

Open Data and its enemies

Dries van Oosten

Debye Institute for NanoMaterials Science
Center for Extreme Matter and Emergent Phenomena

Open data

Why?

Why not?

How?



Universiteit Utrecht

Content

Open data

Why?

Why not?

How?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

FAIR

- ▶ Findable
- ▶ Accessible
- ▶ Interoperable
- ▶ Reusable

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why?

- ▶ Funding agencies (will) demand it

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why?

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

Why do we fear opening our data?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

Why do we fear opening our data?

- ▶ What if someone does better with our data?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

What if someone does better with our data?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

What if someone does better with our data?

- ▶ Good for science, right?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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!

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

What if someone does something horrible with our data?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Why not

What if someone does something horrible with our data?

- ▶ Monkey business: people (sadly) photoshop data

Open data

Why?

Why not?

How?



Open data

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?

How?



Universiteit Utrecht

Open data

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

Why?

Why not?

How?



Open data

Why not

Preparing data to be FAIR is a lot of work

Open data

Why?

Why not?

How?



Open data

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?

How?



Open data

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?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

Why not

Interoperability, what does it even mean?

Open data

Why?

Why not?

How?



Open data

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?



Open data

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?

How?



Universiteit Utrecht

Open data

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



Again, we already know we hate that old pc!



Universiteit Utrecht

Open data

Why?

Why not?

How?

Open data

Why not

Interoperability, what does that mean in 10 years?

Open data

Why?

Why not?

How?

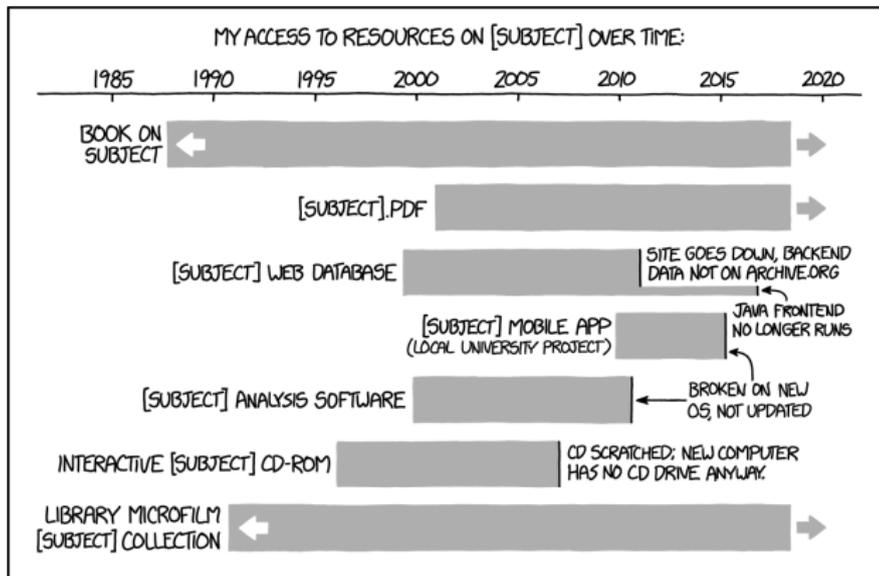


Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

How?

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

How?

- ▶ Easy to use (so we use it)

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?



Universiteit Utrecht

Open data

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)

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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)



- ▶ With data protection, auditing, etc.

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our workflow

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our workflow

- ▶ Data publication right before submission of manuscript

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our workflow

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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

Open data

Why?

Why not?

How?



Open data

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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

Open data

Why?

Why not?

How?



Open data

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

I ask most of this of my students anyway!

Open data

Why?

Why not?

How?



Open data

YODA

Some nerding out

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

Some nerding out

- ▶ Use iRODS (integrated Rule-Oriented Data management System)

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

Some nerding out

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

Some nerding out

- ▶ 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?



Universiteit Utrecht

Open data

YODA

Some nerding out

- ▶ 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?



Universiteit Utrecht

Open data

YODA

Some nerding out

- ▶ 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?



Universiteit Utrecht

Open data

YODA

Some nerding out

- ▶ 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

Open data

Why?

Why not?

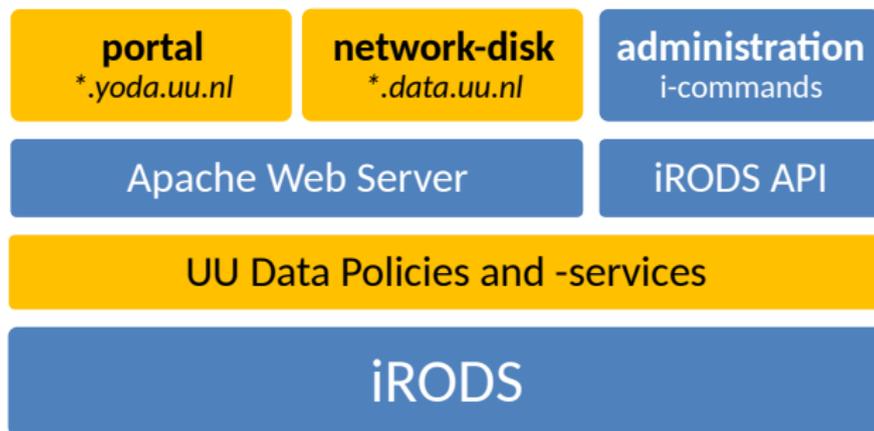
How?



Universiteit Utrecht

Open data

YODA



Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our YODA workflow

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our YODA workflow

- ▶ Store data (looks like a shared volume)

Open data

Why?

Why not?

How?



Open data

Our YODA workflow

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

Open data

Why?

Why not?

How?



Open data

Our YODA workflow

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

Our YODA workflow

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

Open data

Why?

Why not?

How?



Open data

Our YODA workflow

- ▶ 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)

Open data

Why?

Why not?

How?



Open data

Our YODA workflow

- ▶ 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

Open data

Why?

Why not?

How?



Open data

Our YODA workflow

- ▶ 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

Open data

Why?

Why not?

How?



Open data

YODA

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

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

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

- ▶ Under our own control, immediate support (ITS-UU)
- ▶ Low cost (2 €/ TB/ Month)

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

YODA

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

Open data

Why?

Why not?

How?



Open data

YODA

- ▶ 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?



Open data

YODA

- ▶ 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?

Open data

Why?

Why not?

How?



Open data

What are the "enemies" of open data?

Open data

Why?

Why not?

How?



Open data

What are the "enemies" of open data?

- ▶ Our own lack of discipline

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

What are the "enemies" of open data?

- ▶ Our own lack of discipline
- ▶ Assessment policies, these have to be modernized

Open data

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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

Why?

Why not?

How?



Universiteit Utrecht

Open data

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?

How?



Universiteit Utrecht

Open data

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

Why?

Why not?

How?



Universiteit Utrecht

Open data

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

Why?

Why not?

How?



Universiteit Utrecht