A Beginner's Guide to Reasoning: How to Reason Your Way to Better Data

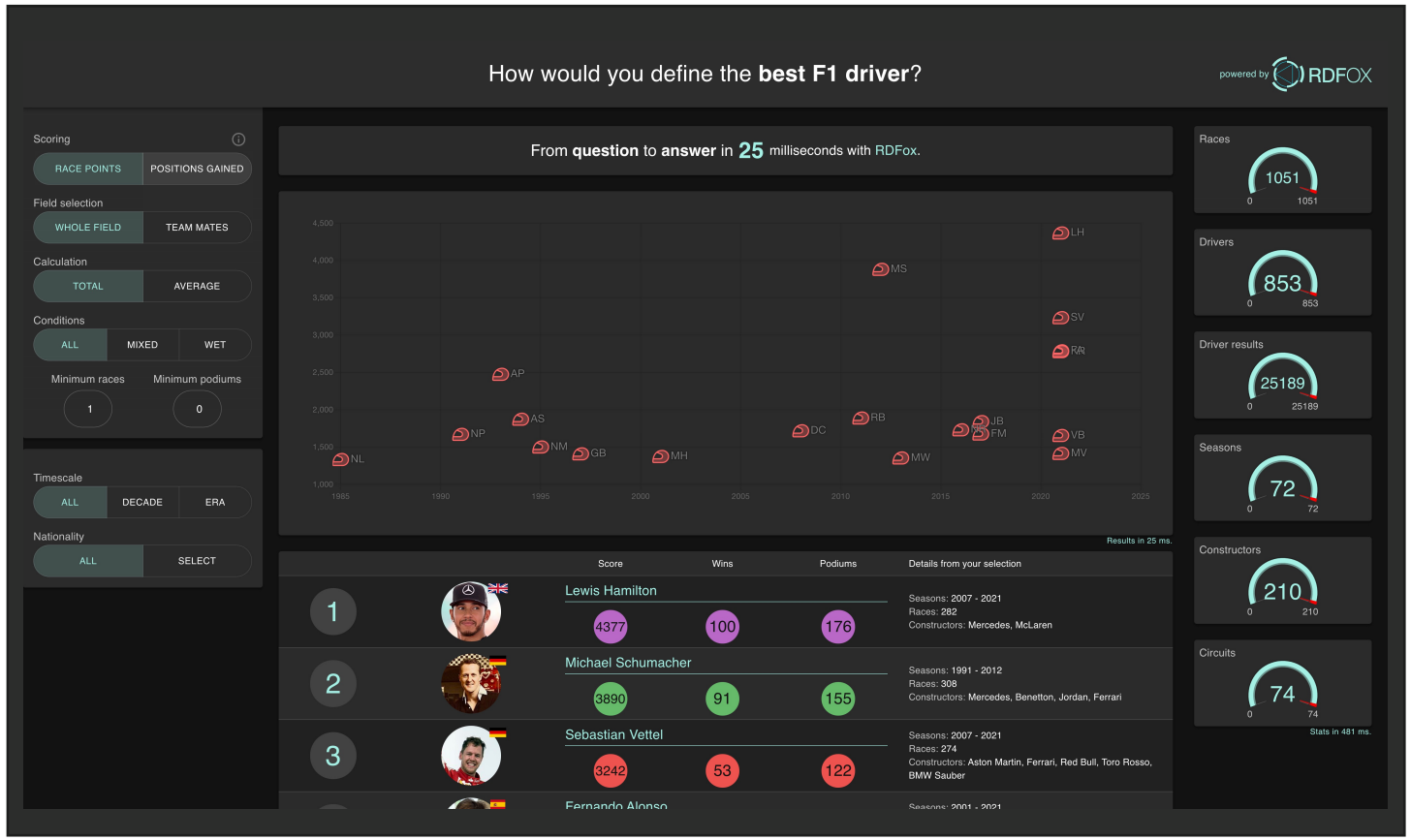
You'll need:

- RDFox Downloaded
  https://www.oxfordsemantic.tech/downloads
- An RDFox License
  https://oxfordsemantictech.typeform.com/to/opwK7fBn
- The Tutorial Files
  https://github.com/OxfordSemantic/RDFoxWorkshop-KGC2022

All links in the chat

+ your favorite IDE!

Who is the Greatest Formula One Driver of All Time?



Controls and filters for the scoring system.

Statistics about the data used to form the results.

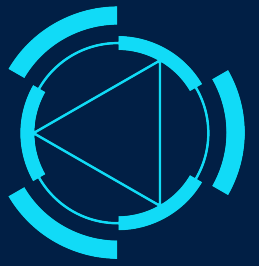
This matches the filters.

http://f1.rdfox.tech

Try it for yourself!

© Oxford Semantic Technologies 2022

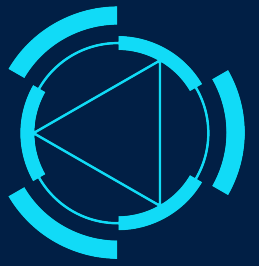
Objectives



By the end of the class everyone will:

- know how to write OWL axioms in Turtle syntax
- know how to add OWL axioms to RDFox datastores
- know how to write and import Datalog rules
- understand how RDFox performs reasoning
- know common uses of rules

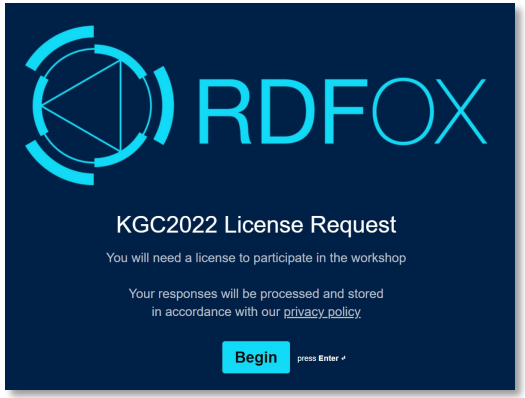
Workshop Structure



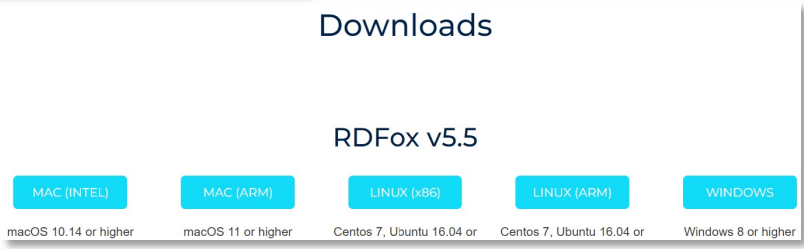
- The repository contains an RDFox script file “start.rdfox” with all the commands you need - copy them one by one as we go
- Each axiom/rule has an associated query (mentioned on the slide). Run it before and after reasoning to see the difference!
- Some examples are exercises and you will need to fill in the gaps
- Answer files have “completed” in their names

