



# One Superhero vs. Team of Heroes

Professionalization, decentralization and cooperation  
in investigative information science

Ohad Zohar, Ph.D.

CEO & Founder

Omnisol Information Systems



11/5/2017



# Introduction

Investigative information science is at a tipping point:

- Web scraping tools
- The semantic-web
- Advanced data processing
- Crowd sourcing platforms
- Visualization tools

# Introduction



Information specialists can now accomplish what only a few years ago only big intelligence companies could. But they need to be “superheroes”.

Alternatively, a carefully managed team of diverse specialists could achieve the same.

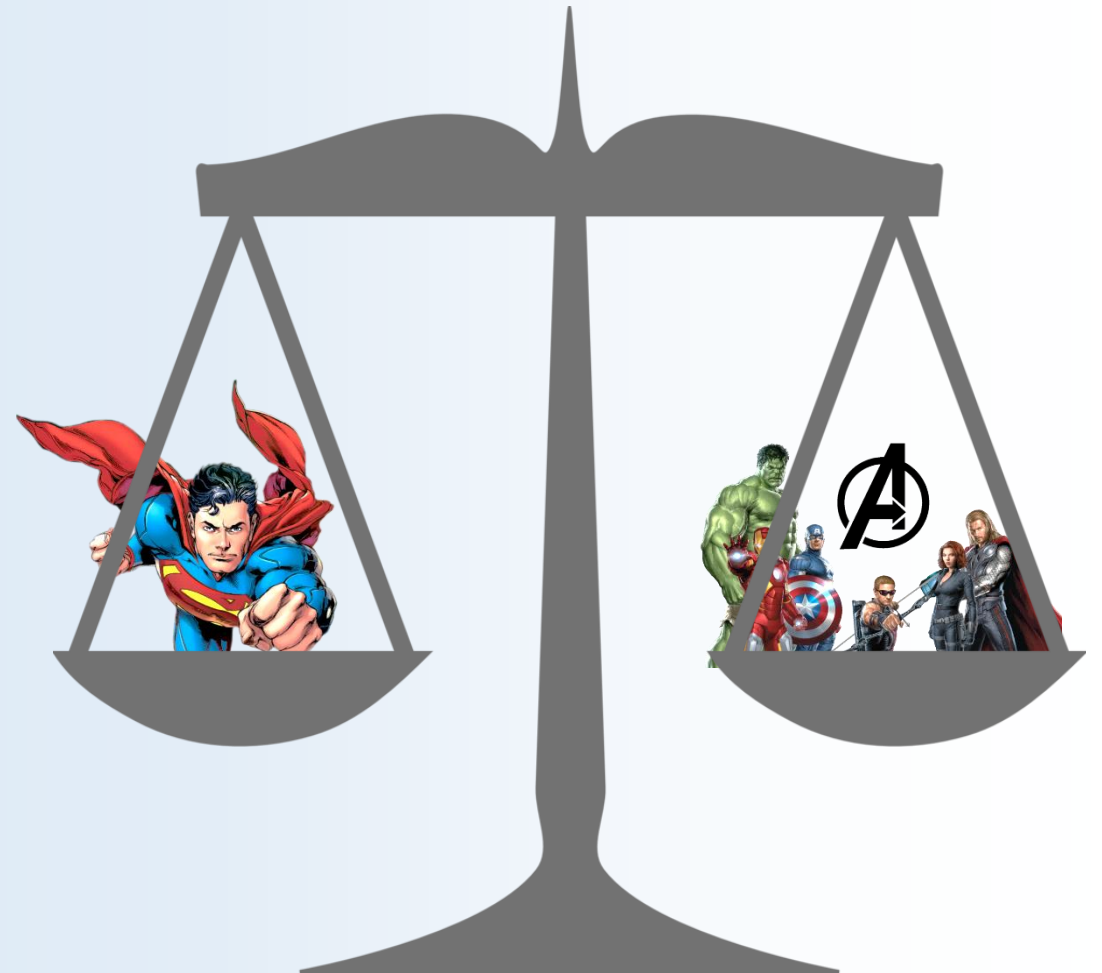


# Heroic Investigation

What can we say about **superhero movies** vs. a **team of heroes** movies?

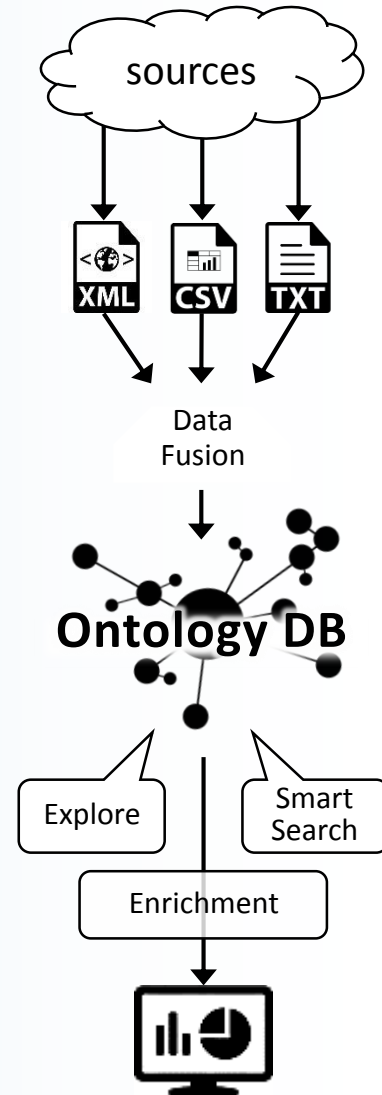
- Which type is more popular?
- Which is made more?
- Which earns more money?

