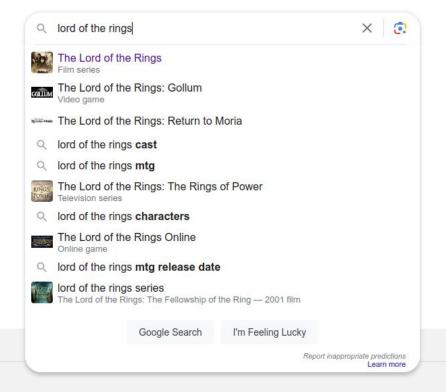
# Semantic Web and its Application in Research Data Management

JSON-LD, RO-Crate, schema.org, ARC

## What is Semantic Web?

Motivation



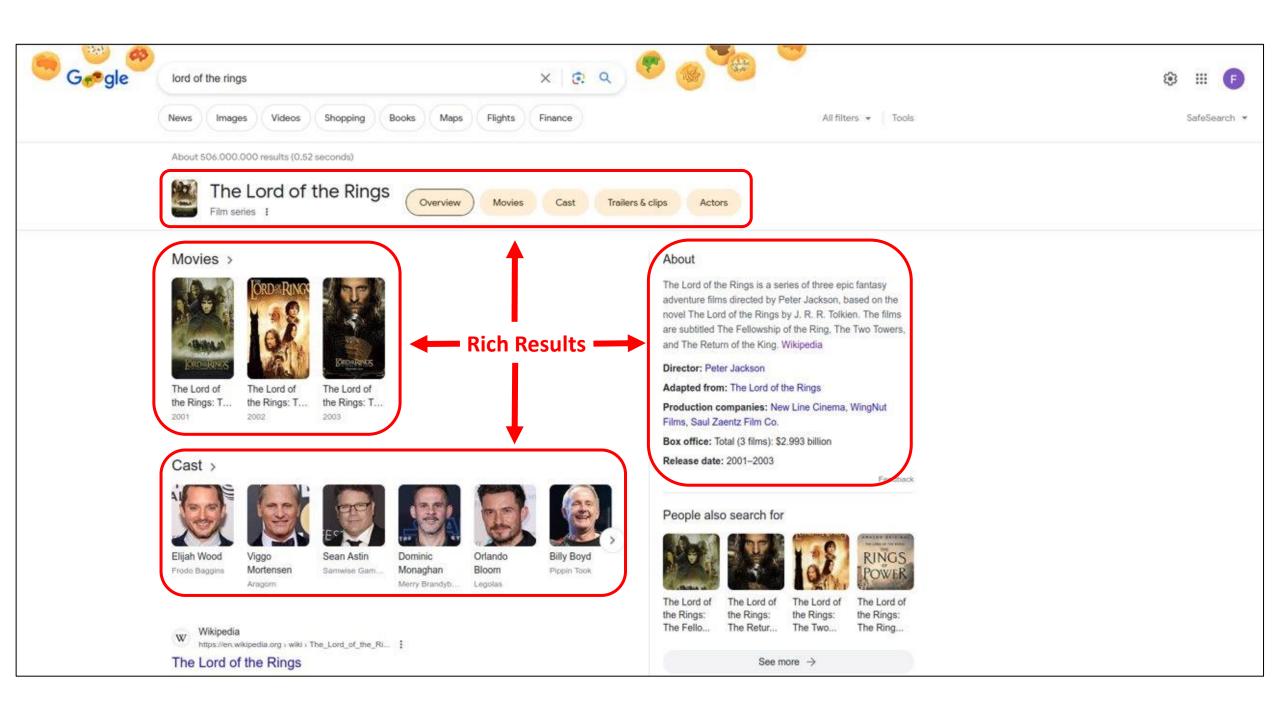


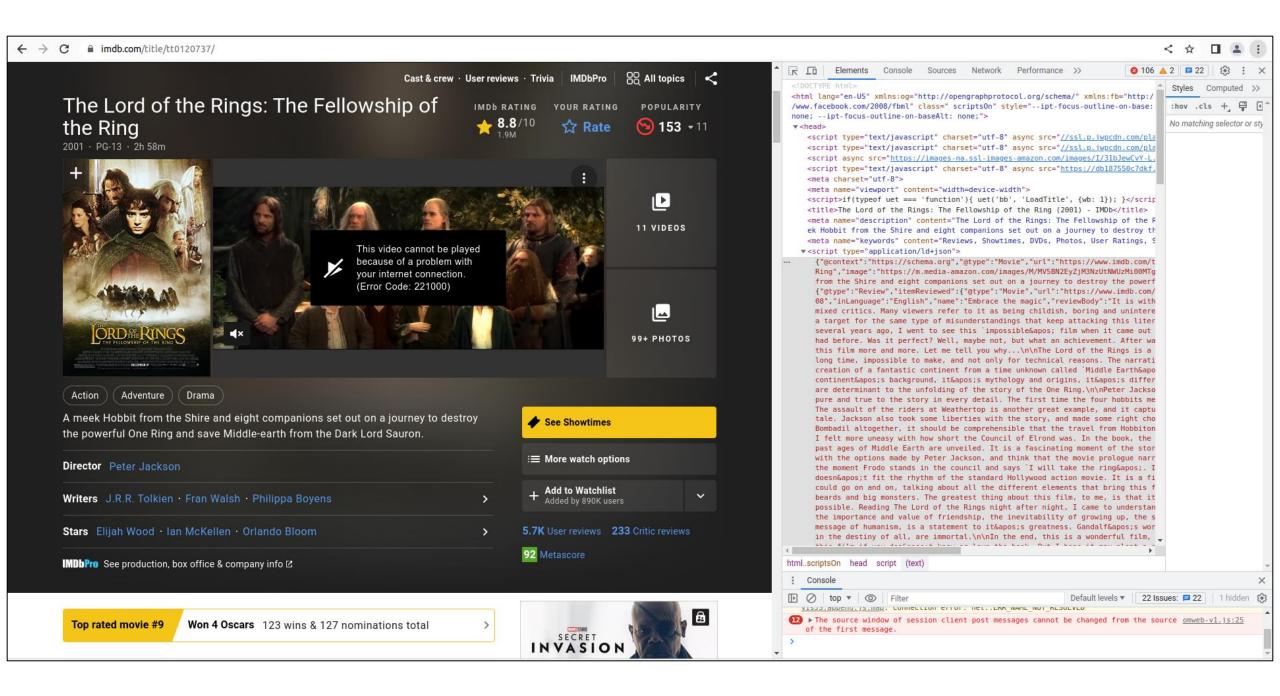
Germany

About Store

Advertising

Business How Search works





```
"@context": "https://schema.org",
2
         "@type": "Movie",
3
         "url": "https://www.imdb.com/title/tt0120737/",
4
         "name": "The Lord of the Rings: The Fellowship of the Ring",
5
         "image": "https://m.media-amazon.com/images/M/MV5BN2EyZjM3NzUtNWUZMi00MTgxLWI0NTctMzY4M2VlOTdjZWRiXkEyXkFqcGdeQXVyNDUzOTQ5MjY@. V1 .jpg",
         "description": "A meek Hobbit from the Shire and eight companions set out on a journey to destroy the powerful One Ring and save Middle-earth from the Dark Lord Sauron.",
8 >
         "review": { --
22
         },
23
         "aggregateRating": {
             "@type": "AggregateRating",
24
25
             "ratingCount": 1925669,
             "bestRating": 10,
26
27
             "worstRating": 1,
             "ratingValue": 8.8
28
29
30
         "contentRating": "PG-13",
31
         "genre": [
             "Action",
32
33
             "Adventure",
34
             "Drama"
35
36
         "datePublished": "2001-12-19",
         "keywords": "ring,quest,hobbit,middle earth,elf",
37
         "trailer": {--
38 >
51
         },
         "actor": [
52
53
                 "@type": "Person",
54
55
                 "url": "https://www.imdb.com/name/nm0000704/",
                 "name": "Elijah Wood"
56
57
58
59
                 "@type": "Person",
                 "url": "https://www.imdb.com/name/nm0005212/",
60
61
                 "name": "Ian McKellen"
62
63
                 "@type": "Person",
64
65
                 "url": "https://www.imdb.com/name/nm0089217/",
                 "name": "Orlando Bloom"
66
67
68
         "director": [
69
70
                 "@type": "Person",
71
                 "url": "https://www.imdb.com/name/nm0001392/",
72
                 "name": "Peter Jackson"
73
74
75
```