Requirements

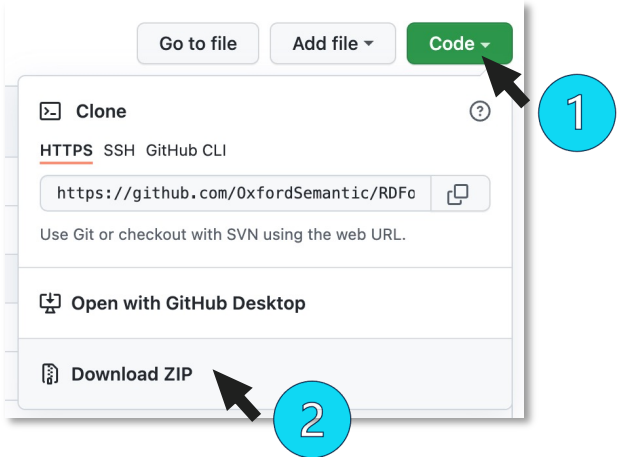
A. Get an RDFox License
https://oxfordsemantictech.typeform.com/to/opwK7fBn



B. Download RDFox (& unzip)
https://www.oxfordsemantic.tech/downloads

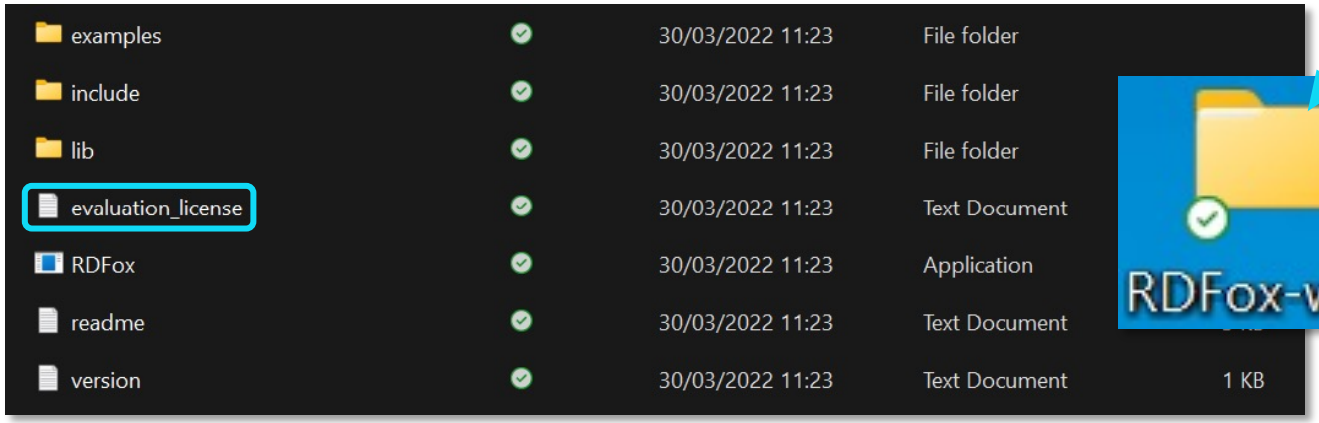


C. Download the class materials from Github:
https://github.com/OxfordSemantic/RDFoxWorkshop-KGC2022



D. Open your IDE of choice (VS Code etc.)
https://code.visualstudio.com/

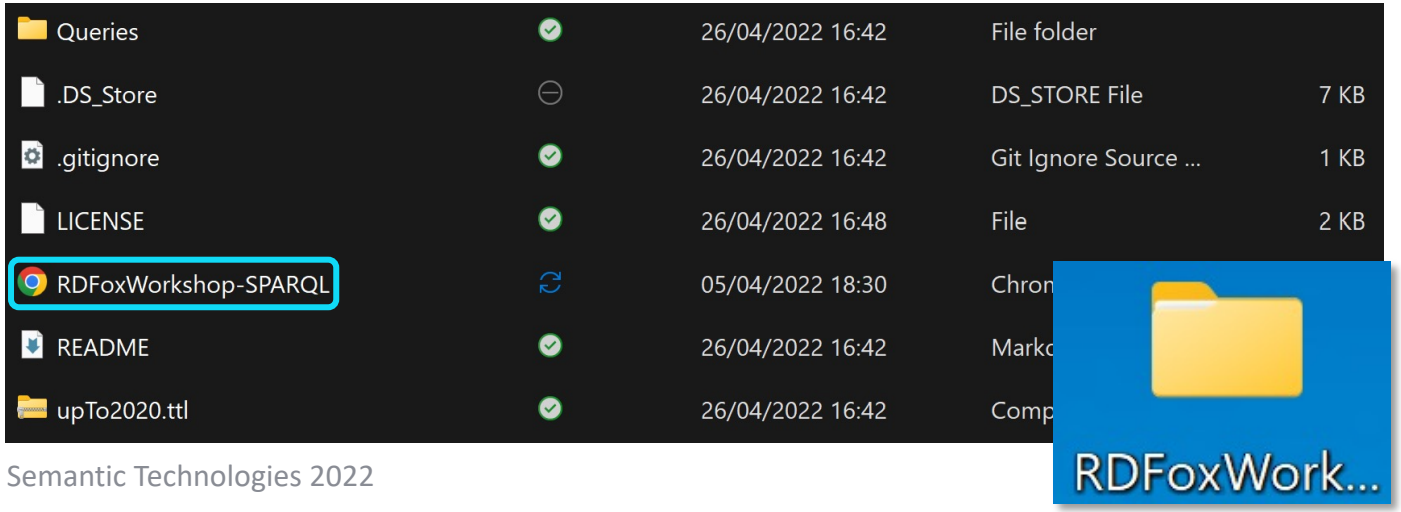
Requirements



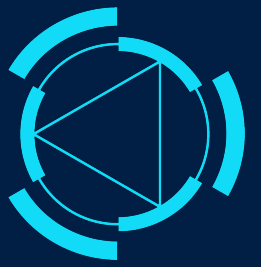
Make sure to put your license **INSIDE** the **RDFox folder**.

Open the class PDF so you can follow along!

You will need to follow the links provided.



Starting RDFox



- Open a terminal in the IDE
- In the terminal, run the following

  Mac: `<Path/to>/RDFox sandbox`

  Windows: `<Path/to>/RDFox.exe sandbox`
- The RDFox server is now running…

If your file path contains spaces put “quotation marks” around it.

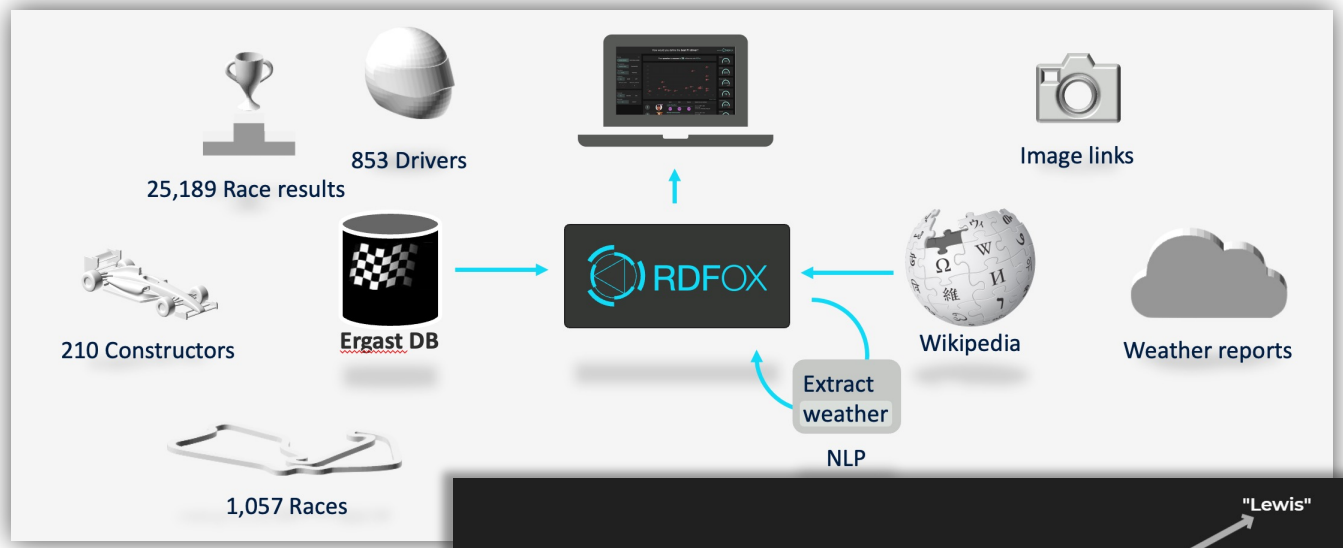
You may need to use back slashes \ or double up the slashes // depending on your setup.

