



# BUILDING A GRAPH BASED INFRASTRUCTURE

HAKEEM SHITTU

RYAN TURINSKY

# INTRODUCTIONS



Hakeem Shittu

- Principal Engineer at OCLC
- Woodwork hobbyist



Ryan Turinsky

- Lead Software Developer at OCLC
- Father of two girls
- Dallas Cowboys fan
- Marvel fan



• Located in Dublin, OH. We're a nonprofit global library organization that provides shared technology services, original research, and community programs for its membership and the library community at large.



• Worldcat is the world's most comprehensive database of information about library collections.

- 600 million bibliographic records
- 3.2 billion article records
- 1.7 billion holdings/authority records

• Catalogers, data quality specialists, and library staff members constantly enhance/enrich Worldcat records with new and corrected information

# BACKGROUND - MELLON GRANT

OCLC has been awarded a grant from The Andrew W. Mellon Foundation to develop a shared "Entity Management Infrastructure" that will support linked data management initiatives underway in the library and scholarly communications community. When complete, this infrastructure will be jointly curated by the community and OCLC, and will ultimately make scholarly materials more connected and discoverable on the web.

The two-year grant, for \$2.436 million, will support work on the project that will run from January 2020 to December 2021. The Mellon grant funding represents approximately half of the total cost of the Entity Management Infrastructure project. OCLC is contributing the remaining half of the required investment.



# LINKED DATA

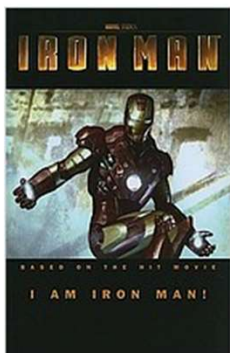
The background is a dark blue gradient with a subtle pattern of white stars. Faint, light blue circular patterns with arrows are visible in the corners, suggesting a technical or data-related theme.


[Add to list](#)

[Add tags](#)

[Write a review](#)

Rate this item:



## Iron Man. I am Iron Man

Author: [Peter David](#)

Publisher: New York : Marvel, [2010] ©2010

Edition/Format:  Print book : Fiction : English : Graphic novel [View all editions and formats](#)

Summary: When Tony Stark is wounded and taken prisoner in the Middle East, he creates a suit of armor to keep him alive and escape, and plans to right the wrongs he committed as an arms dealer by becoming the superhero Iron Man.

Rating:  (not yet rated)  0 with reviews - Be the first.

Subjects [Iron Man \(Fictitious character\) -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes.](#)

More like this

 [Similar Items](#)

### Get a Copy

[Find a copy in the library.](#)

[AbeBooks](#)

\$1.74

[Amazon](#)

\$13.97

# Following the Links

WorldCat®

Advanced Search Find a Library

<< Return to Search Results

Add to list Add tags Write a review Rate this item: ☆☆☆☆

Cite/Export Print E-mail Share Permalink

**Iron Man. I am Iron Man**

Author: [Peter David](#)

Publisher: New York : [illegible] ©2010

Edition/Format: Print book : Fiction

Summary: When Tony Stark  
alive and escape,  
Iron Man.

Rating: ☆☆☆☆ (not yet)

Subjects: [Iron Man \(Fiction\)](#)  
[Superheroes -- C](#)  
[Superheroes.](#)

More like this Similar Items

Get a Copy

[Find a copy in the library](#)

[AbeBooks](#) \$1.74

Create lists, bibliographies and reviews: [Sign in](#) or [create a free account](#)

WorldCat®

[au:David, Peter](#)

Advanced Search Find a Library

Search results for 'au:David, Peter'

Open Content

Open Access

Format

All Formats (375,376)

Article (218488)

Chapter (32705)

Downloadable article (27514)

Manuscript Article (1)

Book (55402)

Print book (39665)

eBook (14557)

Thesis/dissertation (2902)

Microform (527)

Manuscript (286)

Results 1-10 of about 375,376 (35 seconds)

Select All Clear All

Save to: [New List] Save

Sort by: Relevance Save Search

1. [Writing for comics with Peter David](#)  
by Peter David  
Print book  
Language: English  
Publisher: Cincinnati, Ohio : Impact Books, [2006] ©2006

2. [Supergirl](#)  
by Peter David; Gary Frank; Chuck Dixon; Terry Dodson; Mike Manley; Cam Smith; Karl C Story; John Nyberg; Gene D'Angelo; Megan McDowell, (Colorist); Pat Prentice; Albert de Guzman; Gaspar Saladino; Leonard Kirk; Prentis Rollins; Tom Ziuko; Patrick Martin, (Comic book artist); Robin Riggs; Digital Chameleon (Firm).

# HUMAN READABLE LINKING!





# NOT SO GREAT FOR COMPUTERS

WorldCat®


Search

Advanced Search Find a Library

<< Return to Search Results


Cite/Export Print E-mail Share Permalink

Add to list Add tags Write a review Rate this item: ☆☆☆☆☆

 **Iron Man. I am Iron Man**

Author: [Peter David](#) ?

Publisher: New York : Marvel, [2010] ©2010

Edition/Format:  Print book : Fiction : English : Graphic novel [View all editions and formats](#) ?

Summary: When Tony Stark is wounded and taken prisoner in the Middle East, he creates a suit of armor to keep him alive and escape, and plans to right the wrongs he committed as an arms dealer by becoming the superhero Iron Man.