What else is interesting to figure out?



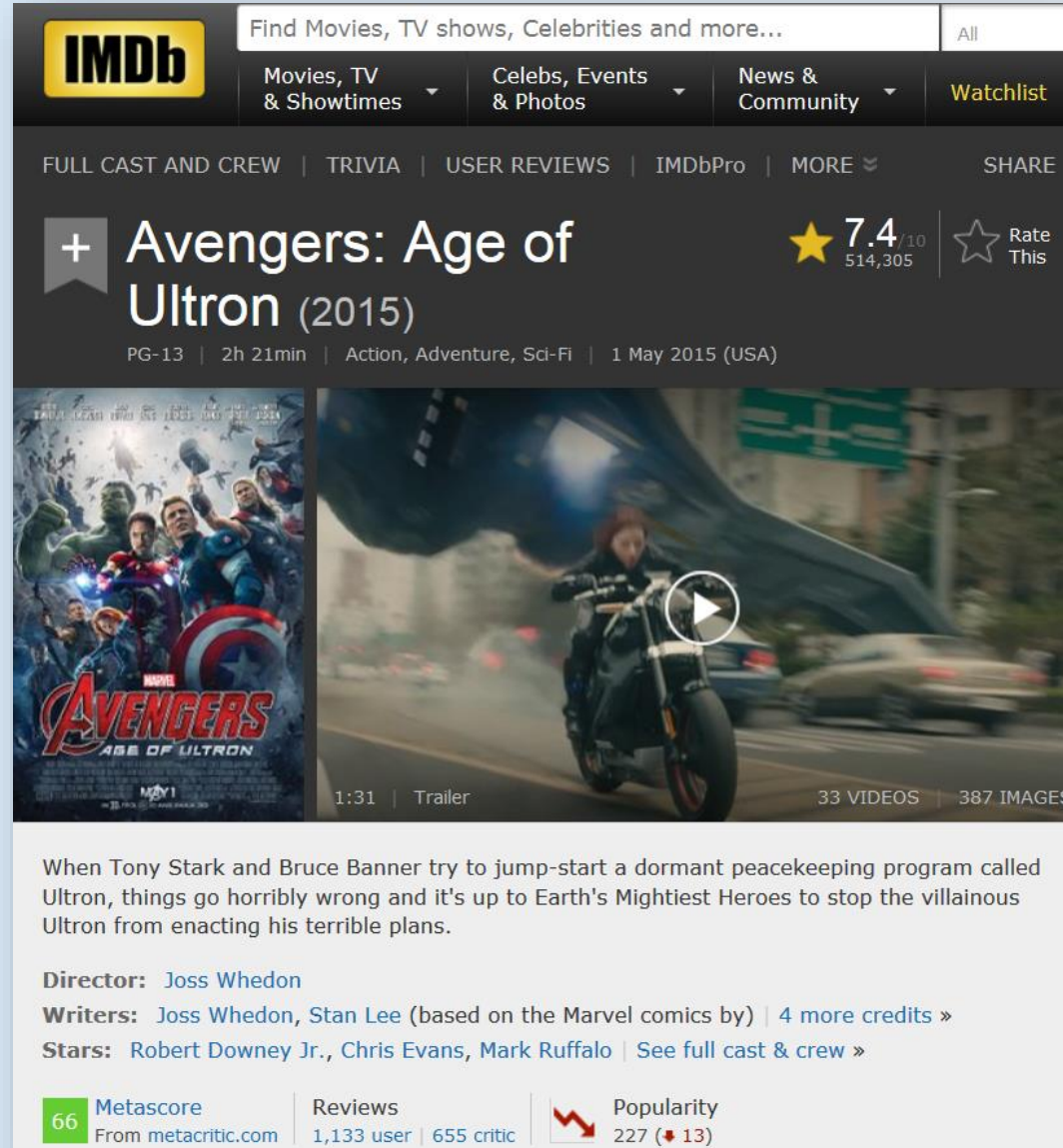
# Methodology

1. Identify sources
2. Extract relevant data
3. Build ontology and load in the data
4. Explore the data
5. Process the data
6. Identify missing data
7. Generate infographics



# IMDb – Internet Movie Database

<http://www.imdb.com>



The screenshot shows the IMDb page for the movie "Avengers: Age of Ultron (2015)". The page includes the IMDb logo, a search bar, and navigation tabs for Movies, TV & Showtimes, Celebs, Events & Photos, News & Community, and Watchlist. The movie's title is prominently displayed with a rating of 7.4/10 and 514,305 votes. Below the title, it lists the PG-13 rating, runtime (2h 21min), genres (Action, Adventure, Sci-Fi), and release date (1 May 2015 (USA)). A movie poster and a trailer thumbnail are shown. The synopsis describes the plot where Tony Stark and Bruce Banner attempt to start a peacekeeping program called Ultron, which goes wrong. Credits for Director (Joss Whedon), Writers (Joss Whedon, Stan Lee), and Stars (Robert Downey Jr., Chris Evans, Mark Ruffalo) are listed. At the bottom, there are sections for Metascore (66), Reviews (1,133 user, 655 critic), and Popularity (227, 13).

