?



Open data How?

Open data

Why?

Why not

How?



How?

► Easy to use (so we use it)

Open data

Why?

Why not?

How?



How?

Easy to use (so we use it)

▶ In our own influence sphere (so it will continue to exist)

Open data

Why?

Why not?



How?

- Easy to use (so we use it)
- ▶ In our own influence sphere (so it will continue to exist)
- Using open standards (preferably an existing one)

Why?

Why not i



How?

- Easy to use (so we use it)
- In our own influence sphere (so it will continue to exist)
- Using open standards (preferably an existing one)

HOW STANDARDS PROUFERATE:

(SEE AIC OMPGERS, CHIPPORTER ENCOUNCE, INSTITUT MESSANING, ETC)

IH?! RIDICULOUS!

WE NEED TO DEVELOP

ONE UNIVERSAL STANDARD

THAT COVERS EVERYONES

USE CASES. YEAH!

STANDARDS.

STANDARDS.

Open dat Why? Why pot?



How?

▶ Easy to use (so we use it)

In our own influence sphere (so it will continue to exist)

Using open standards (preferably an existing one)

HOW STANDARDS PROLIFERATE:

(SEE A/C CHARGERS, GHÁROCHE BIOLODAG, INSTINIT MESAGANG, EIC)

H-Y?: RIDICULOUS!

WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONES
USE CASES. YEAH!

THERE ARE
IN COMPETING
STANDARDS.

With data protection, auditing, etc.

Open data Why? Why not?



Our workflow

Open data

Why?

Why not



Our workflow

▶ Data publication right before submission of manuscript

Open data

Why?

Why not?



Our workflow

Open data

Why?

Why not

- Data publication right before submission of manuscript
- ► Contains all data and all Python scripts to generate graphs



Our workflow

Open data

Why?

Why not'

- Data publication right before submission of manuscript
- ► Contains all data and all Python scripts to generate graphs
- Python scripts are used as roadmap



Our workflow

Open data

Why?

Why not

- Data publication right before submission of manuscript
- ► Contains all data and all Python scripts to generate graphs
- Python scripts are used as roadmap
- Scripts require no user input



Our workflow

Open data

Why?

Why not

- Data publication right before submission of manuscript
- Contains all data and all Python scripts to generate graphs
- Python scripts are used as roadmap
- Scripts require no user input
- All scripts and data under Creative Commons license



Our workflow

- Data publication right before submission of manuscript
- Contains all data and all Python scripts to generate graphs
- Python scripts are used as roadmap
- Scripts require no user input
- All scripts and data under Creative Commons license
- No proprietary software used



Our workflow

Open data

Why?

Why not

How?

- Data publication right before submission of manuscript
- Contains all data and all Python scripts to generate graphs
- Python scripts are used as roadmap
- Scripts require no user input
- All scripts and data under Creative Commons license
- No proprietary software used

I ask most of this of my students anyway!



Some nerding out

Open data

Why?

Why not?



Open data

Why?

Why not?

How?

Some nerding out

 Use iRODS (integrated Rule-Oriented Data management System)



Open data

Why?

Why not?

How?

- Use iRODS (integrated Rule-Oriented Data management System)
 - data grid software



Open data

Why?

Why not

How?

- Use iRODS (integrated Rule-Oriented Data management System)
 - data grid software
 - order and use data spread across multiple locations



Open data

Why?

Why not?

How?

- Use iRODS (integrated Rule-Oriented Data management System)
 - data grid software
 - order and use data spread across multiple locations
 - · with their metadata



Open data

Why?

Why not?

How?

- Use iRODS (integrated Rule-Oriented Data management System)
 - data grid software
 - order and use data spread across multiple locations
 - · with their metadata
- Rules to determine who can do what etc.



Open data

Why?

Why not?

How?

- Use iRODS (integrated Rule-Oriented Data management System)
 - data grid software
 - order and use data spread across multiple locations
 - · with their metadata
- Rules to determine who can do what etc.
- Interfaces that hide that complexity



portal
*.yoda.uu.nl

network-disk
*.data.uu.nl

administration
i-commands

iRODS API

UU Data Policies and -services

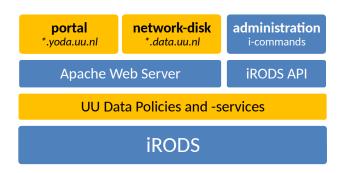
iRODS

Open data

Why?

Why not





▶ Blue: open source projects

Why?
Why not



portal
*.yoda.uu.nl

network-disk
*.data.uu.nl

Apache Web Server

iRODS API

UU Data Policies and -services

iRODS

Open data

vvny

Why not

How?

▶ Blue: open source projects

Yellow: UU developed (also open sourced)

Universiteit Utrecht

Our YODA workflow

Open data

Why?

Why not?



Our YODA workflow

Store data (looks like a shared volume)

Open data

Why?

Why not?