## What is Semantic Web?

**Technologies** 

### JSON: JavaScript Object Notation

- Human- and machine-readable format
- Objects are simple lists of key-value pairs
- Attribute identifiers are strings
- Values are strings, numerical values, lists or objects

```
"@context": "https://schema.org",
2
         "@type": "Movie",
3
         "url": "https://www.imdb.com/title/tt0120737/",
4
         "name": "The Lord of the Rings: The Fellowship of the Ring",
5
         "image": "https://m.media-amazon.com/images/M/MV5BN2EyZjM3NzUtNWUZMi00MTgxLWI0NTctMzY4M2VlOTdjZWRiXkEyXkFqcGdeQXVyNDUzOTQ5MjY@. V1 .jpg",
         "description": "A meek Hobbit from the Shire and eight companions set out on a journey to destroy the powerful One Ring and save Middle-earth from the Dark Lord Sauron.",
8 >
         "review": { --
22
         },
23
         "aggregateRating": {
             "@type": "AggregateRating",
24
25
             "ratingCount": 1925669,
             "bestRating": 10,
26
27
             "worstRating": 1,
             "ratingValue": 8.8
28
29
30
         "contentRating": "PG-13",
31
         "genre": [
             "Action",
32
33
             "Adventure",
34
             "Drama"
35
36
         "datePublished": "2001-12-19",
         "keywords": "ring,quest,hobbit,middle earth,elf",
37
         "trailer": {--
38 >
51
         },
         "actor": [
52
53
                 "@type": "Person",
54
55
                 "url": "https://www.imdb.com/name/nm0000704/",
                 "name": "Elijah Wood"
56
57
58
59
                 "@type": "Person",
                 "url": "https://www.imdb.com/name/nm0005212/",
60
61
                 "name": "Ian McKellen"
62
63
                 "@type": "Person",
64
65
                 "url": "https://www.imdb.com/name/nm0089217/",
                 "name": "Orlando Bloom"
66
67
68
         "director": [
69
70
                 "@type": "Person",
71
                 "url": "https://www.imdb.com/name/nm0001392/",
72
                 "name": "Peter Jackson"
73
74
75
```