IMDb Find Movies, TV shows, Celebrities and more... All

Movies, TV & Showtimes Celebs, Events & Photos News & Community Watchlist

FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE | SHARE

**Avengers: Age of Ultron (2015)** ★ 7.4/10 514,305 Rate This

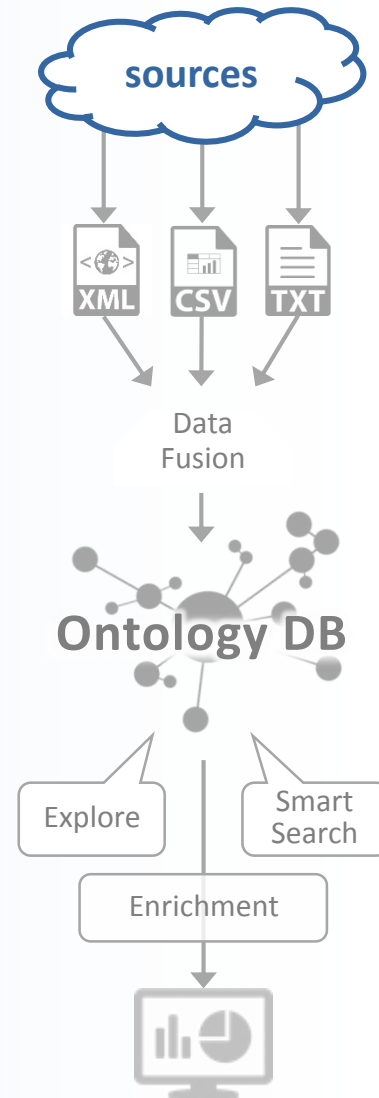
PG-13 | 2h 21min | Action, Adventure, Sci-Fi | 1 May 2015 (USA)

1:31 | Trailer 33 VIDEOS | 387 IMAGES

When Tony Stark and Bruce Banner try to jump-start a dormant peacekeeping program called Ultron, things go horribly wrong and it's up to Earth's Mightiest Heroes to stop the villainous Ultron from enacting his terrible plans.

**Director:** Joss Whedon  
**Writers:** Joss Whedon, Stan Lee (based on the Marvel comics by) | 4 more credits »  
**Stars:** Robert Downey Jr., Chris Evans, Mark Ruffalo | See full cast & crew »

66 Metascore From metacritic.com  
Reviews 1,133 user | 655 critic  
Popularity 227 (13)





# RottenTomatoes

<https://www.rottentomatoes.com>



**Rotten Tomatoes** What's the Tomatometer®? RT Critics

Search movies, TV, actors, more...

MOVIES & DVDS ▾ TV ▾ NEWS ▾ TICKETS & SHOWTIMES

TRENDING ON RT Guardians of the Galaxy Reviews Dunkirk Trailer Summer Movie Calendar Groot's Five Favorite Films FOLLOW US [Facebook] [Twitter] [Instagram] [Snapchat] [YouTube]

HOME > MARVEL CINEMATIC UNIVERSE > AVENGERS: AGE OF ULTRON

IN THEATERS

DVD & STREAMING

TV SHOWS

**TOP DVD & STREAMING**

- 6% Rings
- 31% A Dog's Purpose
- 99% I Am Not Your Negro
- 42% Gold
- 97% The Salesman (Forusha...

View All >

**NEW ON DVD THIS WEEK**

- 9% Fifty Shades Darker
- 100% Things to Come (L'avenir)
- 33% A Fantastic Fear of Ever...
- 77% A Street Cat Named Bob
- 100% Hounds of Love

View All >

**AVENGERS: AGE OF ULTRON** (2015)

Part of the Collection: *Marvel Cinematic Universe* View Collection >

**TOMATOMETER** ?

**75%**

Average Rating: 6.7/10  
Reviews Counted: 312  
Fresh: 233  
Rotten: 79

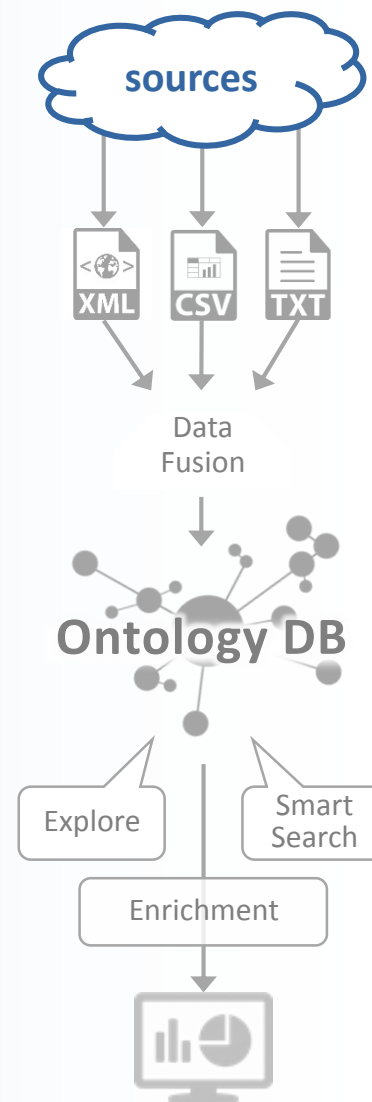
**All Critics | Top Critics**

Critics Consensus: Exuberant and eye-popping, *Avengers: Age of Ultron* serves as an overstuffed but mostly satisfying sequel, reuniting its predecessor's unwieldy cast with a few new additions and a worthy foe.

