



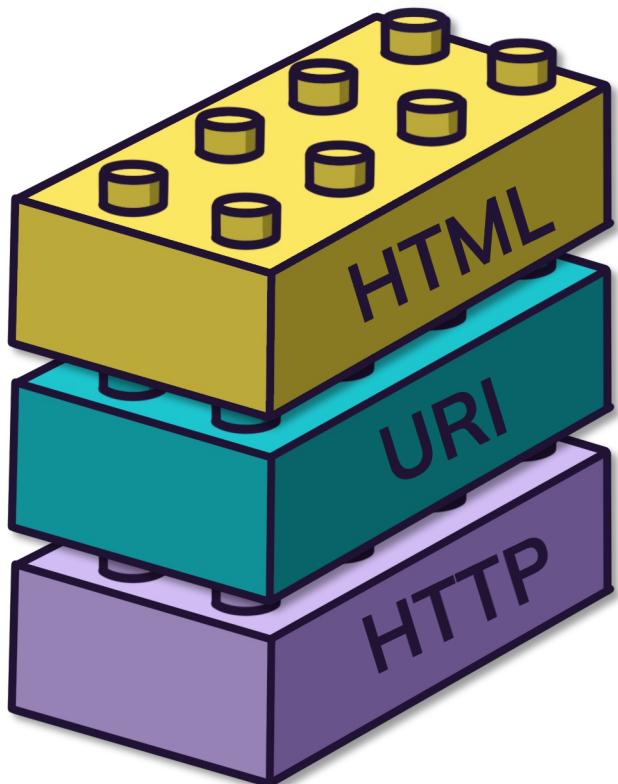
# The Carpentries Incubator

## Identifier and Web Locations

How to Identify and Retrieve Data on the World Wide Web

Fundamentals of Scientific Metadata: Why Context Matters

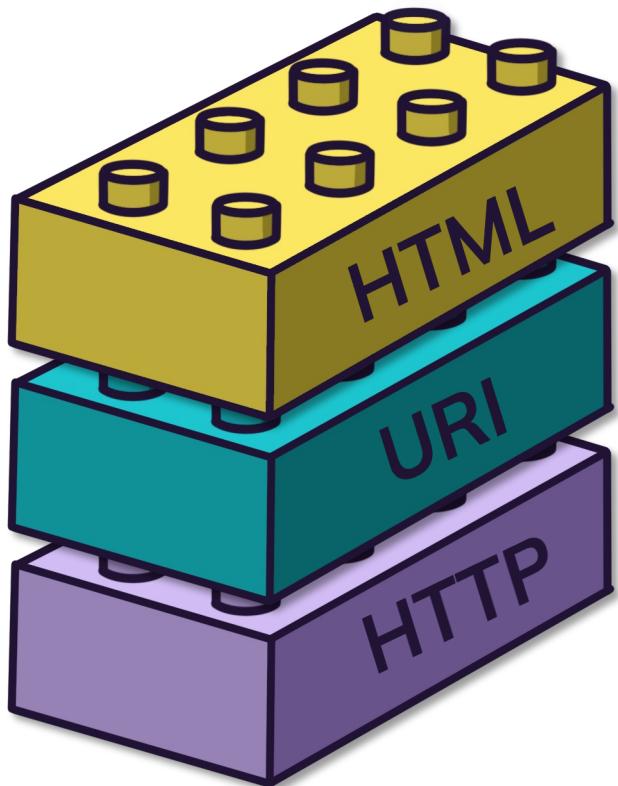
# World Wide Web building blocks



HyperText Markup Language

Uniform Resource Identifiers

HyperText Transfer Protocol



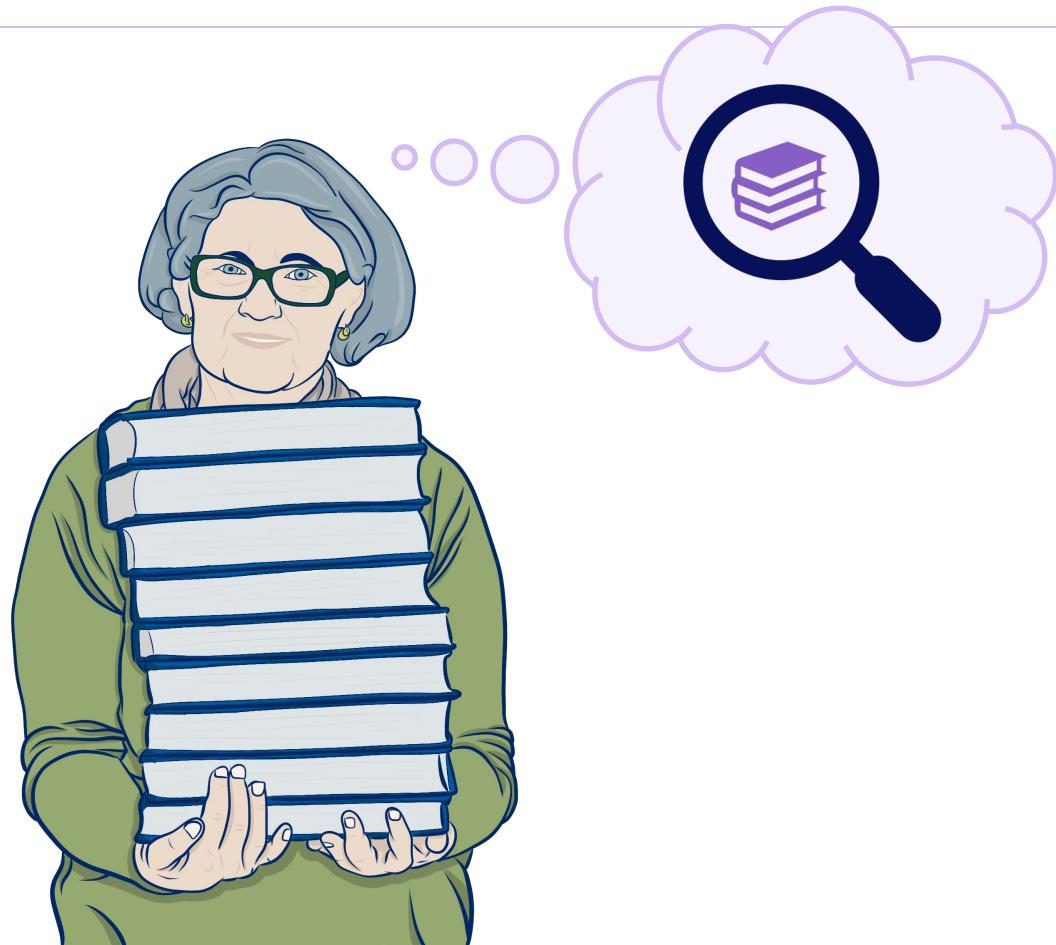
HyperText Markup Language

Uniform Resource Identifiers

HyperText Transfer Protocol



# Uniform Resource Identifier - Librarian



# Uniform Resource Identifier - Librarian





# Uniform Resource Identifier – Molecular Biology





# Uniform Resource Identifier – Molecular Biology





# Unique Identification Codes



- globally unique and persistent identification
- central, world or community wide registration
- referenceable, unambiguous



Identifier in your  
research area



# Web Identifier





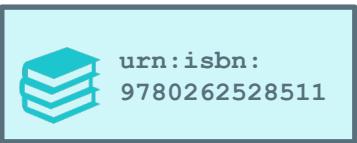
# URN – URL – URI





# URN – URL – URI

## URN



## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol





# URN – URL – URI

## URN



## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol





# URN – URL – URI

## URN



## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol

## URL

UNIFORM RESOURCE  
LOCATOR

- address
- series of characters
- comm. protocol



# URN – URL – URI

## URN

 urn:isbn:  
9780262528511

## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol

## URL

UNIFORM RESOURCE  
LOCATOR

- address
- series of characters
- comm. protocol

## URL

 [https://orcid.org/  
0000-0003-3043-  
5657](https://orcid.org/0000-0003-3043-5657)

 [https://arabidops  
is.org/servlets/T  
airObject?id=1360  
48&type=locus](https://arabidopsis.org/servlets/TairObject?id=136048&type=locus)