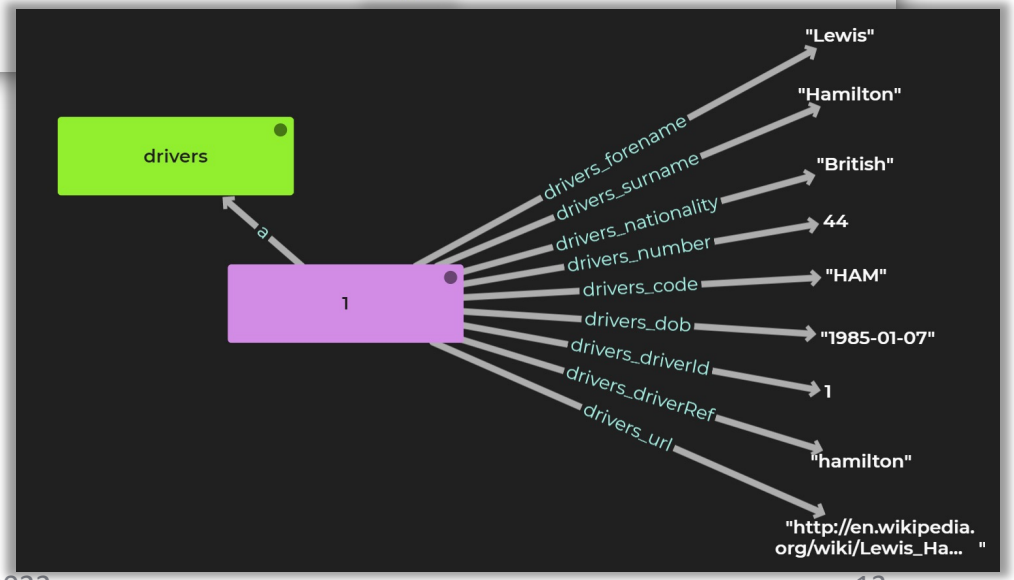
```
Source code for RDFox v1.0 Copyright 2013 Oxford University Innovation Limited and subsequent improvements Copyright 2017-2021 by Oxford Semantic Techn
ologies Limited.
This copy of RDFox is licensed for Developer use to Tom Vout (tom.vout@oxfordsemantic.tech) of OST until 07-Jun-2022 16:01:44
This system is equipped with 16.9 GB of RAM, and RDFox is configured to use at most 15.2 GB (89.9% of the total).
Currently, 2.8 GB (18.4% of the amount allocated to RDFox) appear to be available on the system.
Since RDFox is a RAM-based system, its performance can suffer when other running processes use a lot of memory.


A new server connection was opened as role 'guest' and stored with name 'sc1'.
>
```

https://docs.oxfordsemantic.tech/command-line-reference.html#starting-the-rdfox-process

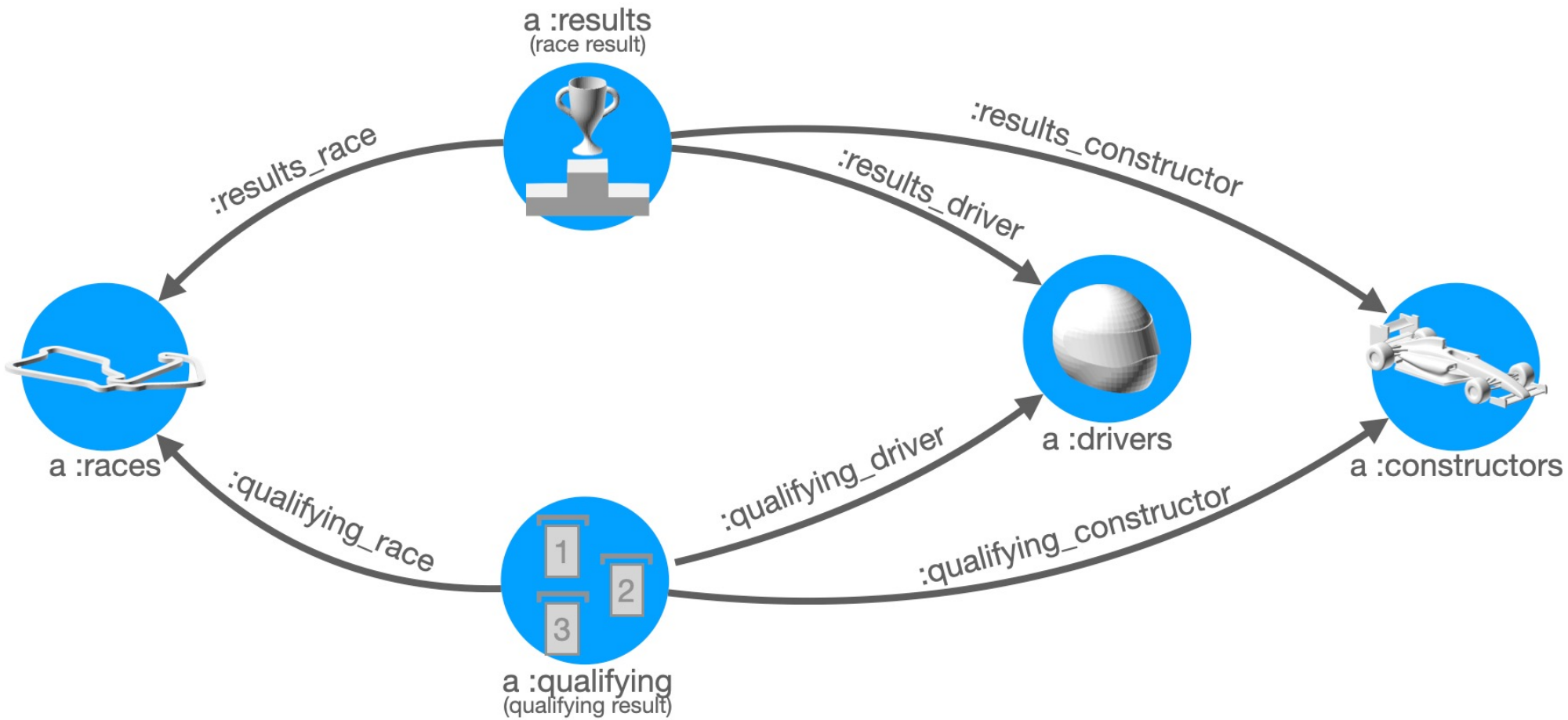
The data for today's class

- Race data from the fan-run Ergast database
  - Drivers
  - Races
  - Race results
  - Constructors