**AUDIENCE SCORE** ?

**83%**  
liked it

Average Rating: 4.1/5  
User Ratings: 283,508



# Movie Lists

<http://www.listchallenges.com/every-superhero-movie>

## Every Superhero Movie

Every superhero ever made, except for straight to dvd, tv movies, and old serials. Full length, theatrical releases only. Also sticking to American/Canadian complicated to try to figure out all the foreign superhero films.

Added by [Jay\\_man](#)

Avg. score: 55 of 90 (61%)

9,870 users · 22,150 views

★ Required scores: 1, 39, 51, 61, 71

[Share](#) [Pin](#) [Share](#) [Stumble](#)

[Checklist](#)

[Stats](#)

[Leaders](#)

[Vote](#)



How many have you seen?



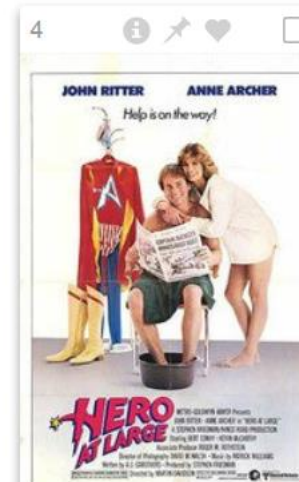
Superman and the Mole Men (1951)



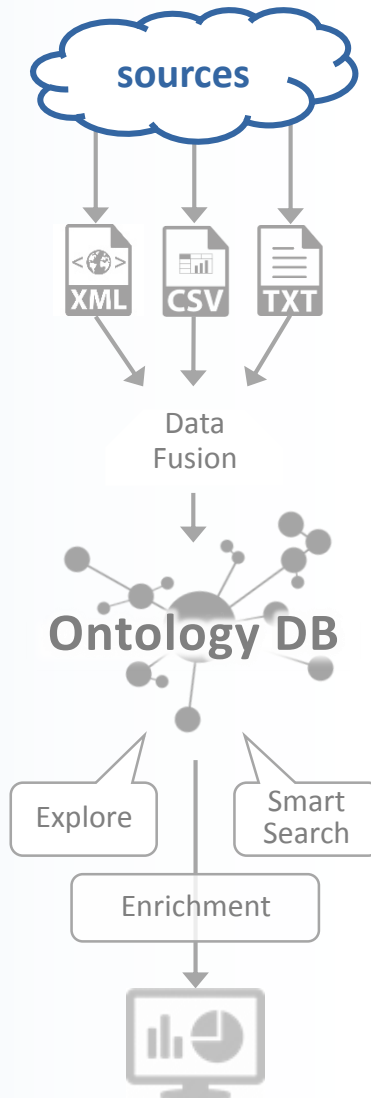
Batman (1966)



Superman (1978)