# URN – URL – URI

## URN

 urn:isbn:  
9780262528511

## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol

## URL

UNIFORM RESOURCE  
LOCATOR

- address
- series of characters
- comm. protocol

## URL

  
`https://orcid.org`  
/0000-0003-3043-  
5657

  
`https://arabidops`  
is.org/servlets/T  
airObject?id=1360  
48&type=locus



# URN – URL – URI

## URN



## URI

UNIFORM RESOURCE IDENTIFIER

## URN

UNIFORM RESOURCE  
NAME

- unique identifier
- series of characters
- **no** protocol

## URL

UNIFORM RESOURCE  
LOCATOR

- address
- series of characters
- comm. protocol

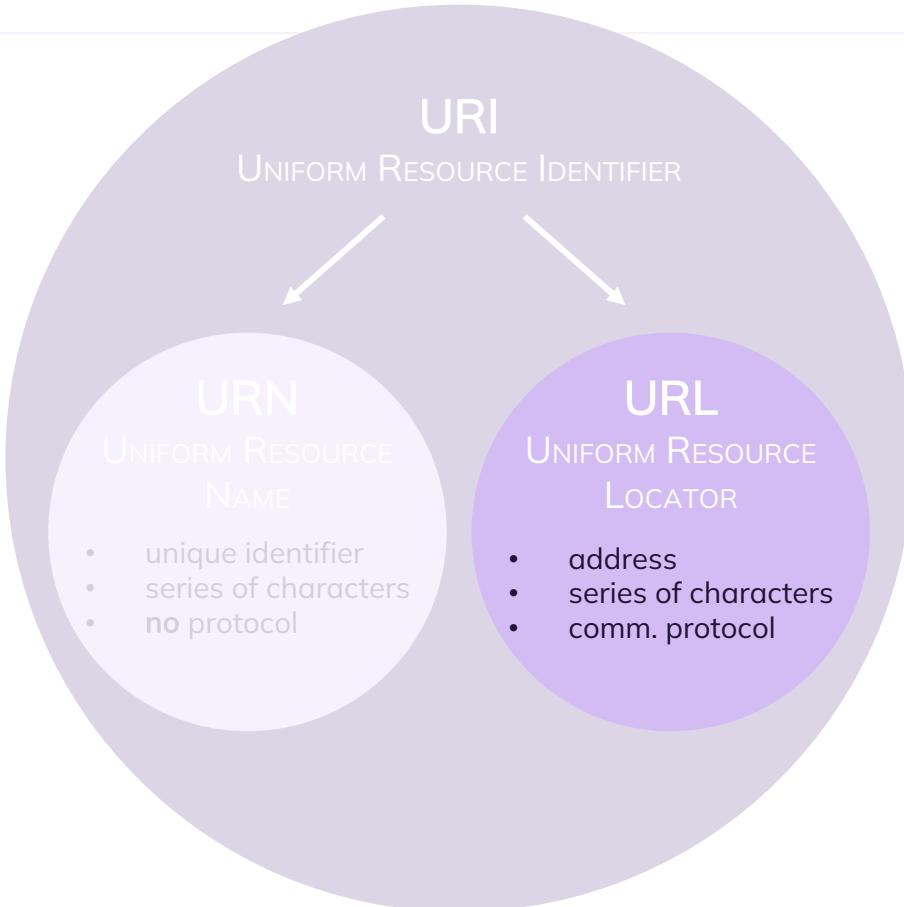
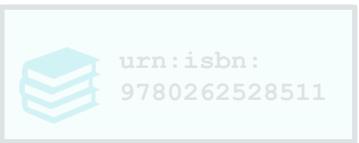
## URL