Rating: ☆☆☆☆☆ (not yet rated) 0 with reviews - Be the first.

Subjects: [Iron Man \(Fictitious character\) -- Comic books, strips, etc. -- Fiction](#) ?  
[Superheroes -- Comic books, strips, etc. -- Fiction](#).  
[Superheroes](#).


More like this [Similar Items](#)

Get a Copy

[Find a copy in the library](#)

[AbeBooks](#) \$1.74

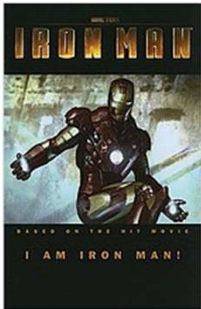




[Advanced Search](#)
[Find a Library](#)

[<< Return to Search Results](#)
Cite/Export
Print
E-mail
Share
Permalink

Add to list
Add tags
Write a review
Rate this item:
☆☆☆☆☆



### Iron Man. I am Iron Man

Author: [Peter David](#)

Publisher: New York : Marvel, [2010] ©2010

Edition/Format: Print book : Fiction : English : Graphic novel [View all editions and formats](#)

Summary: When Tony Stark is wounded and taken prisoner in the Middle East, he creates a suit of armor to keep him alive and escape, and plans to right the wrongs he committed as an arms dealer by becoming the superhero Iron Man.

Rating: ☆☆☆☆☆ (not yet rated) [0 with reviews - Be the first.](#)

Subjects: [Iron Man \(Fictitious character\) -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes.](#)

More like this [Similar Items](#)

#### Get a Copy

[Find a copy in the library](#)

[AbeBooks](#)
\$1.74

[Amazon](#)
\$13.97

Title	Iron Man: I Am Iron Man
Creator	Peter David
Publisher	Marvel
Date Published	2010

Title	Iron Man: I Am Iron Man
Creator	<a href="#">Peter David</a>
Publisher	<a href="#">Marvel</a>
Date Published	2010