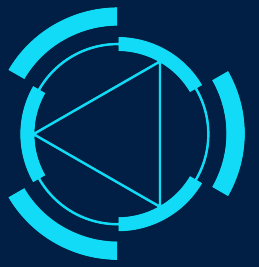




Sample data



Sample data



- we import real F1 race data from file “upTo2020.ttl”
- we also import test data that showcases how OWL axioms work (file “sample.ttl”)

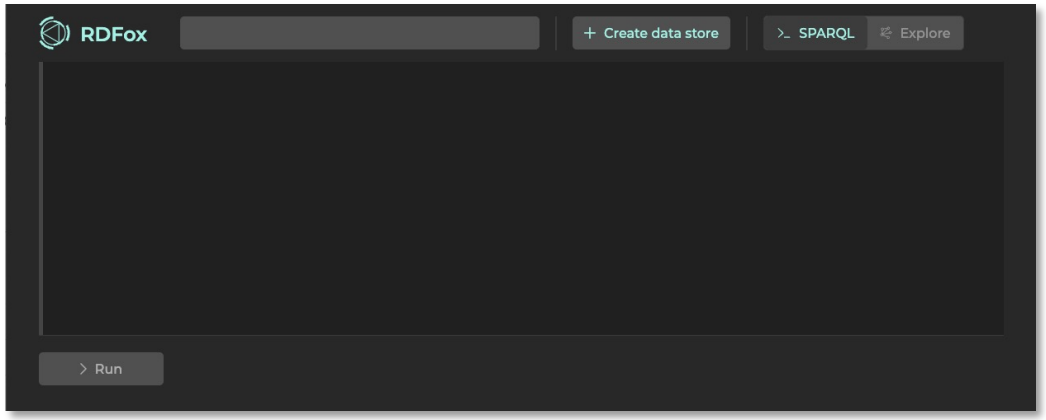
```
:drivers\/123456789 # does not belong to any class
    :drivers_forename "John";
    :drivers_surname "Smith" .

:results\/123456789 a :results;
    :results_driver :drivers\/123456789, :drivers\/1; # two drivers instead of one
    :drivers_forename "John"; # not supposed to be here - this is a driver property
    # missing constructor
    :results_race :races\/123456789;
    :results_positionOrder "1"^^xsd:integer .

:results\/123456789-b a :win .
```

The REST Endpoint and UI

- In the terminal, run:
  `endpoint start`

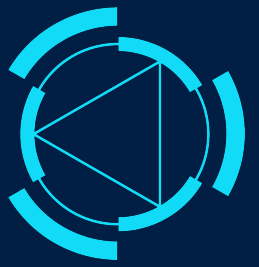
```
> endpoint start

WARNING: The RDFox endpoint is running with no transport layer security (TLS). This could allow attackers to
steal information including role passwords.
         See the endpoint.channel variable and related variables in the description of the RDFox endpoint for
 details of how to set up TLS.

The REST endpoint was successfully started at port number/service name 12110 with 7 threads.
>
```

- The RDFox endpoint is now running, so we can use the web UI
- Open a browser and go to: localhost:12110/console
- This will show an empty console

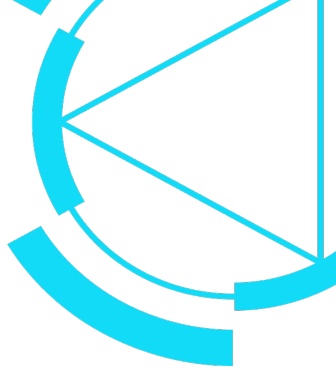


https://docs.oxfordsemantic.tech/command-line-reference.html#endpoint

RDFox Commands

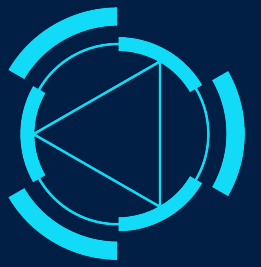


- `endpoint start` - enables UI at localhost:12110/console (default)
- `dstore create [dstore_name]` - creates a data store
- `active [dstore_name]` - sets active datastore
- `set output [out/filename]` - directs output to shell or file
- `prefix [name]: <[iri]>` - saves prefix declaration
- `import [+/-]? [> graph_iri]? [filename]` - import/remove data or rule file
- `importaxioms [source_graph]? [target_graph]? [+/-]?` – parse source graph for axioms to add to target graph
- `evaluate [filename]` - evaluate a SPARQL query from file
- `info [rulestats/axioms/...]?` - print information about data store
- `quit` - exit RDFox



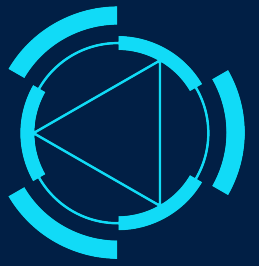
OWL reasoning

Web Ontology Language (OWL)

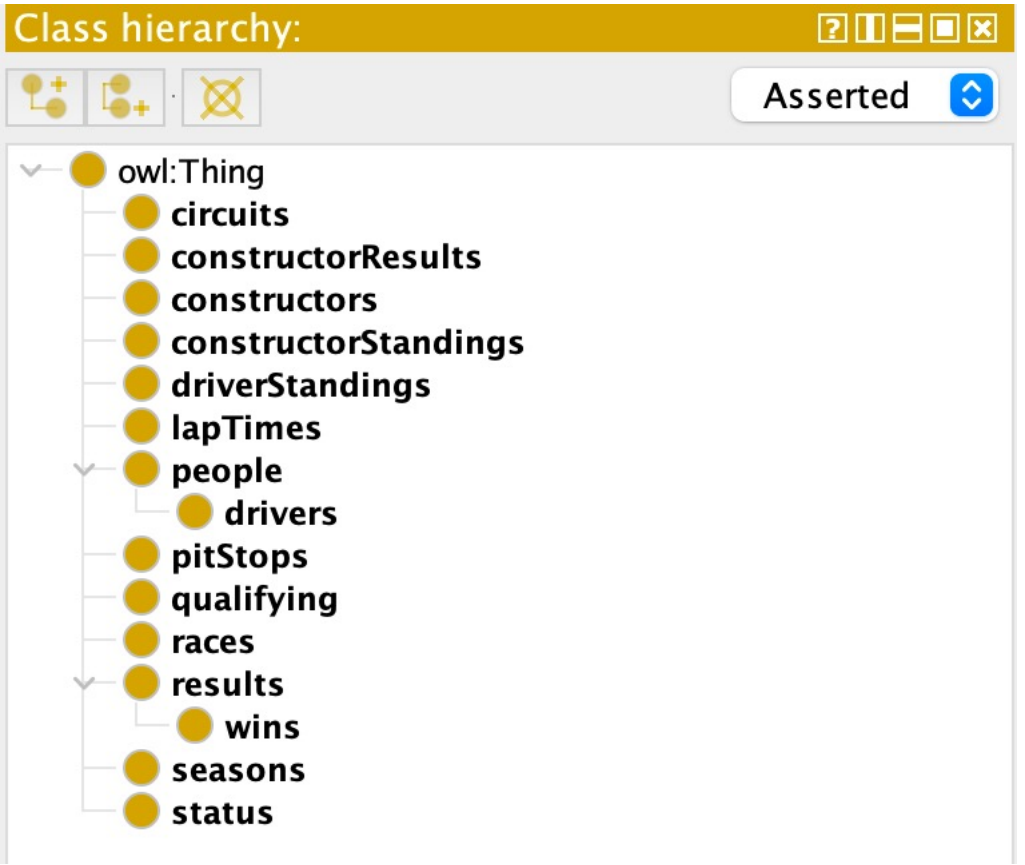


- standard created by W3C, RDFox supports the profile OWL 2 RL
- axioms = assertions about classes, properties or individuals
- ontology = collection of axioms
- ontologies are themselves graphs and so can be written in many different syntaxes - we will use Turtle
- internally, RDFox translates axioms into Datalog rules