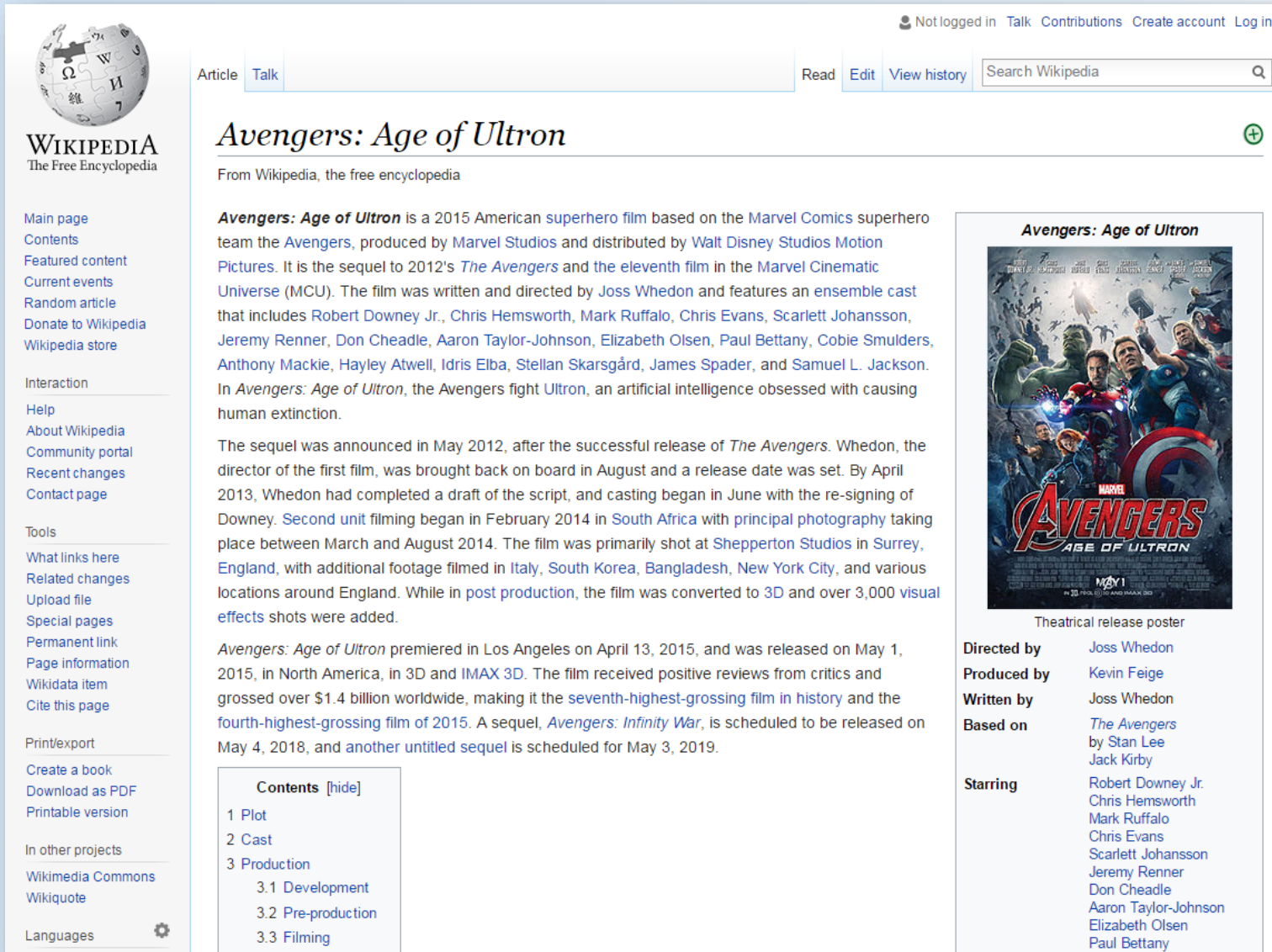


Hero at Large (1980)





# Wikipedia and DBpedia



The screenshot shows the Wikipedia article for "Avengers: Age of Ultron". The page includes the Wikipedia logo, navigation links, and the article text. The article describes the film as a 2015 American superhero film based on the Marvel Comics superhero team the Avengers, produced by Marvel Studios and distributed by Walt Disney Studios Motion Pictures. It is the sequel to 2012's *The Avengers* and the eleventh film in the Marvel Cinematic Universe (MCU). The film was written and directed by Joss Whedon and features an ensemble cast that includes Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, Scarlett Johansson, Jeremy Renner, Don Cheadle, Aaron Taylor-Johnson, Elizabeth Olsen, Paul Bettany, Cobie Smulders, Anthony Mackie, Hayley Atwell, Idris Elba, Stellan Skarsgård, James Spader, and Samuel L. Jackson. In *Avengers: Age of Ultron*, the Avengers fight Ultron, an artificial intelligence obsessed with causing human extinction.

The sequel was announced in May 2012, after the successful release of *The Avengers*. Whedon, the director of the first film, was brought back on board in August and a release date was set. By April 2013, Whedon had completed a draft of the script, and casting began in June with the re-signing of Downey. *Second unit* filming began in February 2014 in South Africa with principal photography taking place between March and August 2014. The film was primarily shot at Shepperton Studios in Surrey, England, with additional footage filmed in Italy, South Korea, Bangladesh, New York City, and various locations around England. While in *post production*, the film was converted to 3D and over 3,000 visual effects shots were added.

*Avengers: Age of Ultron* premiered in Los Angeles on April 13, 2015, and was released on May 1, 2015, in North America, in 3D and IMAX 3D. The film received positive reviews from critics and grossed over \$1.4 billion worldwide, making it the seventh-highest-grossing film in history and the fourth-highest-grossing film of 2015. A sequel, *Avengers: Infinity War*, is scheduled to be released on May 4, 2018, and another untitled sequel is scheduled for May 3, 2019.

**Contents** [hide]

- 1 Plot
- 2 Cast
- 3 Production
  - 3.1 Development
  - 3.2 Pre-production
  - 3.3 Filming

**Avengers: Age of Ultron**

From Wikipedia, the free encyclopedia

*Avengers: Age of Ultron* is a 2015 American superhero film based on the Marvel Comics superhero team the Avengers, produced by Marvel Studios and distributed by Walt Disney Studios Motion Pictures. It is the sequel to 2012's *The Avengers* and the eleventh film in the Marvel Cinematic Universe (MCU). The film was written and directed by Joss Whedon and features an ensemble cast that includes Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, Scarlett Johansson, Jeremy Renner, Don Cheadle, Aaron Taylor-Johnson, Elizabeth Olsen, Paul Bettany, Cobie Smulders, Anthony Mackie, Hayley Atwell, Idris Elba, Stellan Skarsgård, James Spader, and Samuel L. Jackson. In *Avengers: Age of Ultron*, the Avengers fight Ultron, an artificial intelligence obsessed with causing human extinction.

The sequel was announced in May 2012, after the successful release of *The Avengers*. Whedon, the director of the first film, was brought back on board in August and a release date was set. By April 2013, Whedon had completed a draft of the script, and casting began in June with the re-signing of Downey. *Second unit* filming began in February 2014 in South Africa with principal photography taking place between March and August 2014. The film was primarily shot at Shepperton Studios in Surrey, England, with additional footage filmed in Italy, South Korea, Bangladesh, New York City, and various locations around England. While in *post production*, the film was converted to 3D and over 3,000 visual effects shots were added.