name	Peter David
Is a	Person
Date of birth	9/23/1956

name	Marvel
Is a	Business
Inception	1939

Subject

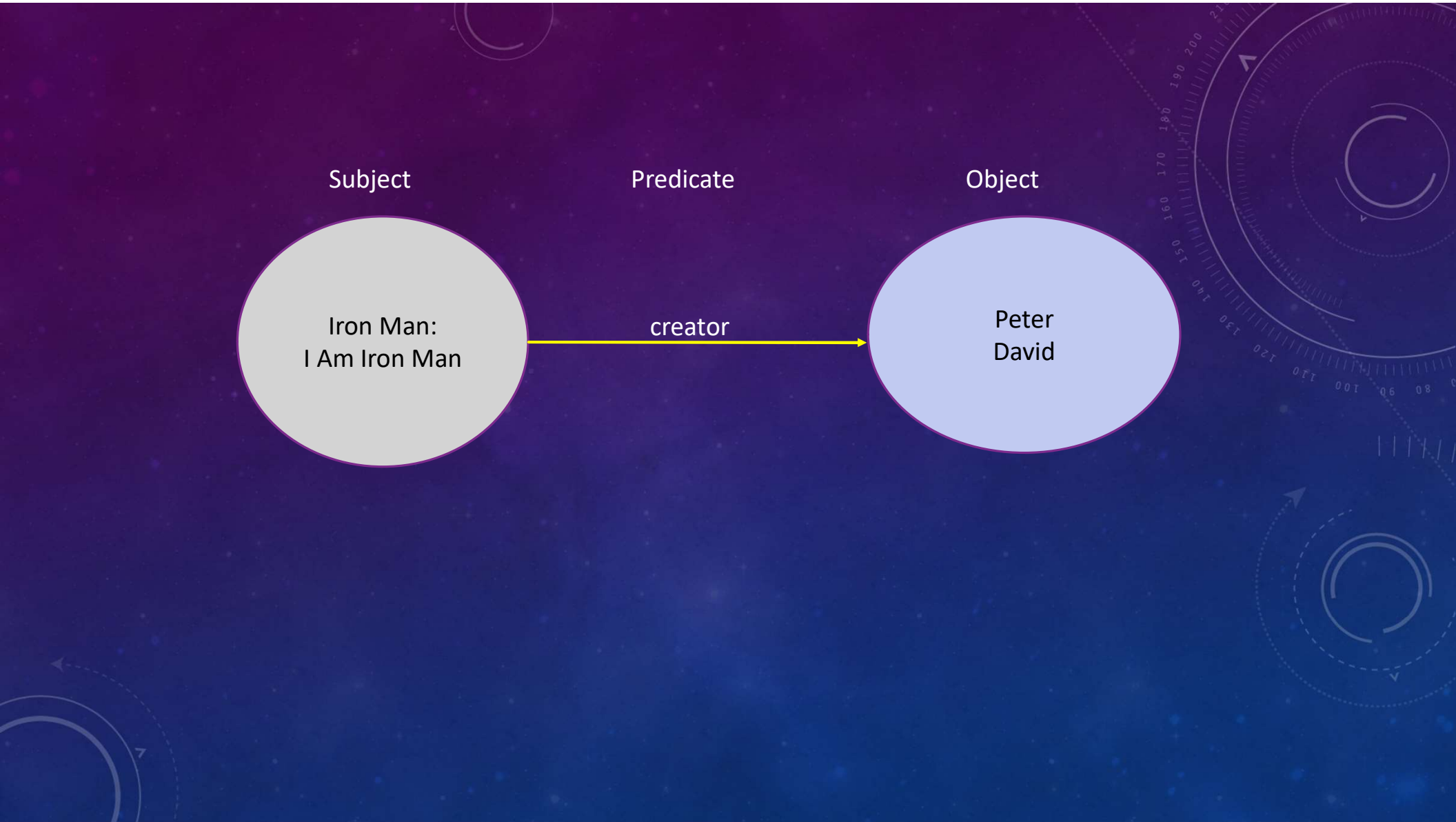
Predicate

Object

Iron Man:  
I Am Iron Man

creator

Peter  
David





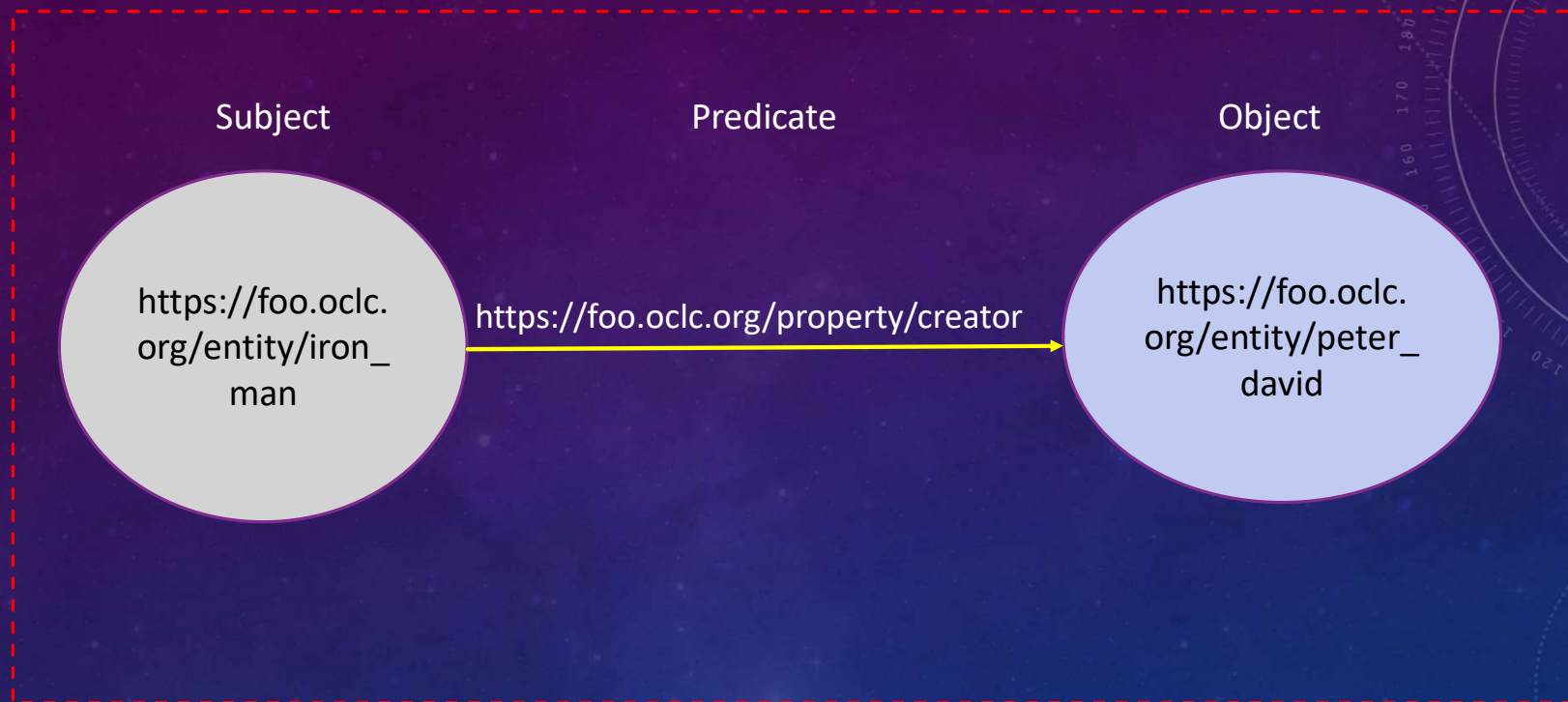
# NEED TO BE SPECIFIC

Iron Man:  
I Am  
Iron Man

creator

Peter  
David





RDF

# RDF FORMATS

- N-Triples

```
<https://foo.oclc.org/entity/123> <http://schema.org/name> "Peter David"@en .  
<https://foo.oclc.org/entity/123> <https://foo.oclc.org/property/author> <https://foo.oclc.org/entity/789> .
```

- JSON-LD

```
{  
  "@context": "http://schema.org/",  
  "type": "Person",  
  "birthDate": "09/23/1956",  
  "jobTitle": "Writer",  
  "name": "Peter David"  
}
```

- N-Quads

```
_:b0 <http://schema.org/birthDate> "09/23/1956"^^<http://schema.org/Date> .  
_:b0 <http://schema.org/jobTitle> "Writer" .  
_:b0 <http://schema.org/name> "Peter David" .  
_:b0 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://schema.org/Person> .
```

# RECAP – LINKED DATA

- Use URIs as names for things
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information, using the standards (RDF)
- Include links to other URIs, so that they can discover more things