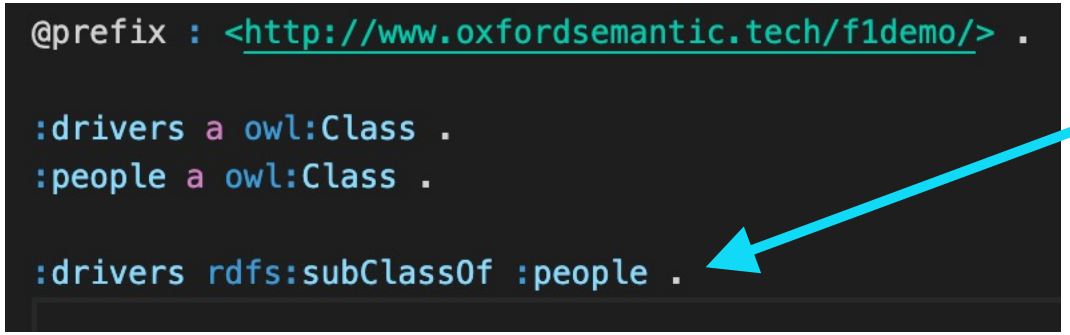
Protégé



- useful program for managing ontologies
- free and open source



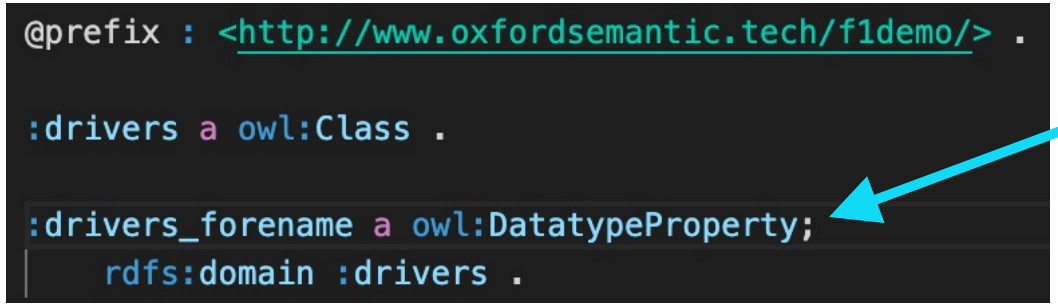
Axiom set 1
Subclasses



this is a completely new class that was not previously present in the data

- declares that `:drivers` and :people are classes
- says that `:drivers` is a subclass of `:people`
- effect: all instances of `:drivers` also become instances of `:people` (check by running query q-a1.rq)

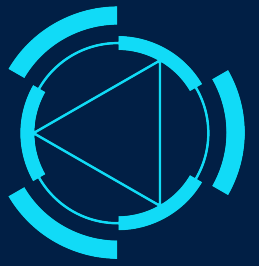
Axiom set 2
Property types and domain/range



OWL defines each property as either a “data” or “object” property

- says the domain of the property `:drivers_forename` is the class `:drivers`
- effect: `:drivers\/123456789` and `:results\/123456789` from sample both become instances of `:drivers` (check by running query q-a2.rq)

Fixing Mistakes



What should I do if I make a mistake?

```
importaxioms : myAxiomsGraph –
```

```
update ! clear graph :myAxiomsGraph
```

```
import > :myAxiomsGraph axioms/a1.ttl …
```

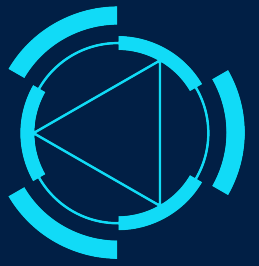
```
importaxioms :myAxiomsGraph
```

Axiom set 3
Disjoint classes

```
@prefix : <http://www.oxfordsemantic.tech/f1demo/> .

:drivers a owl:Class .
:results a owl:Class .

:drivers owl:disjointWith :results .
```

- says that :drivers and `:results` are disjoint classes
- effect: `:results\/123456789` from the sample becomes an instance of `owl:Nothing` (check by running query q-a3.rq) – this is a consequence of axiom set 2

Axiom set 4
Cardinality restrictions



```
@prefix : <http://www.oxfordsemantic.tech/f1demo/> .

:results_driver a owl:ObjectProperty .

:results rdfs:subClassOf [
    a owl:Restriction ;
    owl:onProperty :results_driver ;
    owl:maxCardinality 1
] .
```

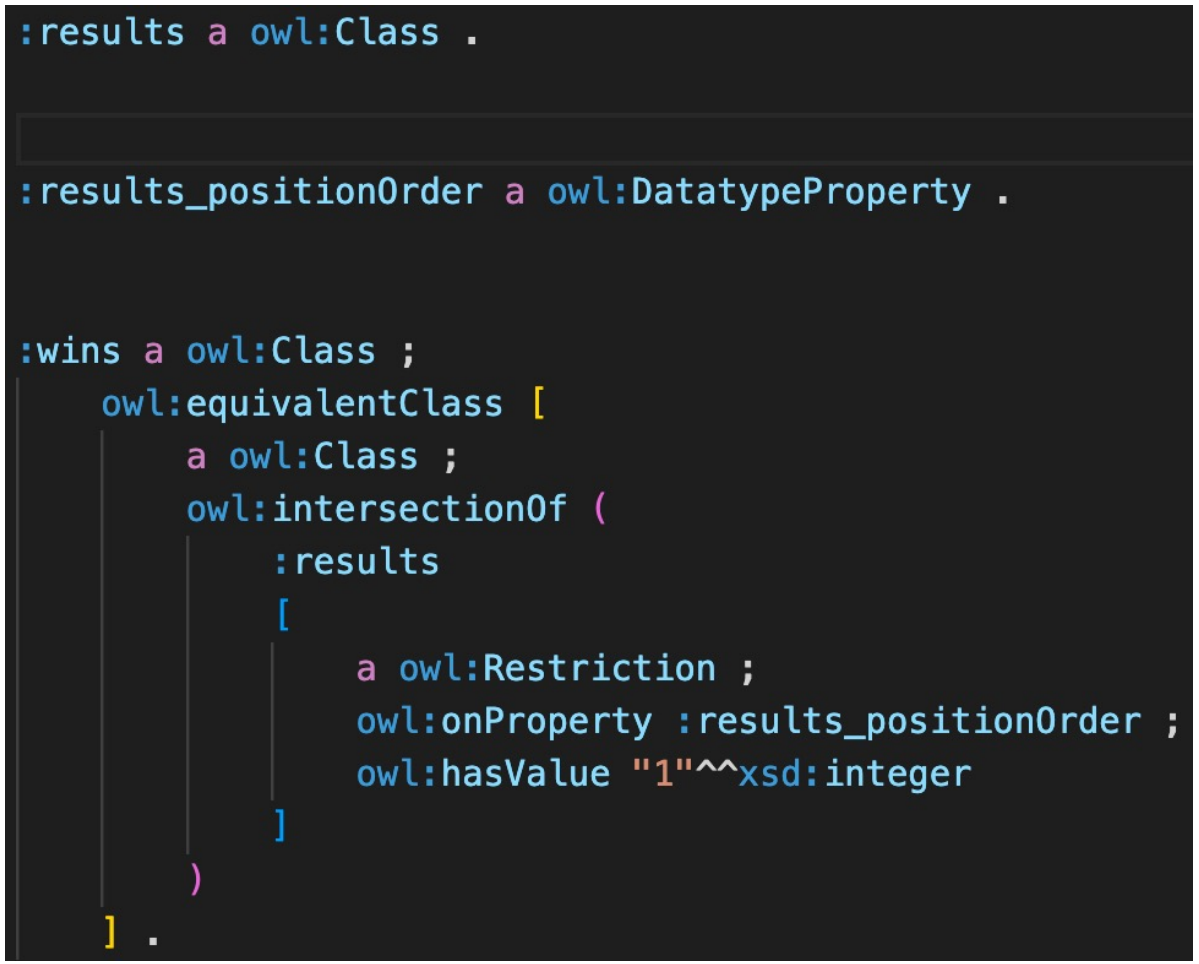
- says that every node of class `:results` has at most one value of `:results_driver`
- effect: property `owl:sameAs` now connects `:driver\/123456789` and `:driver\/1`, as well as every driver who ever raced to themselves (check by running query q-a4.rq)