# URN – URL – URI

## URN



## URL



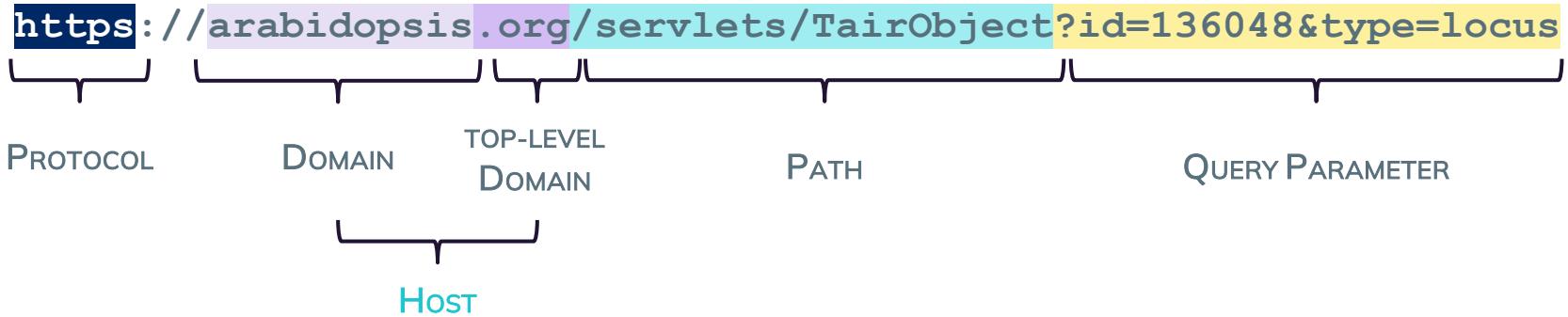


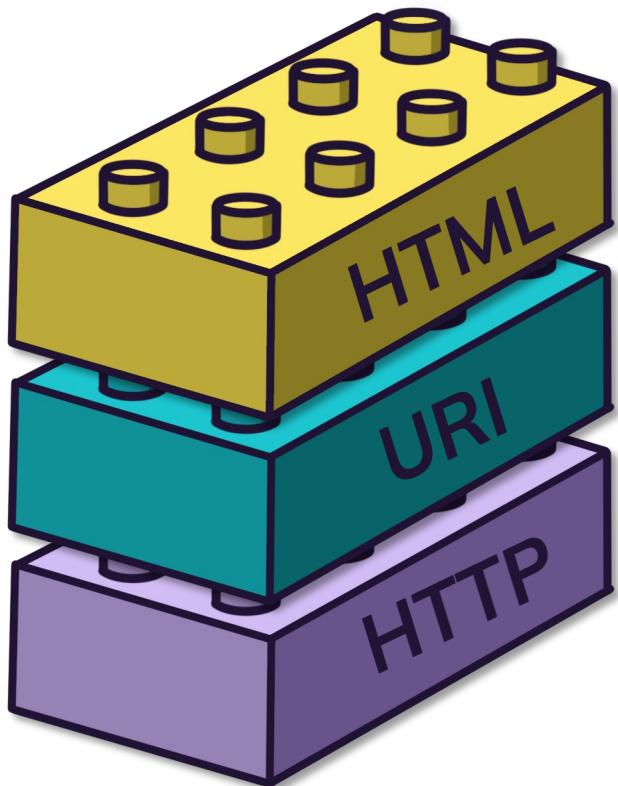
## URL structure

`https://arabidopsis.org/servlets/TairObject?id=136048&type=locus`



# URL structure





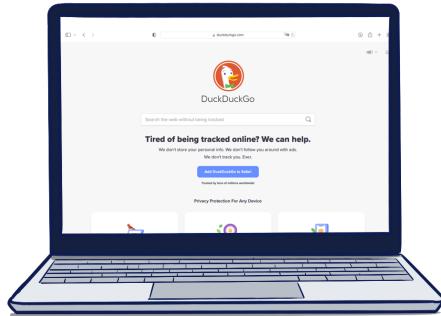