<https://www.w3.org/DesignIssues/LinkedData.html>



# ONTOLOGY

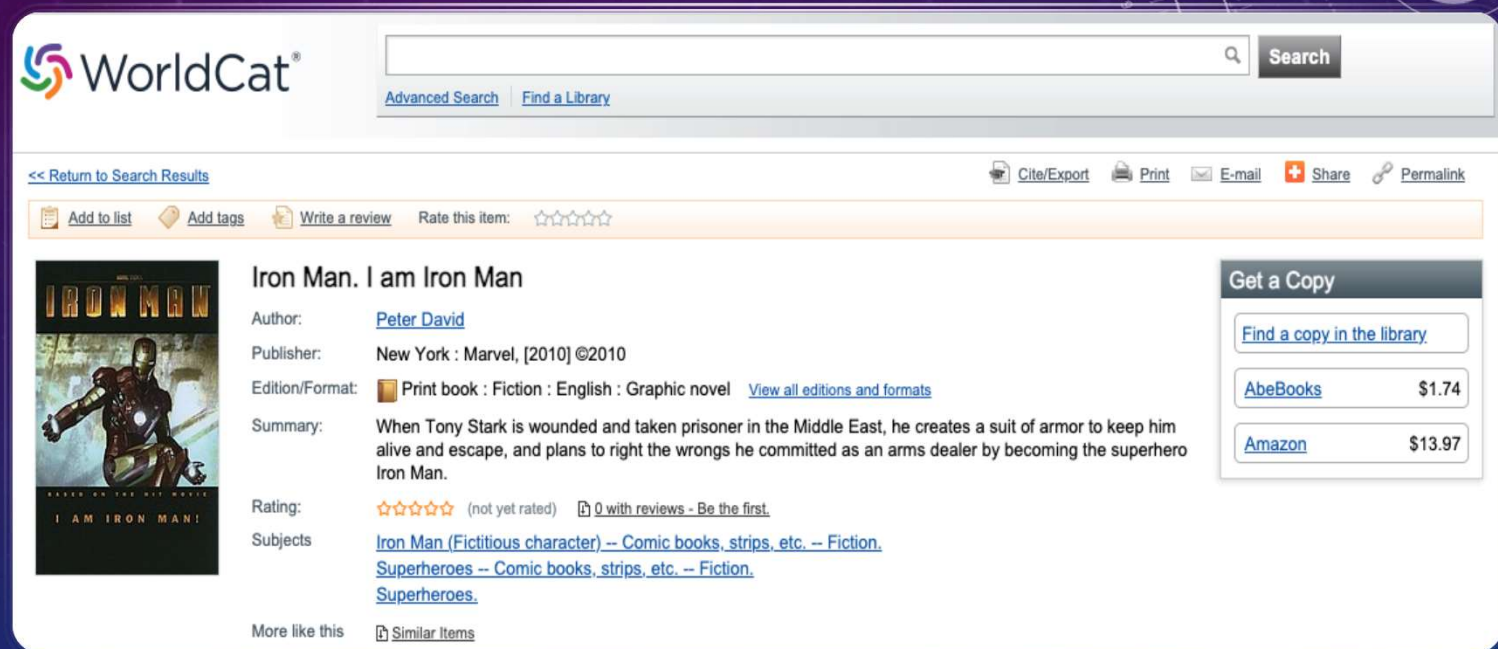


# WHAT ARE ONTOLOGIES?

- Ontologies are semantic data models that define the types of things that exist in a domain and the properties that can be used to describe them
- Generalized data models
- Main components consist of
  - Classes - (Nodes)
  - Relationships - (Edges)
    - Decide on directionality of relationships.
  - Attributes –
    - Required, mandatory and optional attributes
    - Data types - How to describe concepts like DateTime may prove more challenging than expected
- Examples:
  - [Schema.org](http://Schema.org)
  - [Dublin Core](http://Dublin Core)
  - [Friend of a Friend \(foaf\) Ontology](http://Friend of a Friend (foaf) Ontology)
- **An approach to defining an ontology for your knowledge graph**
  - **Start from existing ontologies** – this allows you to use industry standard terms and relationships
  - **Adapt to suit your needs** – extend for the specific nuances of your domain

# BUILDING AN ONTOLOGY - CLASSES

- Books
- Authors
- Publishers
- Locations
- Subjects



The screenshot displays the WorldCat website interface. At the top, the WorldCat logo is on the left, and a search bar with a 'Search' button is on the right. Below the search bar are links for 'Advanced Search' and 'Find a Library'. The main content area shows the book record for 'Iron Man. I am Iron Man' by Peter David. The record includes a book cover image, the title, author, publisher (New York : Marvel, [2010] ©2010), edition/format (Print book : Fiction : English : Graphic novel), a summary, a rating (0 stars, not yet rated), and subjects (Iron Man (Fictitious character) -- Comic books, strips, etc. -- Fiction; Superheroes -- Comic books, strips, etc. -- Fiction; Superheroes). On the right side, there is a 'Get a Copy' section with a link to 'Find a copy in the library' and two purchase options: AbeBooks for \$1.74 and Amazon for \$13.97. At the bottom, there are links for 'More like this' and 'Similar Items'.