Axiom set 5
Property chains axioms

```
@prefix : <http://www.oxfordsemantic.tech/f1demo/> .

:results_driver a owl:ObjectProperty .

:hasRacedIn owl:propertyChainAxiom (
    [owl:inverseOf :results_driver]
    :results_race
) .
```

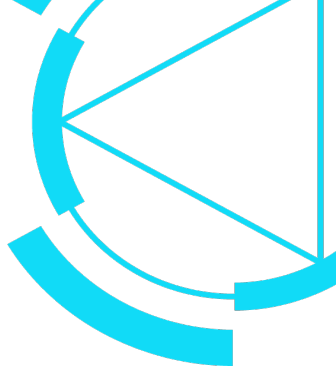
- defines a new property `:hasRacedIn` as a combination of `:results_driver` (inverted) and `:results_race`
- effect: property `:hasRacedIn` now connects drivers and their races (check by running query q-a5.rq)

Axiom set 6
Combining OWL elements



```
:results a owl:Class .


:results_positionOrder a owl:DatatypeProperty .


:wins a owl:Class ;
    owl:equivalentClass [
        a owl:Class ;
        owl:intersectionOf (
            :results
            [
                a owl:Restriction ;
                owl:onProperty :results_positionOrder ;
                owl:hasValue "1"^^xsd:integer
            ]
        )
    ] .
```

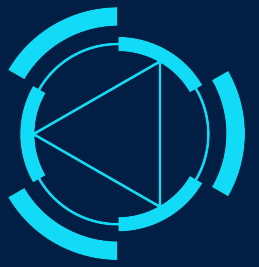
- says that the `:wins` class is the intersection of the `:results` class and nodes that have 1 as the value of their `:results_postionOrder` property
- effect: some instances of `:results` also become instances of `:wins`
  (check by running query q-a6.rq)
- instance of `:wins` from sample gets additional properties
  (check by running query q-a6-extra.rq)



Datalog Reasoning with RDFox

- ❑ Basic rules
- ❑ Aggregates
- ❑ Negation
- ❑ Filters
- ❑ Binds
- ❑ Incremental reasoning

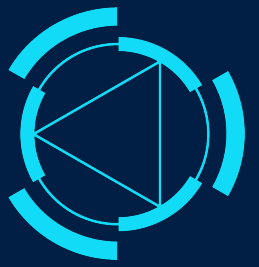
Datalog Rules



[*fact A*] :- [*fact B*]

“A is true whenever B is true”

Rule 1
Basic Rules



```
[?driver, :hasRacedIn, ?race] :-
    [?result, :results_race, ?race],
    [?result, :results_driver, ?driver] .
```

This rule will add a direct link between drivers and the races they raced in. Notice this has the same effect as axiom 5, but expressed more clearly

The same could seemingly be achieved with an equivalent SPARQL update, but rules offer several advantages over write queries.

Check for inferred triples with query q-r1.rq

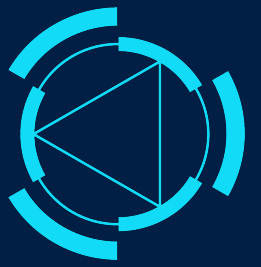
Rule 2
FILTERs in Rules

```
[?driver, :hasPodiumInRace, ?race] :-
    [?result, :results_race, ?race],
    [?result, :results_driver, ?driver],
    [?result, :results_positionOrder, ?positionOrder],
    FILTER(?positionOrder < 4) .
```

We can use FILTERs in rules too, in this case to find the races in which a driver reached the podium.

Check for inferred triples with query q-r2.rq

Rule 3
Aggregate Rule



```
[?driver, :hasRaceCount, ?raceCount] :-
    AGGREGATE(
        [?driver, :hasRacedIn, ?race]
        ON ?driver
        BIND COUNT(?race) AS ?raceCount
        ) .
```

This rule adds a count of races a driver has entered.

Notice we use the previously inferred `:hasRacedIn` property. Rules can be composed together, and RDFox will automatically find the order in which to run them.

Check for inferred triples with query q-r3.rq

Rule 4 - exercise

Aggregate Rule

```
[?___, :hasRaceWinCount, ?___] :-
    AGGREGATE(
        [?___, :results_driver, ?___],
        [?___, :results_positionOrder, 1]
        ON ?driver
        BIND COUNT(?___) AS ?___
        ) .
```

This rule adds a count of races a driver has *won* (provided they won at least 1).
We can put more than one atom *inside* the AGGREGATE statement.

Check for inferred triples with query q-r4.rq

Rule 5
Negation

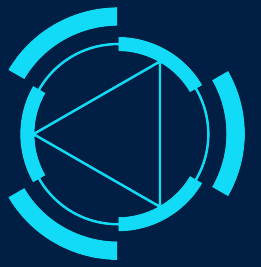
```
[?driver, a, :DriverWithoutPodiums] :-
    [?driver, a, :drivers],
    NOT EXISTS ?race IN (
        [?driver, :hasPodiumInRace, ?race]) .
```

This rule tells us that if a driver does not have a race where they were on the podium, then they are a `:DriverWithoutPodiums`.

So, we are looking for something that is **not** in the data (i.e. no race where the driver was on the podium).

Check for inferred triples with query q-r5.rq

Negation as Failure



Negation as failure is negation understood as the *absence* of a triple.

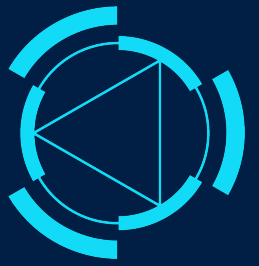
Using Negation as Failure leads to ‘non-monotonic reasoning’- the engine may need to retract some triples instead of just adding them. This adds another dimension to an already complicated problem of reasoning planning, but RDFox handles all of that automatically.