*Avengers: Age of Ultron* premiered in Los Angeles on April 13, 2015, and was released on May 1, 2015, in North America, in 3D and IMAX 3D. The film received positive reviews from critics and grossed over \$1.4 billion worldwide, making it the seventh-highest-grossing film in history and the fourth-highest-grossing film of 2015. A sequel, *Avengers: Infinity War*, is scheduled to be released on May 4, 2018, and another untitled sequel is scheduled for May 3, 2019.

**Avengers: Age of Ultron**

Theatrical release poster

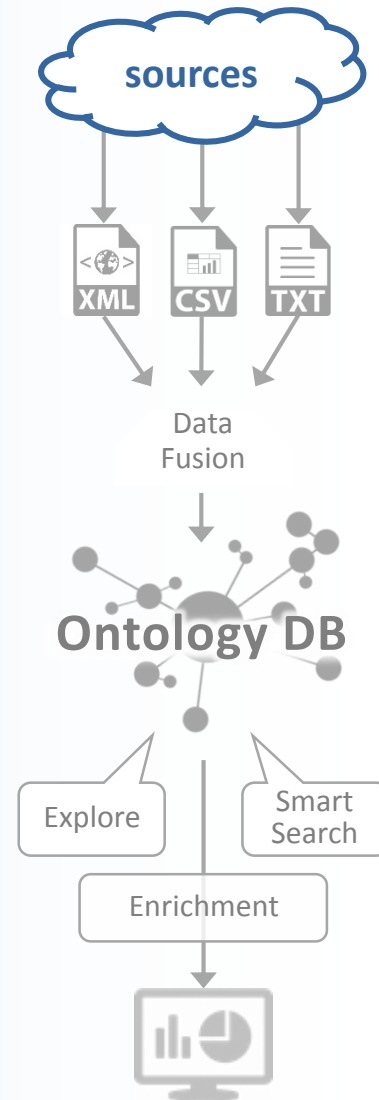
**Directed by** Joss Whedon

**Produced by** Kevin Feige

**Written by** Joss Whedon

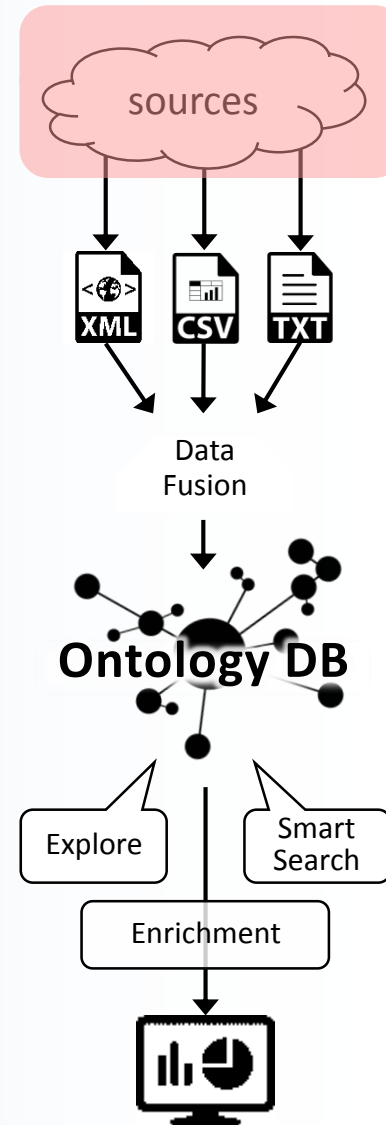
**Based on** *The Avengers* by Stan Lee  
Jack Kirby

**Starring** Robert Downey Jr.  
Chris Hemsworth  
Mark Ruffalo  
Chris Evans  
Scarlett Johansson  
Jeremy Renner  
Don Cheadle  
Aaron Taylor-Johnson  
Elizabeth Olsen  
Paul Bettany



# Team of Heroes – The source Finder

Team Member	Role
Source finder	Locate relevant data sources, evaluate completeness and accuracy. negotiate usage.



# Web Scrapping

The automatic extraction of data from websites into local files or a database

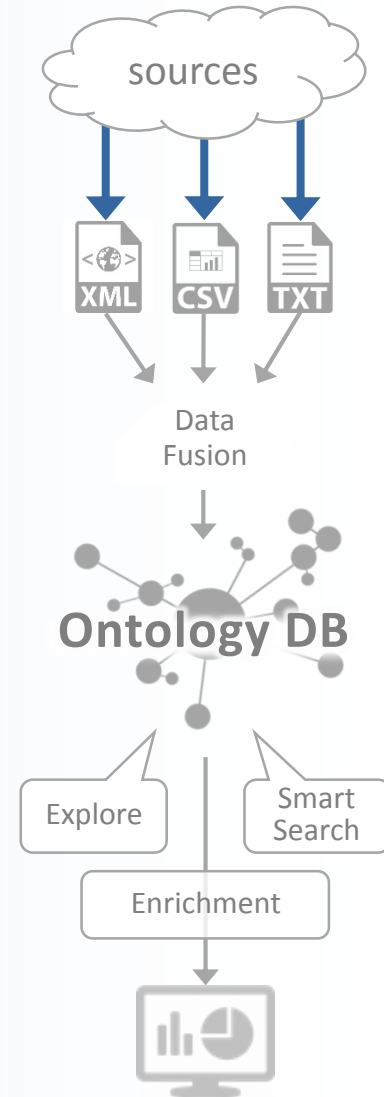
- highlight the required elements and specify a name for each
- The scraper will extract the data into a table
- Many scraping tools and services are available




mozenda



winautomation



# Wikipedia and DBpedia



WIKIPEDIA  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction

Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools

What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Wikidata item  
Cite this page