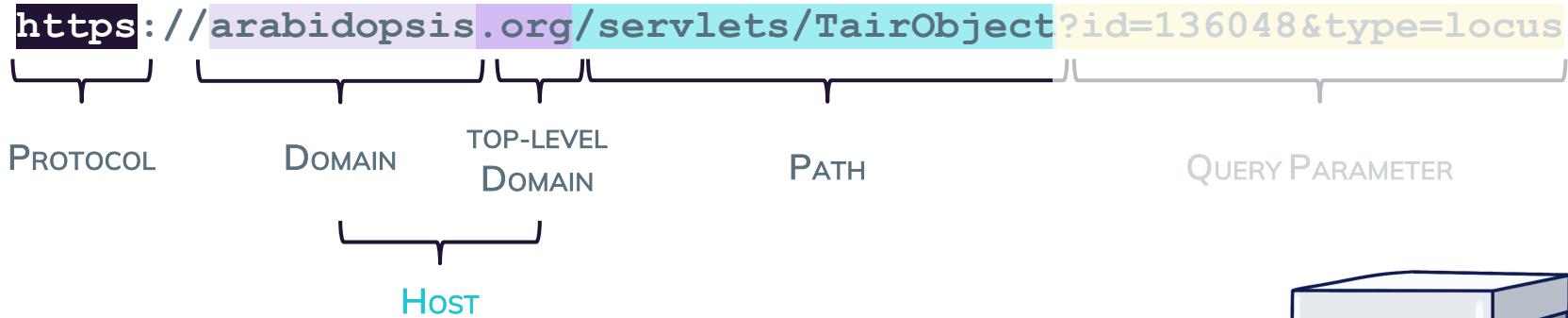
HyperText Markup Language

Uniform Resource Identifiers

HyperText Transfer Protocol



# HTTP request



HTTP Request

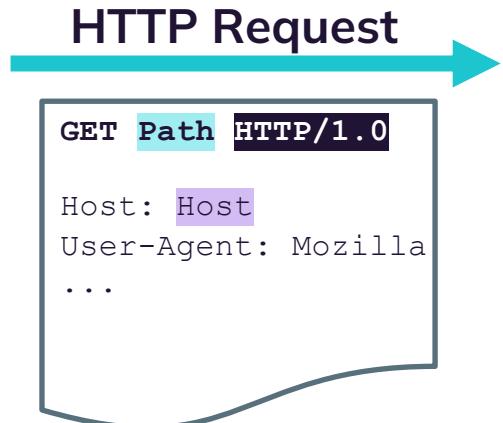
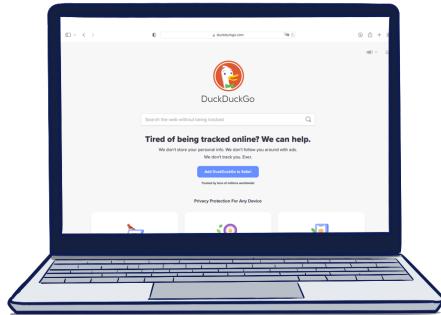
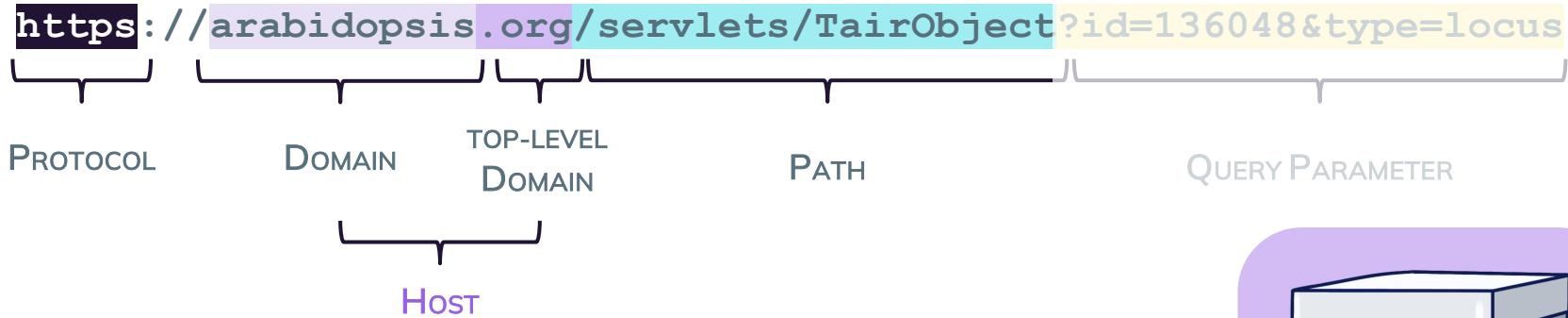
GET Path HTTP/1.0