#### JSON-LD: JSON for Linked Data

- Objects have a unique property "@id", preferably an URI
- Objects have a "@type", referencing some vocabulary
- Types are specified within a "@context"
- Example:

```
Global identifier

"@context": "https://json-ld.org/contexts/person.jsonld",

"@id": "http://dbpedia.org/resource/John_Lennon",

"@type": "Person",

"name": "John Lennon",

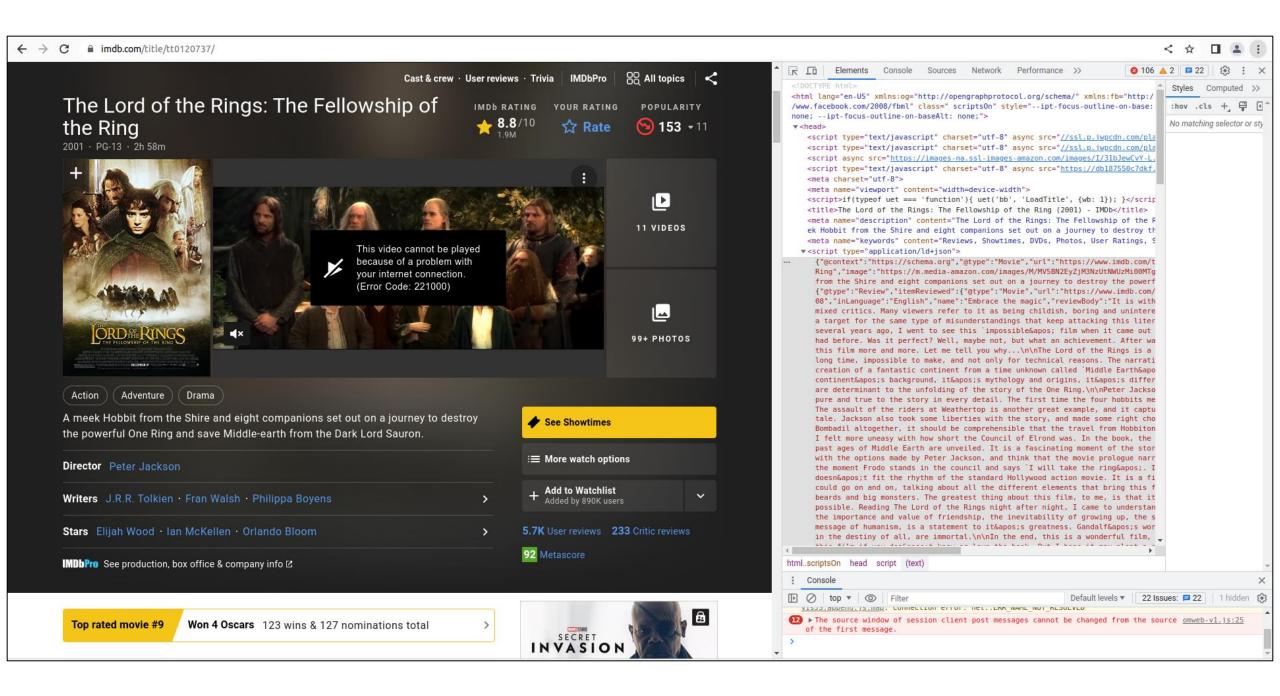
"born": "1940-10-09",

"spouse": "http://dbpedia.org/resource/Cynthia_Lennon"

Type definition
```

### Embedding in Websites

- Google requires JSON-LD markup for rich results
- Must be in the header of html-websites
- Must use schema.org vocabulary
- Google adds additional constraints



### Embedding in Websites

- Google requires JSON-LD markup for rich results
- Must be in the header of html-websites
- Must use schema.org vocabulary
- Google adds additional constraints



#### Movie

A Schema.org Type

Thing > CreativeWork > Movie

A movie.

Property	Expected Type	Description
Properties from Movie		
actor	Person	An actor, e.g. in TV, radio, movie, video games etc., or in an event. Actors can be associated with individual items or with a series, episode, clip. Supersedes actors.
	Country	The country of origin of something, including products as well as creative works such as movie and TV content.
countryOfOrigin		In the case of TV and movie, this would be the country of the principle offices of the production company or individual responsible for the movie. For other kinds of CreativeWork it is difficult to provide fully general guidance, and properties such as contentLocation and locationCreated may be more applicable.
		In the case of products, the country of origin of the product. The exact interpretation of this may vary by context and product type, and cannot be fully enumerated here.
director	Person	A director of e.g. TV, radio, movie, video gaming etc. content, or of an event. Directors can be associated with individual items or with a series, episode, clip. Supersedes directors.
duration	Duration	The duration of the item (movie, audio recording, event, etc.) in ISO 8601 date format.
musicBy	MusicGroup or Person	The composer of the soundtrack.
productionCompany	Organization	The production company or studio responsible for the item, e.g. series, video game, episode etc.
subtitleLanguage	Language or Text	Languages in which subtitles/captions are available, in IETF BCP 47 standard format.
	Text or URL	An EIDR (Entertainment Identifier Registry) identifier representing at the most general/abstract level, a work of film or television.
titleEIDR		For example, the motion picture known as "Ghostbusters" has a titleEIDR of "10.5240/7EC7-228A-510A-053E-CBB8-J". This title (or work) may have several variants, which EIDR calls "edits". See editEIDR.
		Since schema.org types like Movie, TVEpisode, TVSeason, and TVSeries can be used for both works and their multiple expressions, it is possible to use titleEIDR alone (for a general description), or alongside editEIDR for a more edit-specific description.
trailer	VideoObject	The trailer of a movie or TV/radio series, season, episode, etc.
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf

Schema.org Docs Schemas Validate About Q

#### **Person**

A Schema.org Type

Thing > Person

[more...]