Print/export

Create a book  
Download as PDF  
Printable version

In other projects

Wikimedia Commons  
Wikiquote

Languages

Not logged in | Talk | Contributions | Create account | Log in

Article | Talk

Read | Edit | View history

Search Wikipedia

## Avengers: Age of Ultron

From Wikipedia, the free encyclopedia

**Avengers: Age of Ultron** is a 2015 American superhero film based on the Marvel Comics superhero team the *Avengers*, produced by Marvel Studios and distributed by Walt Disney Studios Motion Pictures. It is the sequel to 2012's *The Avengers* and the eleventh film in the Marvel Cinematic Universe (MCU). The film was written and directed by Joss Whedon and features an ensemble cast that includes Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, Scarlett Johansson, Jeremy Renner, Don Cheadle, Aaron Taylor-Johnson, Elizabeth Olsen, Paul Bettany, Cobie Smulders, Anthony Mackie, Hayley Atwell, Idris Elba, Stellan Skarsgård, James Spader, and Samuel L. Jackson. In *Avengers: Age of Ultron*, the Avengers fight *Ultron*, an artificial intelligence obsessed with causing human extinction.


The sequel was announced in May 2012, after the successful release of *The Avengers*. Whedon, the director of the first film, was brought back on board in August and a release date was set. By April 2013, Whedon had completed a draft of the script, and casting began in June with the re-signing of Downey. *Second unit* filming began in February 2014 in South Africa with principal photography taking place between March and August 2014. The film was primarily shot at Shepperton Studios in Surrey, England, with additional footage filmed in Italy, South Korea, Bangladesh, New York City, and various locations around England. While in *post production*, the film was converted to 3D and over 3,000 *visual effects* shots were added.

*Avengers: Age of Ultron* premiered in Los Angeles on April 13, 2015, and was released on May 1, 2015, in North America, in 3D and IMAX 3D. The film received positive reviews from critics and grossed over \$1.4 billion worldwide, making it the *seventh-highest-grossing film in history* and the *fourth-highest-grossing film of 2015*. A sequel, *Avengers: Infinity War*, is scheduled to be released on May 4, 2018, and another *untitled sequel* is scheduled for May 3, 2019.

### Contents [hide]

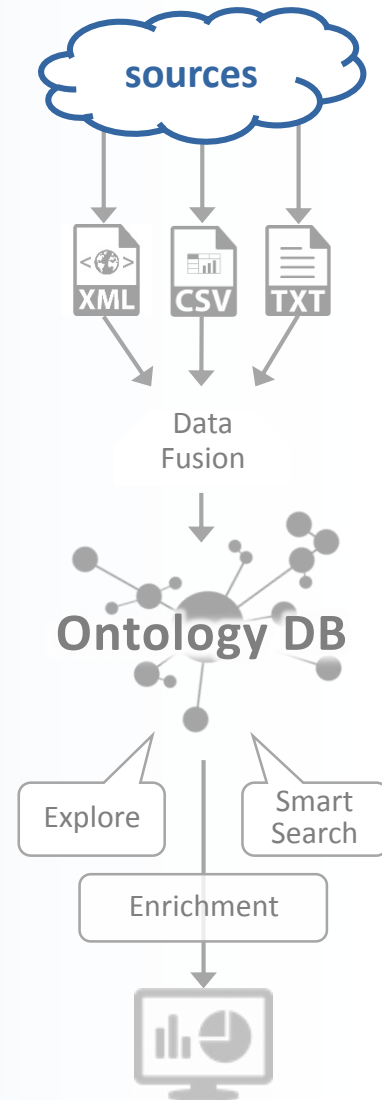
- 1 Plot
- 2 Cast
- 3 Production
  - 3.1 Development
  - 3.2 Pre-production
  - 3.3 Filming