WorldCat®

Advanced Search Find a Library

<< Return to Search Results


Cite/Export Print E-mail Share Permalink

Add to list Add tags Write a review Rate this item: ☆☆☆☆☆

**Iron Man. I am Iron Man**

Author: [Peter David](#)

Publisher: New York : Marvel, [2010] ©2010

Edition/Format:  Print book : Fiction : English : Graphic novel [View all editions and formats](#)

Summary: When Tony Stark is wounded and taken prisoner in the Middle East, he creates a suit of armor to keep him alive and escape, and plans to right the wrongs he committed as an arms dealer by becoming the superhero Iron Man.

Rating: ☆☆☆☆☆ (not yet rated) [0 with reviews - Be the first.](#)

Subjects: [Iron Man \(Fictitious character\) -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes.](#)

More like this [Similar Items](#)

**Get a Copy**

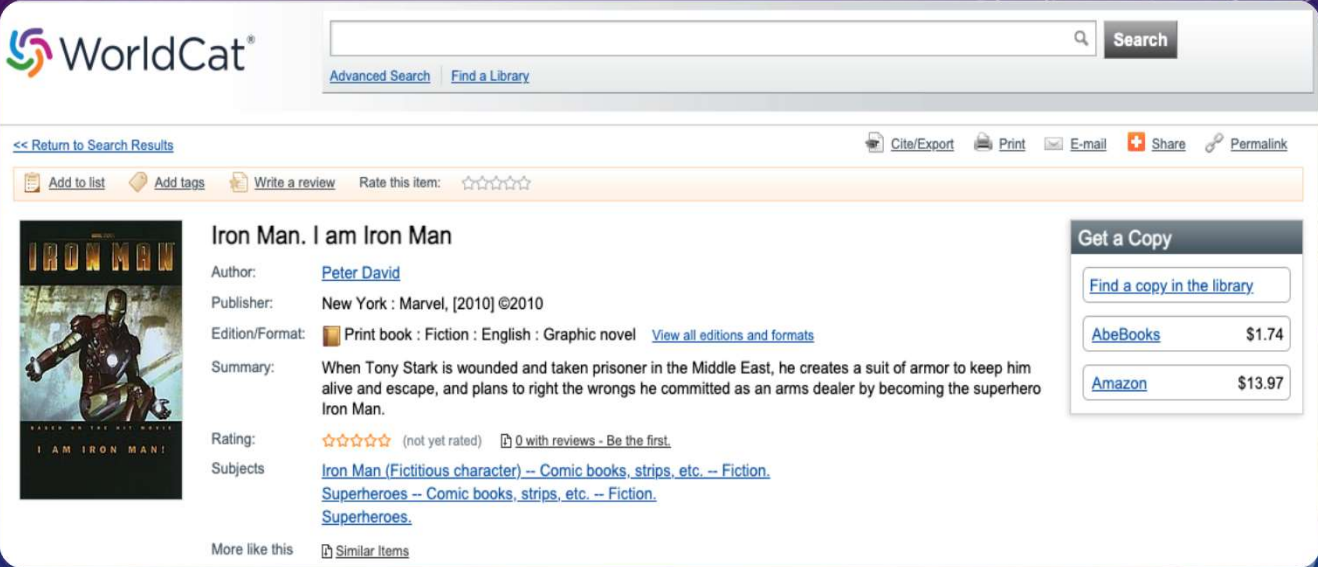
[Find a copy in the library.](#)

[AbeBooks](#) \$1.74

[Amazon](#) \$13.97

# ONTOLOGY – RELATIONSHIPS AND ATTRIBUTES

- Books are created by Authors
- Books are published by Publishers
- Books have subjects
- Books are published on date
- Books have summaries



The screenshot displays the WorldCat website interface. At the top, the WorldCat logo is on the left, and a search bar with a 'Search' button is on the right. Below the search bar are links for 'Advanced Search' and 'Find a Library'. A navigation bar contains links like '<< Return to Search Results', 'Cite/Export', 'Print', 'E-mail', 'Share', and 'Permalink'. Below this is a row of icons for 'Add to list', 'Add tags', 'Write a review', and 'Rate this item:'. The main content area features a book cover for 'Iron Man. I am Iron Man' on the left. To the right of the cover, the title 'Iron Man. I am Iron Man' is displayed, followed by the author 'Peter David', publisher 'New York : Marvel, [2010] ©2010', and edition/format 'Print book : Fiction : English : Graphic novel'. A summary describes Tony Stark's story. The rating section shows '0 with reviews - Be the first.' and subjects include 'Iron Man (Fictitious character) -- Comic books, strips, etc. -- Fiction.' and 'Superheroes -- Comic books, strips, etc. -- Fiction.' A 'Get a Copy' sidebar on the right offers 'Find a copy in the library', 'AbeBooks' at \$1.74, and 'Amazon' at \$13.97. At the bottom, there are links for 'More like this' and 'Similar items'.

WorldCat®

Advanced Search Find a Library

<< Return to Search Results

Cite/Export Print E-mail Share Permalink

Add to list Add tags Write a review Rate this item: ☆☆☆☆☆

**Iron Man. I am Iron Man**

Author: [Peter David](#)

Publisher: New York : Marvel, [2010] ©2010

Edition/Format: Print book : Fiction : English : Graphic novel [View all editions and formats](#)

Summary: When Tony Stark is wounded and taken prisoner in the Middle East, he creates a suit of armor to keep him alive and escape, and plans to right the wrongs he committed as an arms dealer by becoming the superhero Iron Man.