A person (alive, dead, undead, or fictional).

[more...]

Property	Expected Type	Description
Properties from Person		
additionalName	Text	An additional name for a Person, can be used for a middle name.
address	PostalAddress or Text	Physical address of the item.
affiliation	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
alumniOf	EducationalOrganization or Organization	An organization that the person is an alumni of. Inverse property: <mark>alumni</mark>
award	Text	An award won by or for this item. Supersedes awards.
birthDate	Date	Date of birth.
birthPlace	Place	The place where the person was born.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintaine by an organization or business person.
callSign	Text	A callsign, as used in broadcasting and radio communications to identify people, radio and TV stations, or vehicles.
children	Person	A child of the person.
colleague	Person or URL	A colleague of the person. Supersedes colleagues.
contactPoint	ContactPoint	A contact point for a person or organization. Supersedes contactPoints.
deathDate	Date	Date of death.
deathPlace	Place	The place where the person died.
duns	Text	The Dun & Bradstreet DUNS number for identifying an organization or busin person.
email	Text	Email address.
familyName	Text	Family name. In the U.S., the last name of a Person.
faxNumber	Text	The fax number.
follows	Person	The most generic uni-directional social relation.
funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.
funding	Grant	A Grant that directly or indirectly provide funding or sponsorship for this iter See also ownershipFundingInfo. Inverse property: fundedItem

#### JSON-LD: JSON for *Linked Data*

- Unique IDs allow flattened form
- Example:

```
"@context": {
         "@vocab": "http://xmlns.com/foaf/0.1/"
       "@graph":[
           "@id": "http://manu.sporny.org/about#manu",
           "@type": "Person",
           "name": "Manu Sporny",
           "knows": { "@id": "https://greggkellogg.net/foaf#me" }
10
11
12
           "@id": "https://greggkellogg.net/foaf#me",
13
           "@type": "Person",
14
           "name": "Gregg Kellogg"
15
16
17
18
```

#### JSON-LD Contexts

- Aliases can be defined in the "@context" field
- Example:

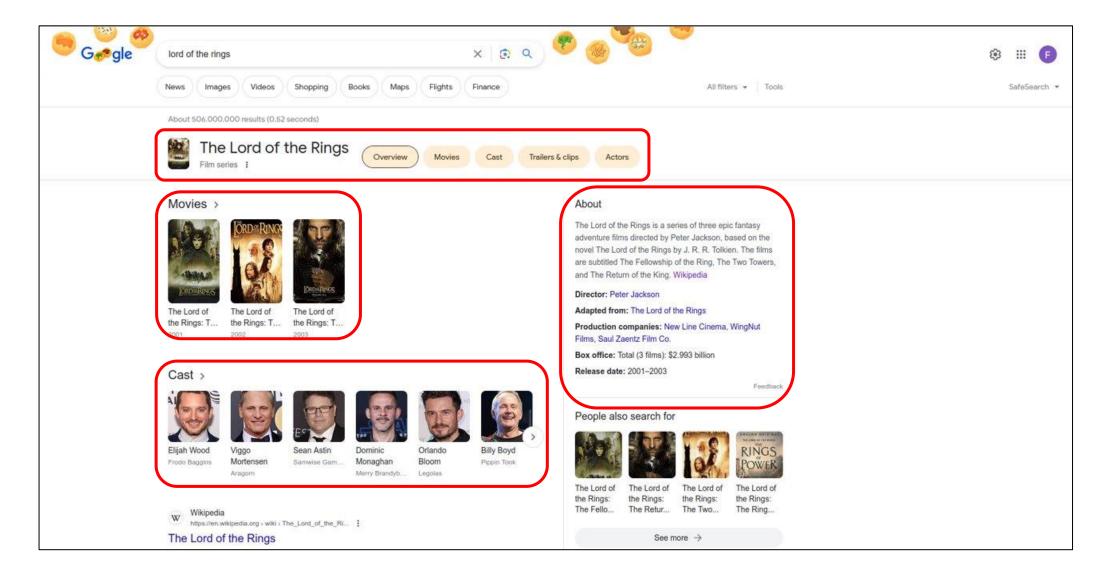
```
"@context": {
         "@vocab": "http://xmlns.com/foaf/0.1/"
       "@id": "http://manu.sporny.org/about#manu",
       "@type": "Person",
 6
       "name": "Manu Sporny",
       "knows": {
 8
         "@id": "https://greggkellogg.net/foaf#me",
 9
         "@type": "Person",
10
         "name": "Gregg Kellogg"
11
12
13
```

```
"@context": {
         "@vocab": "http://xmlns.com/foaf/0.1/"
       "@graph":[
           "@id": "http://manu.sporny.org/about#manu",
           "@type": "Person",
           "name": "Manu Sporny",
           "knows": { "@id": "https://greggkellogg.net/foaf#me" }
10
11
12
           "@id": "https://greggkellogg.net/foaf#me",
13
           "@type": "Person",
14
           "name": "Gregg Kellogg"
15
16
17
18
```