Example:
What if a driver without any podiums *so far* gets a podium next year? In 2021, for instance, George Russell got his first podium. In our data, which is accurate up to and including 2020, George Russell would be a `:DriverWithoutPodiums`. But if we add data from 2021, we will need to retract this fact.

- Exercise: Run the query “q-r5-extra.rq”, then import the data from file “2021-22.ttl” and try again

Rule 6 - exercise

BIND Rules

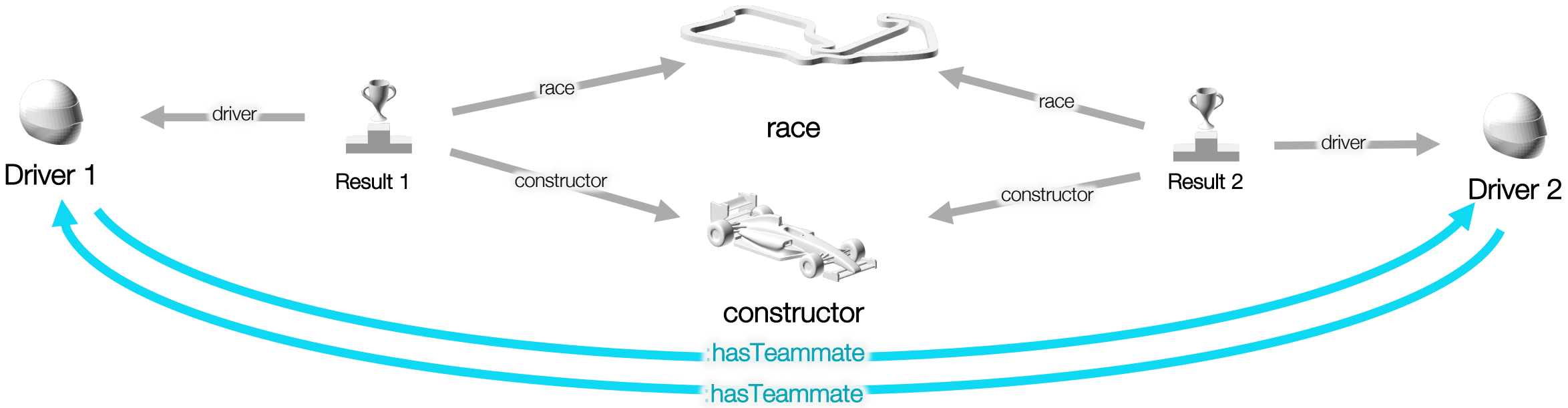


```
[?driver, :hasWinPercentage, ?percentage] :-
    [?___, :hasRaceCount, ?___],
    [?___, :hasRaceWinCount, ?___],
    BIND(?raceWinCount/?raceCount AS ?percentage) .
```

With `BIND` we can use various mathematical functions, string manipulation, regular expression matching, conditional binds, hashing and IRI creation.

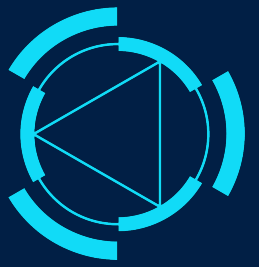
Check for inferred triples with query q-r6.rq

Bonus Exercise



- Write and import a rule to find out if two drivers have been teammates (use the predicate `:hasTeammate`).

Rule 7
Minimum rule

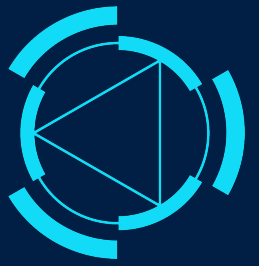


```
PREFIX : <http://www.oxfordsemantic.tech/f1demo/>

[?result1, :hasNextResult, ?result2] :-
    [?result1, a, :results],
    AGGREGATE(
        [?result1, :results_driver, ?driver],
        [?result1, :results_race, ?race1],
        [?race1, :races_date, ?date1],
        BIND(xsd:date(?date1) AS ?xsdDate1),
        [?otherResult, a, :results],
        [?otherResult, :results_driver, ?driver],
        [?otherResult, :results_race, ?otherRace],
        [?otherRace, :races_date, ?otherDate],
        BIND(xsd:date(?otherDate) AS ?otherXsdDate),
        FILTER(?xsdDate1 < ?otherXsdDate)
    ON ?result1 BIND MIN(?otherXsdDate) AS ?minXsdDate
    ),
    [?result1, :results_driver, ?driver],
    [?result2, :results_race, ?race2],
    [?result2, :results_driver, ?driver],
    [?race2, :races_date, ?date2],
    FILTER(xsd:date(?date2) = ?minXsdDate) .
```

We create a link between each result and the next result for the same driver.

Check for inferred triples with query q-r7.rq

Rule 8
Chain seeding

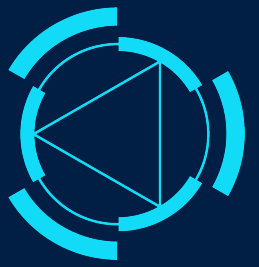


```
BASE <http://www.oxfordsemantic.tech/f1demo/>
PREFIX : <http://www.oxfordsemantic.tech/f1demo/>

[?streak, a, :streaks],
[?streak, :hasLastResult, ?win],
[?streak, :hasLength, 1] :-
    [?win, a, :results],
    [?win, :results_positionOrder, 1],
    NOT EXISTS ?previousWin IN (
        [?previousWin, a, :results],
        [?previousWin, :hasNextResult, ?win],
        [?previousWin, :results_positionOrder, 1]
    ),
    BIND(IRI(CONCAT("streaks/", MD5(STR(?win)))) AS ?streak) .
```

We create new IRIs that represent win streaks (when a driver won N consecutive races they participated in), first for N=1

Check for inferred triples with query q-r8.rq

Rule 9
Chain extension



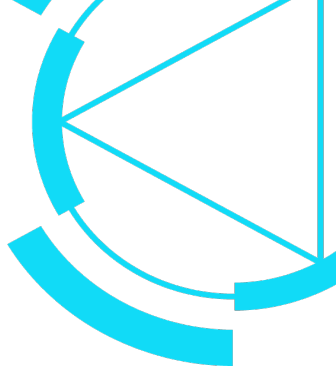
```
BASE <http://www.oxfordsemantic.tech/f1demo/>
PREFIX : <http://www.oxfordsemantic.tech/f1demo/>

[?newStreak, a, :streaks],
[?newStreak, :hasLastResult, ?newWin],
[?newStreak, :hasLength, ?newLength] :-
    [?oldStreak, a, :streaks],
    [?oldStreak, :hasLength, ?oldLength],
    BIND(?oldLength + 1 AS ?newLength),
    [?oldStreak, :hasLastResult, ?oldWin],
    [?oldWin, :hasNextResult, ?newWin],
    [?newWin, :results_positionOrder, 1],
    BIND(IRI(CONCAT("streaks/", MD5(CONCAT(STR(?oldStreak), STR(?newWin))) )) AS ?newStreak) .
```

We create new IRIs that represent win streaks.