Rating: ☆☆☆☆☆ (not yet rated) [0 with reviews - Be the first.](#)

Subjects: [Iron Man \(Fictitious character\) -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes -- Comic books, strips, etc. -- Fiction.](#)  
[Superheroes.](#)

More like this [Similar items](#)

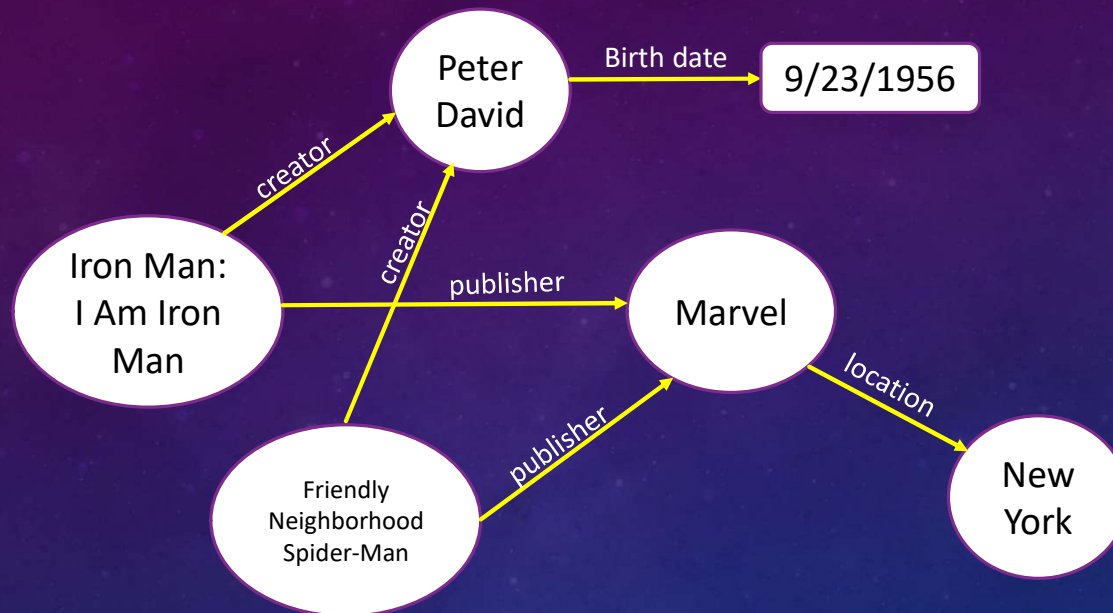
**Get a Copy**

[Find a copy in the library](#)

[AbeBooks](#) \$1.74

[Amazon](#) \$13.97





- Provides ability to answer context specific questions:
  - List of all works created by Peter David and published by Marvel
  - The names of all creators whose works were published by Marvel

# KNOWLEDGE GRAPHS

The background is a gradient of dark blue and purple, resembling a starry night sky. It features several faint, stylized circular patterns and arcs, some with tick marks and arrows, suggesting a technical or scientific theme. The text 'KNOWLEDGE GRAPHS' is centered in a white, sans-serif font.

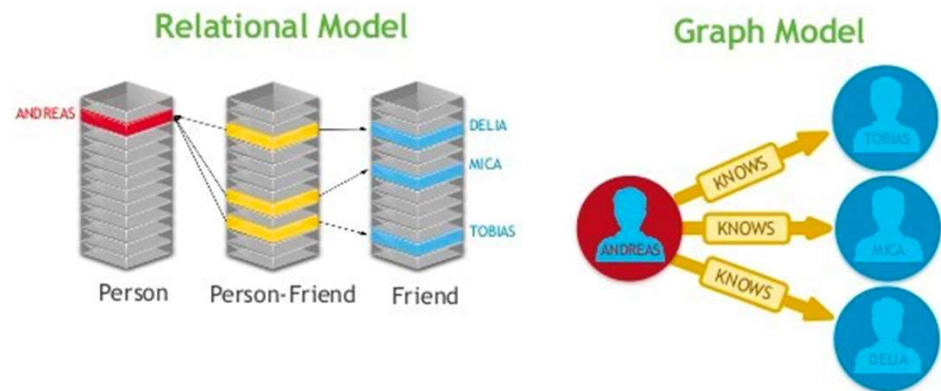
# WHAT IS A KNOWLEDGE GRAPH?

- Represents a network of real-world entities and illustrates the relationship between them
- "Things, not strings" - Amit Singhal
- Consists of data stored in a graph database, with a schema derived from your ontology
- Data is access via a query language
  - [SPARQL](#)
  - [Gremlin](#)
  - [openCypher](#)
- Examples:
  - [Wikidata](#)
  - [Google Knowledge Graph](#)
  - [Facebook Open Graph](#)

# GRAPH DATABASE

- What is this?
- Models
  - Labeled Property Graph (LPG)
  - Resource Description Framework (RDF)
- Query Language
  - OpenCypher (LPG)
  - Gremlin (LPG)
  - SPARQL (RDF)