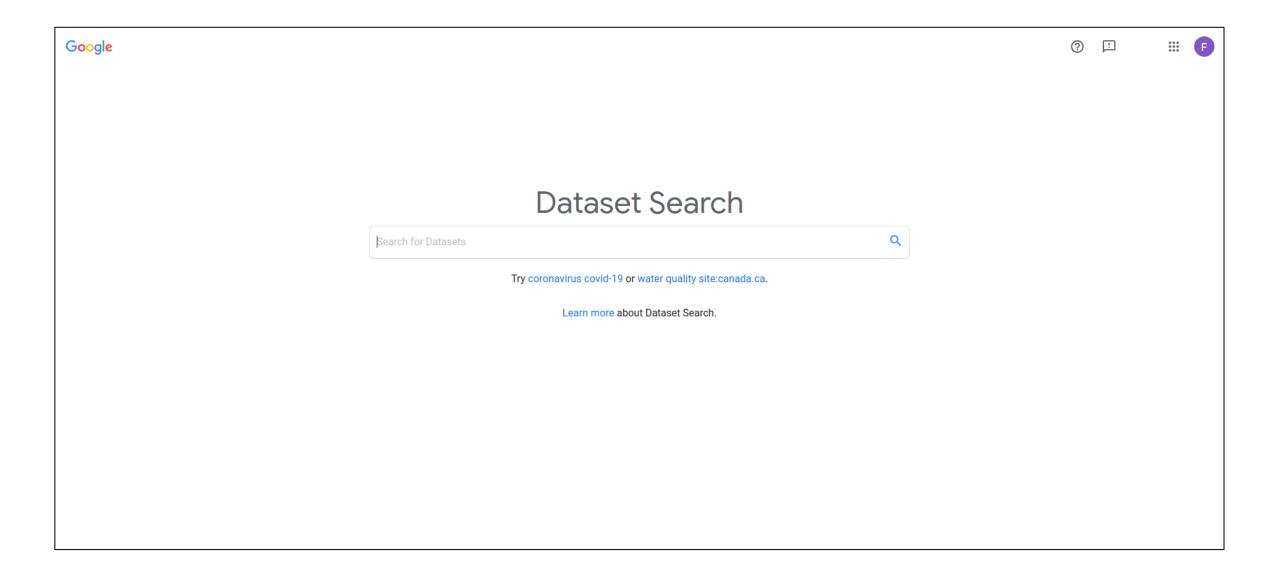
## Semantic Web for Research Data

What has this to do with RDM?