Host: Host  
User-Agent: Mozilla  
...

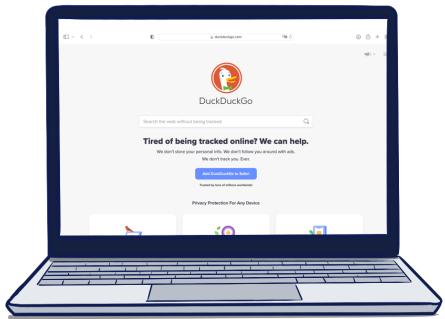




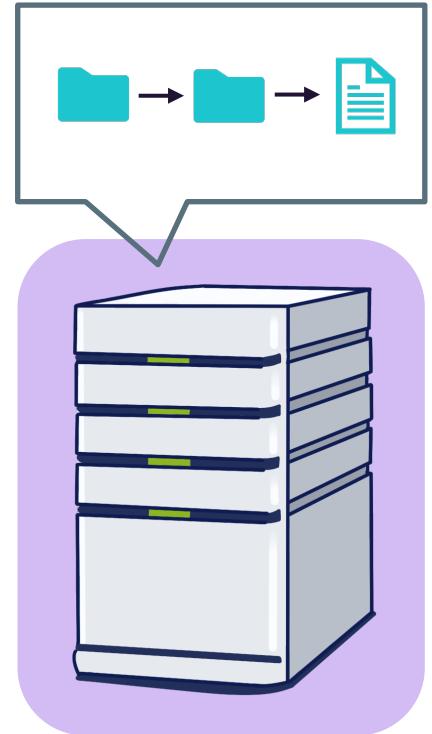
# HTTP request



# HTTP request

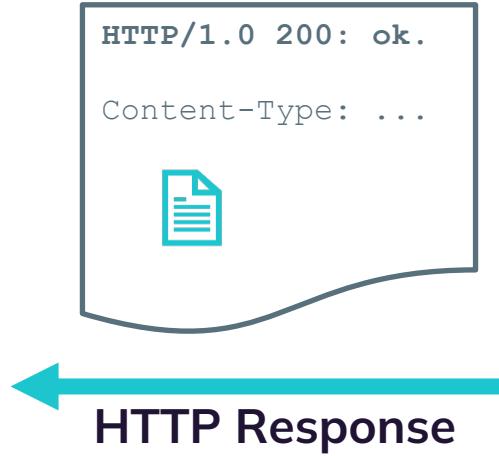
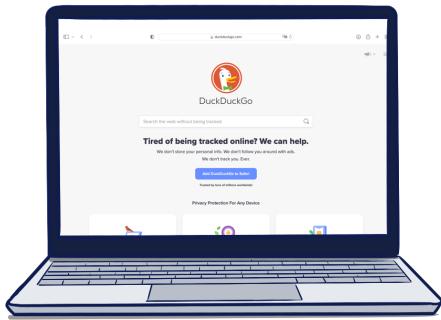


```
GET Path HTTP/1.0  
Host: Host  
User-Agent: Mozilla  
...
```