## Relational Versus Graph Models





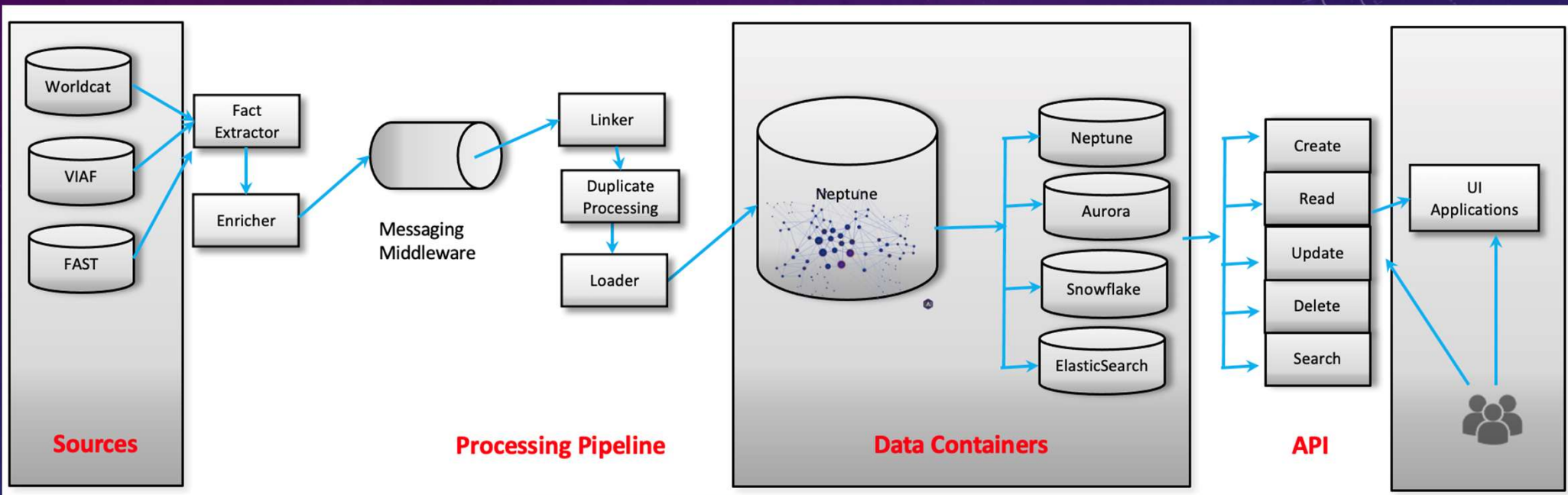
# NEPTUNE

- **AWS Graph DB as a Service**
- **Support for both LPG and RDF**
  - Only 1 mode supported via each instance of the graph
- **Support for up to 100 billion nodes + edges provides needed capacity for scaling**
- Multiple Query Language
  - OpenCypher (LPG)
  - Gremlin (LPG)
  - SPARQL (RDF)
- Sub second performance with complex queries
  - 100k concurrent query capacity yet to be tested

# STRINGS-TO -THINGS PROCESS

- Step 1: Identify your use case, define your vocabulary, create your ontology
- Step 2: Define your tech stack
- Step 3: Decant your initial data, build your update flows
- Step 4: Materialize your views
- Step 5: Build your APIs
- Step 6: Build your UI

# PROCESSING PIPELINE



# INFRASTRUCTURE STACK

- We have an AWS based environment and opted for managed service components.
- Processing engine – Flink, Spark, Storm,
- Messaging Pipeline – Kinesis, Kafka, RabbitMQ
- Graph Database - Neptune (LPG), Neo4j, Stardog, OrientDB
- Data Access Containers -
  - Elastic search
  - Neptune (RDF)
  - Aurora
  - Snowflake
- User Access
  - API – Springboot,
  - UI – React, NextJS



## SELECT Entity Based On Label and Language

sparql?query=PREFIX rdfs:

<http://www.w3.org/2000/01/rdf-schema%23> SELECT  
?s ?p ?o WHERE { ?s rdfs:label "Iron Man"@en . }

```
1 {
2   "head": { "s": 1 },
9   "results": {
10    "bindings": [
11      {
12        "s": {
13          "type": "uri",
14          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGjgvXMYdJ9KyRjv49TDy"
15        },
16      },
17      {
18        "s": {
19          "type": "uri",
20          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGTtMK43VHTJjvcBjFmh3"
21        },
22      },
23      {
24        "s": {
25          "type": "uri",
26          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCFDQFwqVGB3bwCK9TCTq43"
27        },
28      },
29      {
30        "s": {
31          "type": "uri",
32          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGhtfgTpMFvQgg6T3MarYP"
33        },
34      },
35      {
36        "s": {
37          "type": "uri",
38          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCFVBw8frKfbY76xDhtYfg"
39        },
40      },
41      {
42        "s": {
43          "type": "uri",
44          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCXrqbbp7dfyfYtwRFghmBP"
45        },
46      },
47      {
48        "s": {
49          "type": "uri",
50          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCH8947KPqPttjPKXD4gfwP"
51        },
52      },
53      {
54        "s": {
55          "type": "uri",
56          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGb6wxBRRJQQRrWkfwcr4g"
57        },
58      },
59      {
60        "s": {
61          "type": "uri",
62          "value": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGm7tj86ydFwHy7m7XgRJc"
63        },
64      },
65    ]
66  }
67 }
```