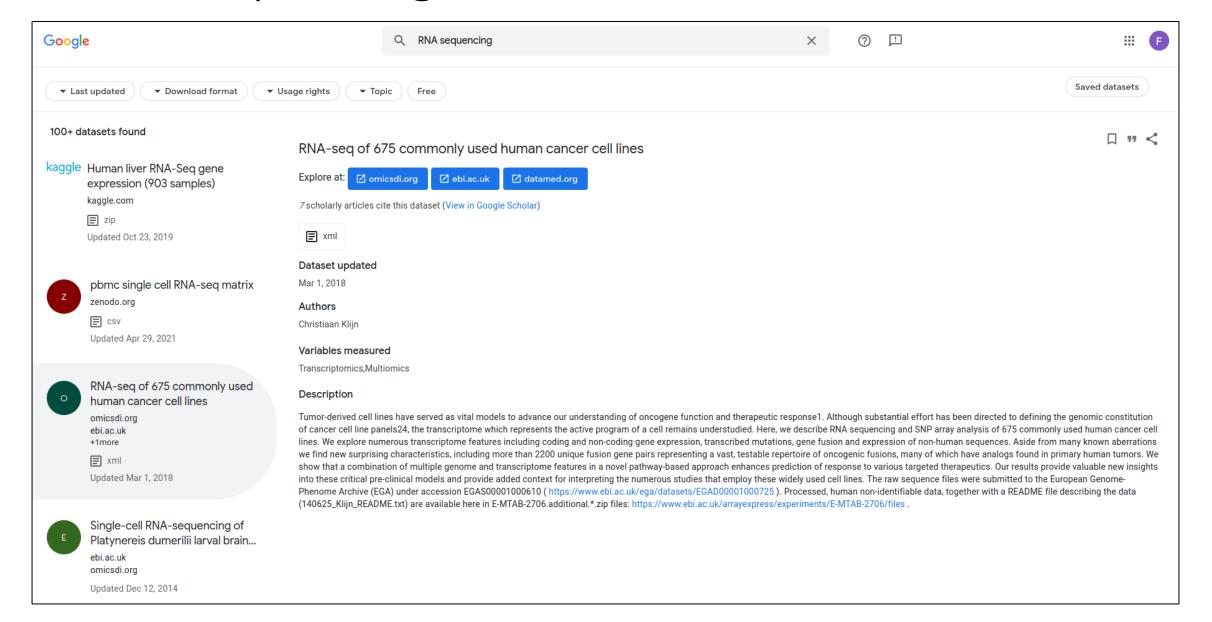
### We want this $\downarrow$ for Research Data



## First Step: Google Dataset Search



## First Step: Google Dataset Search



## Semantic Web for Research Data

**Technologies** 

#### **RO-Crate**

- Filesystem-based format for Datasets based on JSON-LD
- Data entities: files and directories
- Contextual entities: people, publications, organizations, measurement
- Mostly uses schema.org terminology

### Datasets in schema.org

#### Dataset

A Schema.org Type

Thing > CreativeWork > Dataset

A body of structured information describing some topic(s) of interest.

**Expected Type Properties from Dataset** DataDownload A downloadable form of this dataset, at a specific location, in a specific format. This property can be repeated if different variations are available. There is no expectation distribution that different downloadable distributions must contain exactly equivalent information (see also DCAT on this point). Different distributions might include or exclude different subsets of the entire dataset, for example. DataCatalog A data catalog which contains this dataset. Supersedes includedDataCatalog, catalog. includedInDataCatalog Inverse property: dataset Text The International Standard Serial Number (ISSN) that identifies this serial publication. issn You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication. DefinedTerm or A subproperty of measurementTechnique that can be used for specifying specific MeasurementMethodEnum methods, in particular via MeasurementMethodEnum. measurementMethod Text or URL DefinedTerm or A technique, method or technology used in an Observation, StatisticalVariable or MeasurementMethodEnum Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (for datasets, described using Text or variableMeasured; for Observation, a StatisticalVariable). Often but not necessarily URL each variableMeasured will have an explicit representation as (or mapping to) an property such as those defined in Schema.org, or other RDF vocabularies and "knowledge graphs". In that case the subproperty of variableMeasured called measuredProperty is applicable. The measurementTechnique property helps when extra clarification is needed about how a measuredProperty was measured. This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a measurementTechnique full representation of measurement, but can often serve as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: "mass spectrometry" or "nmr spectroscopy" or "colorimetry" or "immunofluorescence". If the variableMeasured is "depression rating", the measurementTechnique could be "Zung Scale" or "HAM-D" or "Beck Depression Inventory". If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding measurementTechnique. The value can also be from an enumeration, organized as a MeasurementMetholdEnumeration. The variableMeasured property can indicate (repeated as necessary) the variables Property or PropertyValue or that are measured in some dataset, either described as text or as pairs of identifier variableMeasured StatisticalVariable or and description using PropertyValue, or more explicitly as a StatisticalVariable. **Properties from CreativeWork** 