Our YODA workflow

Open data

Why?

Why not?

- Store data (looks like a shared volume)
- ► Fill out metadata (through command line or web portal)



Our YODA workflow

Open data

Why?

Why not?

- Store data (looks like a shared volume)
- ► Fill out metadata (through command line or web portal)
- ▶ Lock data



Our YODA workflow

Орен цата

Why?

Why not?

- Store data (looks like a shared volume)
- ► Fill out metadata (through command line or web portal)
- Lock data
- Submit to vault (approved by groupleader)



Our YODA workflow

Open data

Why?

Why not?

- Store data (looks like a shared volume)
- ▶ Fill out metadata (through command line or web portal)
- ▶ Lock data
- Submit to vault (approved by groupleader)
- Submit Publish (approved by data manager)



Our YODA workflow

Open data

Why?

Why not?

- Store data (looks like a shared volume)
- Fill out metadata (through command line or web portal)
- Lock data
- Submit to vault (approved by groupleader)
- Submit Publish (approved by data manager)
- DOI assigned, leads to landing page with metadata



Our YODA workflow

A/I 2

vviiy:

Why not?

- Store data (looks like a shared volume)
- Fill out metadata (through command line or web portal)
- Lock data
- Submit to vault (approved by groupleader)
- Submit Publish (approved by data manager)
- DOI assigned, leads to landing page with metadata
- And data itself, if you wish



Our YODA workflow

Open data

Why?

Why not?

- Store data (looks like a shared volume)
- Fill out metadata (through command line or web portal)
- ▶ Lock data
- Submit to vault (approved by groupleader)
- Submit Publish (approved by data manager)
- ▶ DOI assigned, leads to landing page with metadata
- And data itself, if you wish
- Cite data in your manuscript



https://www.doi.org/10.24416/UU01-MZ5ONP

Open data

Why?

Why not?



https://www.doi.org/10.24416/UU01-MZ5ONE

► You can download that data

Open data

Why?

Why not?



https://www.doi.org/10.24416/UU01-MZ5ONP

- ► You can download that data
- ▶ Look for the figure that caught your eye

Open data

vvny?

Why not?



► https://www.doi.org/10.24416/UU01-MZ5ONF

- You can download that data
- ▶ Look for the figure that caught your eye
- ▶ Find the accompanying Python script

Open data

vvny?

Why not



► https://www.doi.org/10.24416/UU01-MZ5ONF

- You can download that data
- ▶ Look for the figure that caught your eye
- Find the accompanying Python script
- And run it (or modify it first)

Open data

vvriy:

Why not



► https://www.doi.org/10.24416/UU01-MZ5ONF

- You can download that data
- Look for the figure that caught your eye
- Find the accompanying Python script
- And run it (or modify it first)
- Use it in your work (just cite the data package)

Open data

vvny!

Why not



Open data

Why?

Why not



► Under our own control, immediate support (ITS-UU)

Open data

Why?

Why not?



Open data

Why?

Why not?

How?

▶ Under our own control, immediate support (ITS-UU)

Low cost (2 €/ TB/ Month)



Open data

Why?

Why not?

- ▶ Under our own control, immediate support (ITS-UU)
- Low cost (2 €/ TB/ Month)
- Redundant (across at least two locations)



Open data

Why?

Why not?

- Under our own control, immediate support (ITS-UU)
- Low cost (2 €/ TB/ Month)
- Redundant (across at least two locations)
- ▶ So easy to use there is no reason not to!



Open data

Why?

Why not?

How?

- Under our own control, immediate support (ITS-UU)
- Low cost (2 €/ TB/ Month)
- Redundant (across at least two locations)
- So easy to use there is no reason not to!

Does this solve everything?



What are the "enemies" of open data?

Open data

Why?

Why not



What are the "enemies" of open data?

Our own lack of discipline

Open data

Why?

Why not?



What are the "enemies" of open data?

Our own lack of discipline

Assessment policies, these have to be modernized

Why?

Why not?



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - See related discussion on Open Access publishing

Open data

Why?

Why not



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - · See related discussion on Open Access publishing
- Propriety software/formats

Open data

Why?

Why not



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - · See related discussion on Open Access publishing
- Propriety software/formats
 - These are by definition not interoperable

Open data

vvny?

vvny not?



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - · See related discussion on Open Access publishing
- Propriety software/formats
 - These are by definition not interoperable
 - But some hard- or software might force you

Open data

Why?

Why not!



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - · See related discussion on Open Access publishing
- Propriety software/formats
 - These are by definition not interoperable
 - · But some hard- or software might force you
 - In the long term we should not accept this

Open data

vviiy:

vvny not?



What are the "enemies" of open data?

- Our own lack of discipline
- Assessment policies, these have to be modernized
 - · See related discussion on Open Access publishing
- Propriety software/formats
 - These are by definition not interoperable
 - · But some hard- or software might force you
 - · In the long term we should not accept this

....

Why?