List longest win streaks in history by running query q-r9.rq

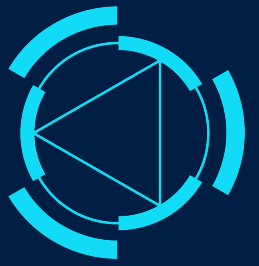
Bonus Exercise

- Write a rule that creates a new class for streaks which cannot be extended



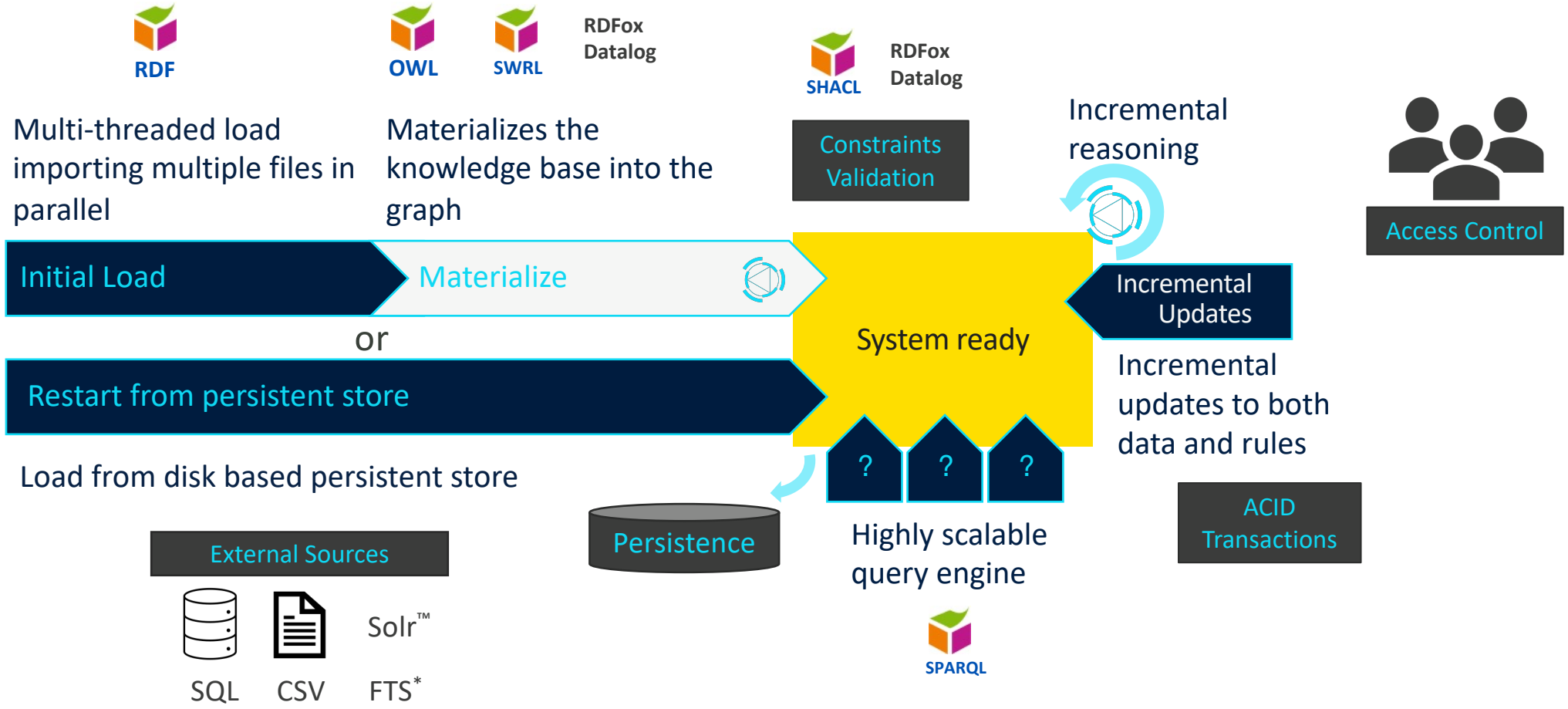
RDFox: How does it work?

Incremental Reasoning



- One of the key strengths of RDFox
- Data can be imported at any time, and the RDFox will automatically compute the necessary inferences
- This all happens *incrementally*, i.e. without the need to reboot
- Incremental reasoning open the door to many novel use cases which were previously impractical

How does it work?

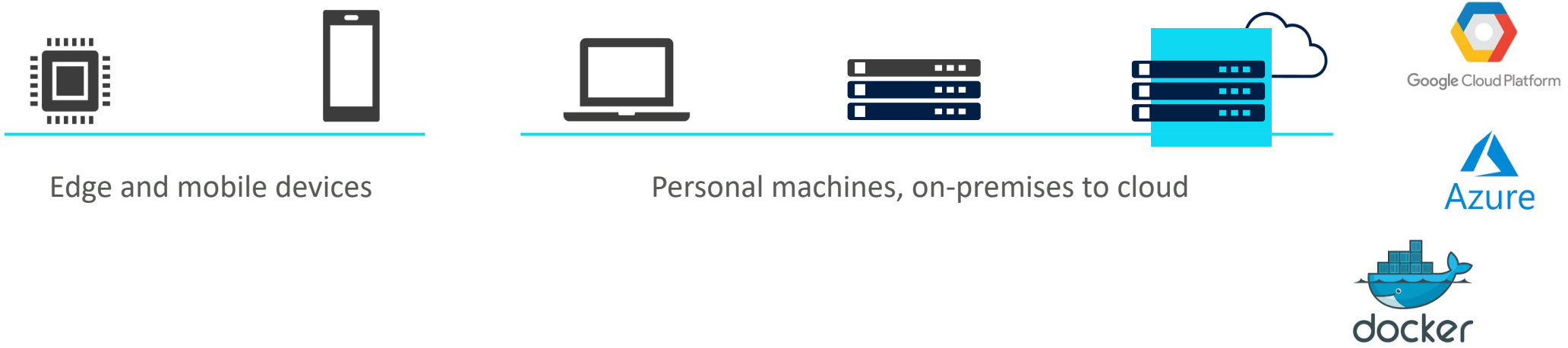


RDFox vs. Other Solutions

Feature	Example	RDFox	Materialisation Competitors	Query Rewriting Competitors
Ontological reasoning	"List all financial instruments / derivative instruments / futures." [class inferencing]	✓	✓*	✓**
Graph analytics	"What is the nearest electrical switch for a circuit?" [negation + recursion]	✓		
Data Analytics	"What is the total volume and value of trades?" [aggregation + arithmetic]	✓		
Local constraints validation	"No trades over a limit & no traders without trades!" [filters + negation]	✓	Partial (SHACL)	
Global constraint validation	"Every component must be connected to a power source!" [negation + recursion]	✓		
All of the above	"What is the total value of assets owned through holdings?"	✓		

*** many times slower** than RDFox
**** orders of magnitude slower** than RDFox

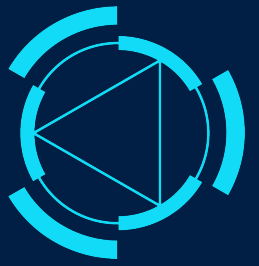
Deployment Options



Edge and mobile devices

Personal machines, on-premises to cloud

Your final exercise…



Click on the link below to start filling out our quick survey. It won't take you more than 3 minutes!

https://oxfordsemantictech.typeform.com/to/qcHnRESb