#### CreativeWork

A Schema.org Type

Thing > CreativeWork

[more...]

The most generic kind of creative work, including books, movies, photographs, software programs, etc.

Property	Expected Type	Description
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf
abstract	Text	An abstract is a short description that summarizes a CreativeWork.
accessMode	Text	The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Values should be drawn from the approved vocabulary.
accessModeSufficient	ItemList	A list of single or combined accessModes that are sufficient to understand all the intellectual content of a resource. Values should be drawn from the approved vocabulary.
accessibilityAPI	Text	Indicates that the resource is compatible with the referenced accessibility API. Values should be drawn from the approved vocabulary.
accessibilityControl	Text	Identifies input methods that are sufficient to fully control the described resource. Values should be drawn from the approved vocabulary.
accessibilityFeature	Text	Content features of the resource, such as accessible media, alternatives and supported enhancements for accessibility. Values should be drawn from the approved vocabulary.
accessibilityHazard	Text	A characteristic of the described resource that is physiologically dangerous to some users. Related to WCAG 2.0 guideline 2.3. Values should be drawn from the approved vocabulary.
accessibilitySummary	Text	A human-readable summary of specific accessibility features or deficiencies, consistent with the other accessibility metadata but expressing subtleties such as "short descriptions are present but long descriptions will be needed for non-visual users" or "short descriptions are present and no long descriptions are needed."
accountablePerson	Person	Specifies the Person that is legally accountable for the CreativeWork.
acquireLicensePage	CreativeWork or URL	Indicates a page documenting how licenses can be purchased or otherwise acquired, for the current item.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
alternativeHeadline	Text	A secondary title of the CreativeWork.
archivedAt	URL or WebPage	Indicates a page or other link involved in archival of a CreativeWork. In the case of MediaReview, the items in a MediaReviewItem may often become inaccessible, but be archived by archival, journalistic, activist, or law enforcement organizations. In such cases, the referenced page may not directly publish the content.
assesses	DefinedTerm or Text	The item being described is intended to assess the competency or learning outcome defined by the referenced term.
associatedMedia	MediaObject	A media object that encodes this CreativeWork. This property is a synonym for encoding.
audience	Audience	An intended audience, i.e. a group for whom something was created. Supersedes serviceAudience.
audio	AudioObject or Clip or MusicRecording	An embedded audio object.
author	Organization or Person	The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.
award	Text	An award won by or for this item. Supersedes awards.
character	Person	Fictional person connected with a creative work.
citation	CreativeWork or Text	A citation or reference to another creative work, such as another publication, web page, scholarly article, etc.
comment	Comment	Comments, typically from users.
commentCount	Integer	The number of comments this CreativeWork (e.g. Article, Question or Answer) has received. This is most applicable to works published in Web sites with commenting system; additional

### What metadata should be annotated?

#### Barebone metadata

- E.g. authors, publications, description, organism studied
- Easy to set up
- Vocabulary needed already available in schema.org
- Important information may be lost

#### All available metadata in full detail

- E.g. detailed protocols used, computational workflows with language/parameters/etc.
- Maybe overly complicated
- Specific vocabulary required (ontologies)
- Users can search for any interesting characteristics

#### Exercise

- Write JSON-LD markup for your own Investigation with at least
  - Author
  - Title
  - Description
- Test your markup in schema.org and google test tools
  - Schema Markup Validator: <a href="https://validator.schema.org/">https://validator.schema.org/</a>
  - Google Rich Result Test: <a href="https://search.google.com/test/rich-results">https://search.google.com/test/rich-results</a>