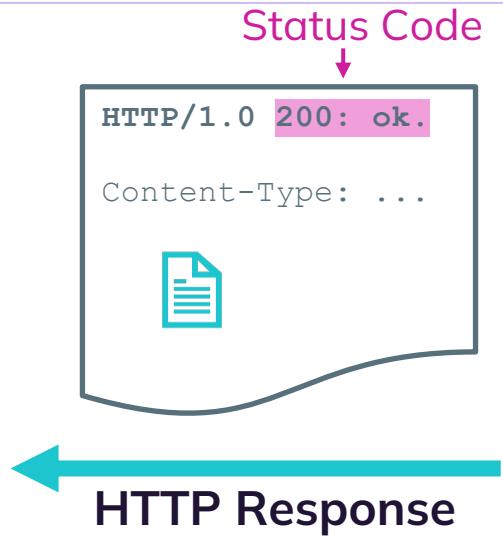
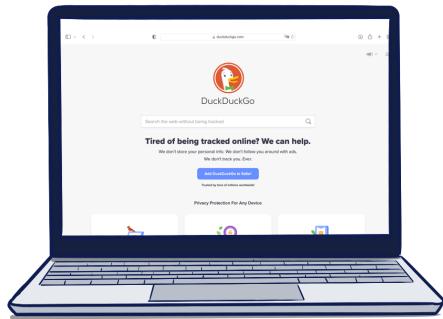