# API – SPARQL ENDPOINT

```

1  {
2  {
3  "@context": { "@ 1653 " },
5680 "@graph": [
5681 {
5682   "id": "https://entity-viewer-play.dev.oclc.org/entity/E39PCGCYJBwxWcxp34VRRxvtmm",
5683   "description": {
5684     "en": "Graphic novels associated with Peter David about Iron Man (Fictitious character)"
5685   },
5686   "label": {
5687     "en": "Iron Man I am Iron Man"
5688   },
5689   "altLabel": {
5690     "en": [
5691       "I am Iron Man",
5692       "Iron Man"
5693     ]
5694   },
5695   "wdt:P142": "10141687024#en",
5696   "wdt:P15": [
5697     "\"I am Iron Man Iron Man\" (en)"
5698   ],
5699   "wdt:P233": [
5700     "2010-1-1T00:00:00Z"
5701   ],
5702   "wdt:P33": [
5703     "wd:Q141"
5704   ],
5705   "wdt:P40": [
5706     "wd:E39PBjqKdTq7PWXrcJChjM9VYP"
5707   ],
5708   "wdt:P5": [
5709     "wd:Q10"
5710   ],
5711   "wdt:P97": [
5712     "1144656476"
5713   ]
5714 }
5715 ]
5716 }
5717 ]

```

API – READ ENDPOINT

```

1- {
2-   {
3-     "entity": {
4-       "id": "E39PCGCYJBwxWcxp34VRRxvtmm",
5-       "labels": {
6-         "en": { = 2 = }
7-       },
8-       "descriptions": {
9-         "en": { = 2 = }
10-      },
11-       "aliases": {
12-         "en": [ = 2 = ]
13-       },
14-       "claims": {
15-         "P33": [
16-           { = 5 = }
17-         ],
18-         "P5": [
19-           { = 5 = }
20-         ],
21-         "P15": [
22-           { = 5 = }
23-         ],
24-         "P233": [
25-           { = 5 = }
26-         ],
27-         "P142": [
28-           { = 5 = }
29-         ],
30-         "P40": [
31-           { = 5 = }
32-         ],
33-         "P97": [
34-           { = 5 = }
35-         ]
36-       },
37-       "sitelinks": {},
38-       "lastrevid": 88410882
39-     },
40-     "scoreComposite": {
41-       "composite": 47.0,
42-       "completeness": {
43-         "weightValue": 25.0,
44-         "weightedScore": 22.0
45-       },
46-       "confidence": {
47-         "weightValue": 25.0,
48-         "weightedScore": 25.0
49-       },
50-       "disambiguation": {
51-         "weightValue": 50.0,
52-         "weightedScore": 0.0
53-       }
54-     },
55-     "superType": "work",
56-     "entityStatus": {
57-       "status": "active",
58-       "rerouteid": null
59-     }
60-   }
61- }

```

# API – READ ENDPOINT

[Back to search](#)

Top

Key Information

Entity Subtypes

Significant Dates

Characteristics

External Sources

Related Persons

ViewEdit

Actions

47

Iron Man I am Iron Man

Copy URI

Entity ID: E39PCGCYJBwxWcxp34VRRxvtmm

Entity Type: Work

Description: Graphic novels associated with Peter David about Iron Man (Fictitious character)

Key Information

Language of Content	Entity Label	Description	Aliases
English	Iron Man I am Iron Man	Graphic novels associated with Peter David about Iron Man (...	2

Entity Subtypes

Instance Of	Rank
text	

Related Persons

Peter David

creator

2 references



Top

Key Information

Entity Subtypes

Significant Dates

External Sources

ViewEdit

Actions

67

Peter David

Copy URI

Entity ID: E39PBJqKdTq7PWXrcJChjM9VYP

Entity Type: Person

Description: American writer of comic books, novels, television, movies and video games

External Sources		6-10 of 22
Source Type	Source ID	
GND ID	<a href="#">129101486</a>	
GRATEVE ID	<a href="#">73593</a>	
IMDb ID	<a href="#">nm0203035</a>	
ISNI	<a href="#">0000 0001 0802 9148</a>	
J9U ID	987007404459205171	

# CONSIDERATIONS

- Quality
  - Issue: How can I verify the correctness of claims
  - Resolution Approach: Provenance, Trusted sources
- Issues: Allowing competing claims into the graph
- Example: Multiple dates of birth for a single individual
- Resolution: Quality scores are provided for each entity reflecting how much they adhere to the ontology constraints
- Issue: Prevent/Identify duplicate entities
- Resolution Approach: Done carefully with different rules in place for each entity type. ML tools (Rosette) useful here

# CONSIDERATIONS (CONTD)

- Scale
  - Issue: Query performance is impacted as graph size increases
  - Resolution: We analyzed our uses cases and moved certain responses to materialized view that may be more suitable. Lucene allows quick text search, while Snowflake provides analytics capabilities moving these needs off Neptune
- Linking
  - Issue: How do you determine a record is connected to the right "Peter David"
  - Resolution Approach: LPG references metadata for linking
- Bi-directional linking
  - Issue: For certain reflexive attributes (like siblings) updating a node may require updating all related nodes

# REFERENCES

- [https://www.youtube.com/watch?v=4x\\_xzT5eF5Q&ab\\_channel=ManuSporny](https://www.youtube.com/watch?v=4x_xzT5eF5Q&ab_channel=ManuSporny)
- <https://medium.com/wallscope/understanding-linked-data-formats-rdf-xml-vs-turtle-vs-n-triples-eb931dbe9827>
- <https://enterprise-knowledge.com/whats-the-difference-between-an-ontology-and-a-knowledge-graph/>
- <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/#:~:text=Linked%20Data%20is%20a%20set,is%20an%20example%20of%20LOD>
- <https://blog.devgenius.io/graph-database-vs-relational-database-70f6156f7415>





# THANK YOU

- What is known has been shared...(and linked)