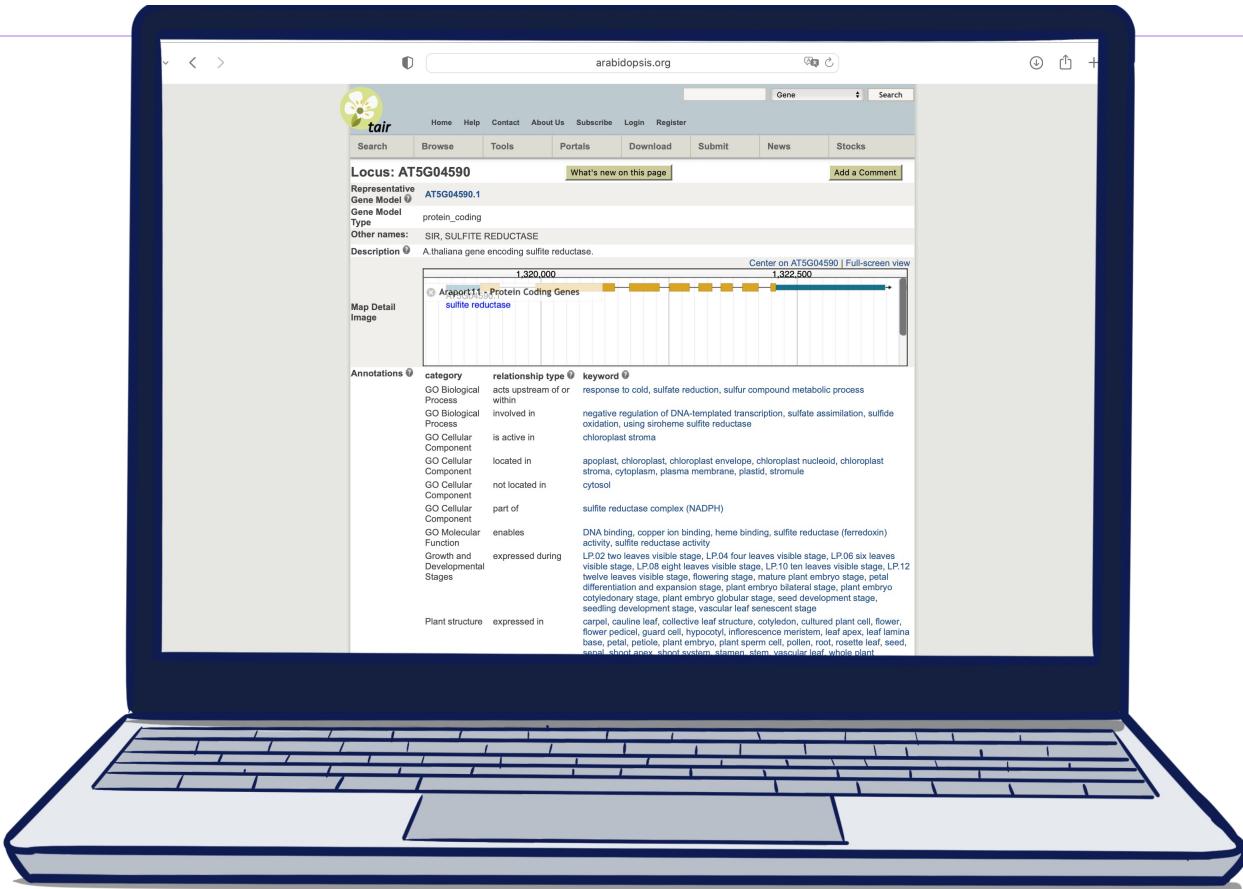
# HTTP response

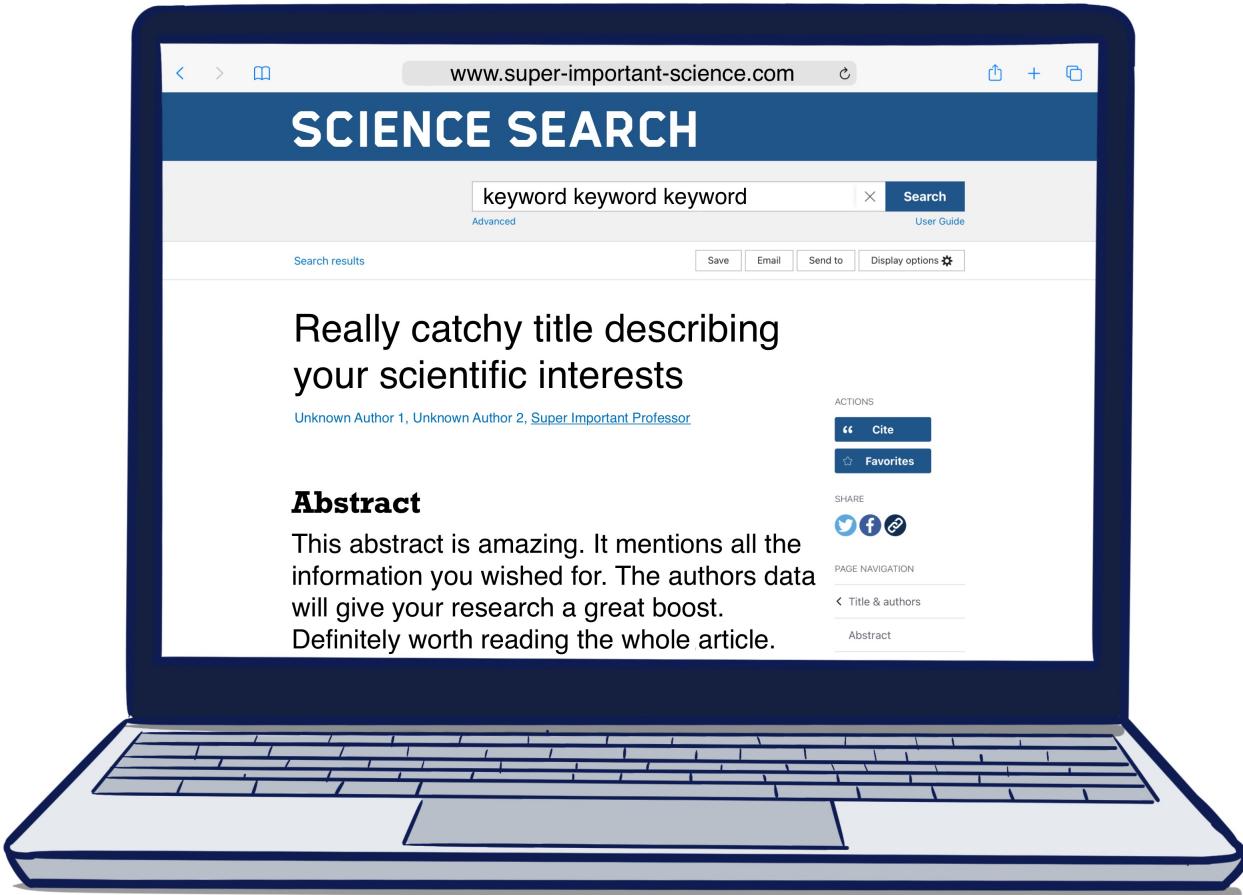


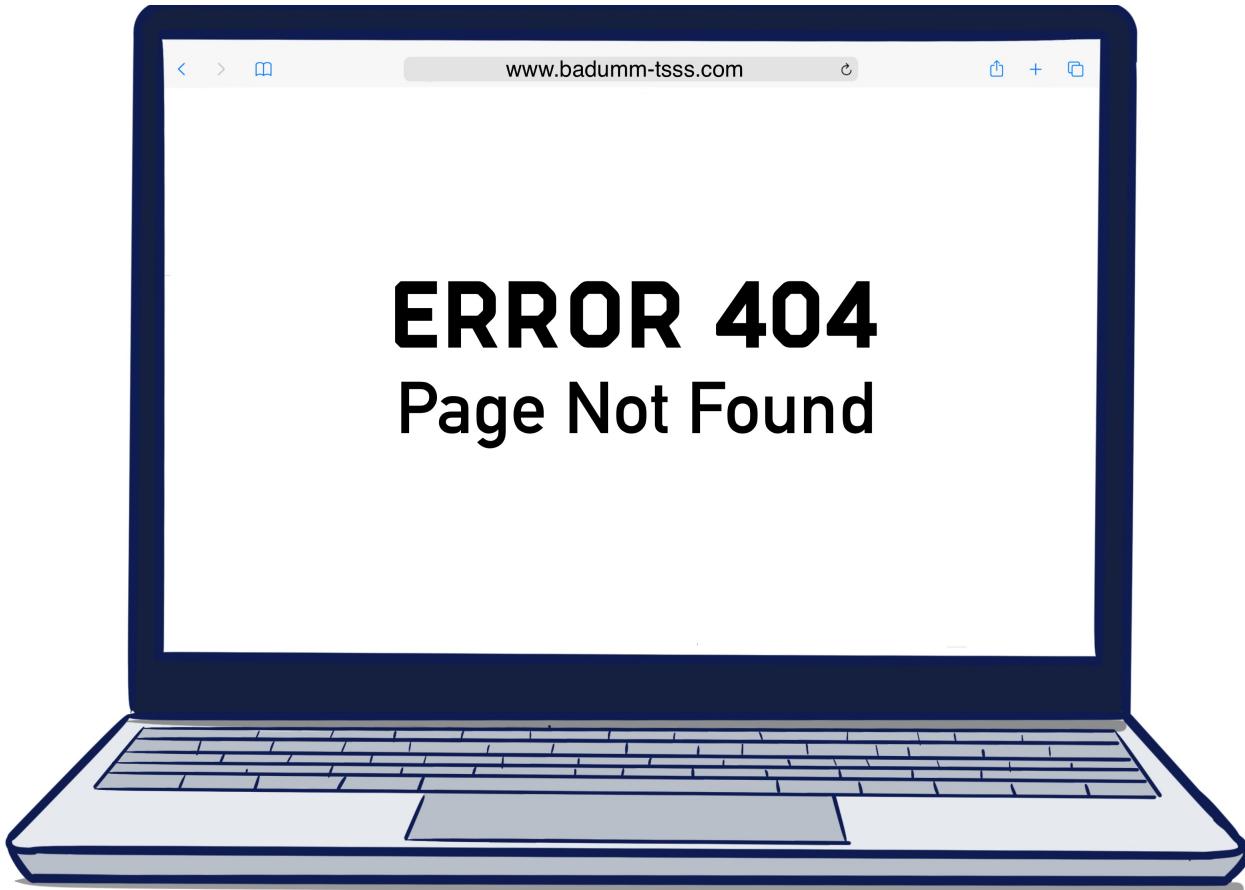


# HTTP response







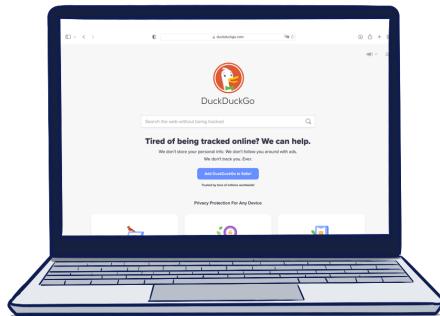




# HTTP request

`https://arabidopsis.org/servlets/TairObject?id=136048&type=locus`

PROTOCOL            HOST            PATH            QUERY PARAMETER



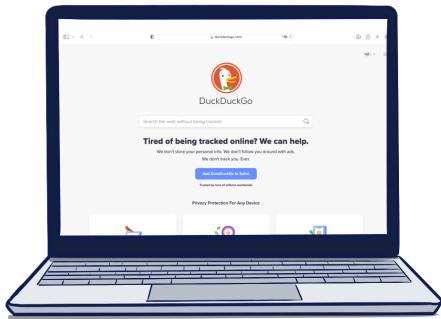
HTTP Request

GET Path HTTP/1.0

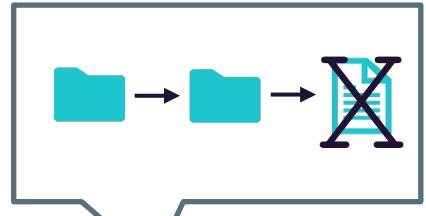
Host: Host  
User-Agent: Mozilla  
...



# ERROR 404: Not Found

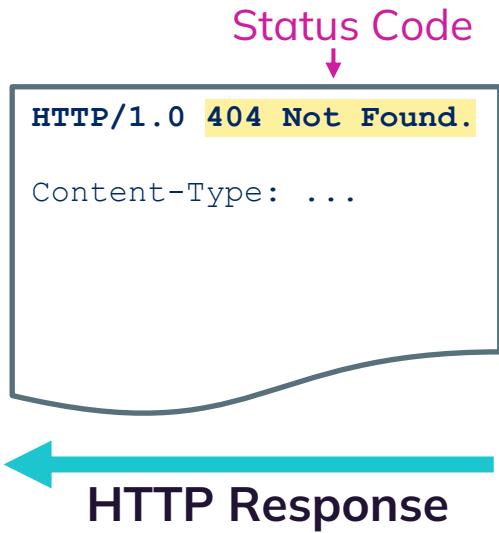


```
GET Path HTTP/1.0  
  
Host: Host  
User-Agent: Mozilla  
...
```





# ERROR 404: Not Found



# Error 404

## Demonstration