### Avengers: Age of Ultron



Theatrical release poster

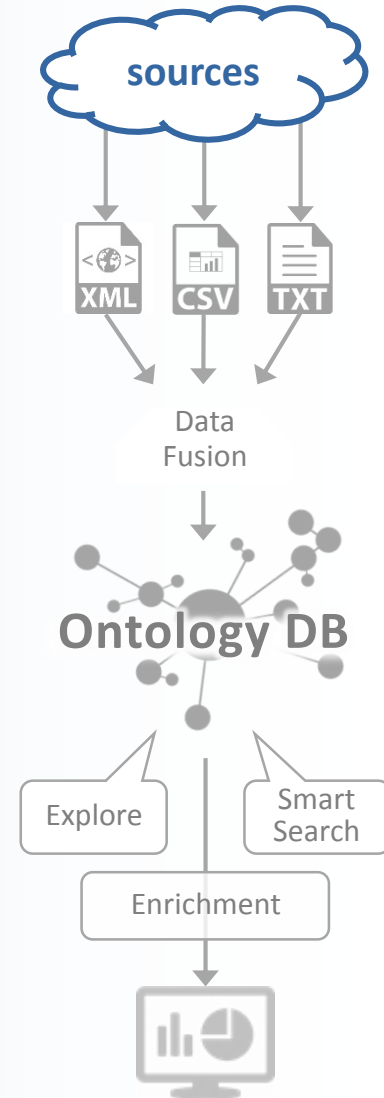
**Directed by** Joss Whedon  
**Produced by** Kevin Feige  
**Written by** Joss Whedon  
**Based on** *The Avengers* by Stan Lee  
Jack Kirby  
**Starring** Robert Downey Jr.  
Chris Hemsworth  
Mark Ruffalo  
Chris Evans  
Scarlett Johansson  
Jeremy Renner  
Don Cheadle  
Aaron Taylor-Johnson  
Elizabeth Olsen  
Paul Bettany



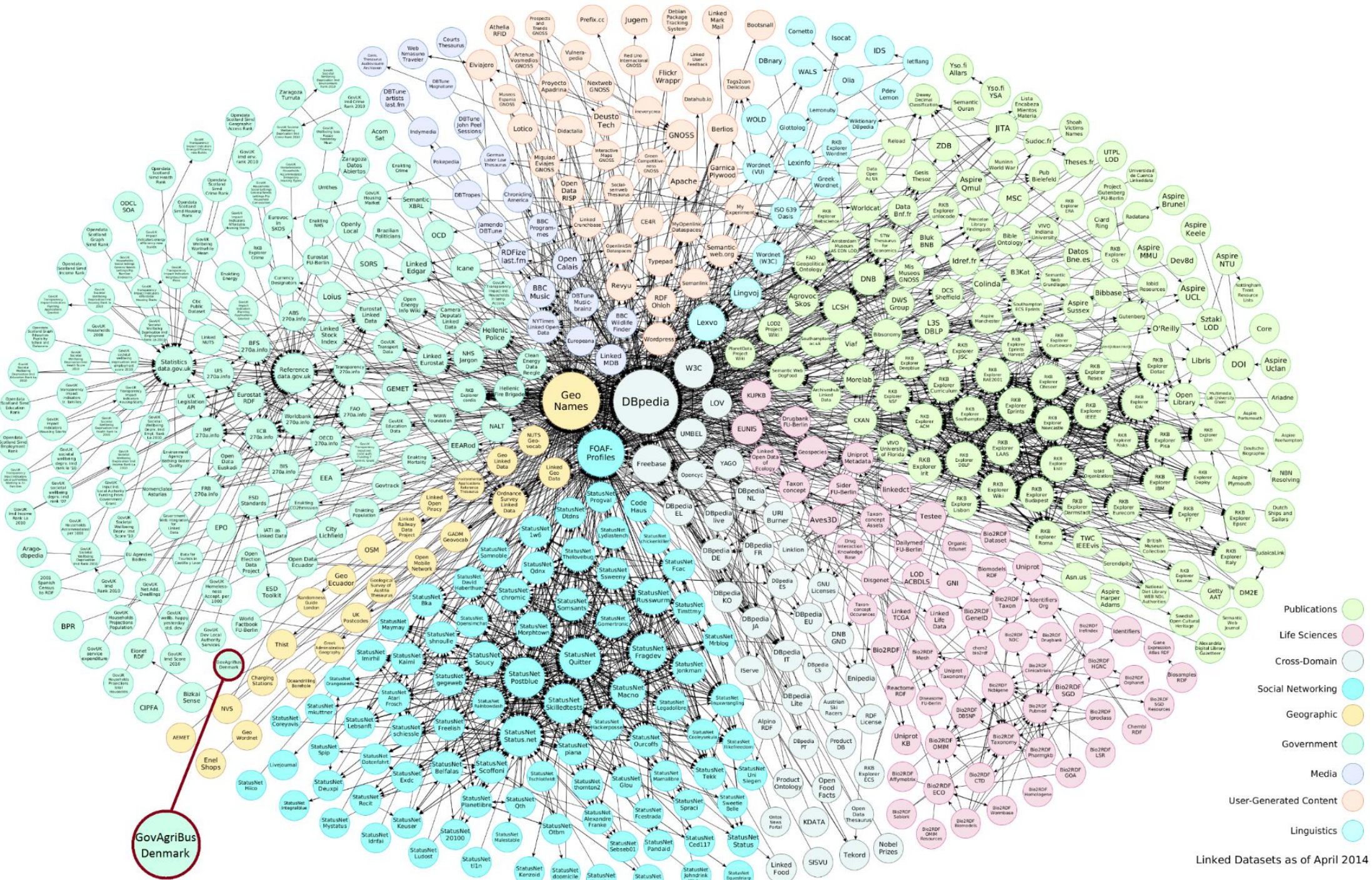


# The Semantic Web

- A “web of **data**” vs. a “web of **documents**”
- Hundreds of interconnected, freely available databases
  - DBpedia (~3.5M concepts)
  - GeoNames (> 7.5M locations)
- SPARQL - The standard querying language



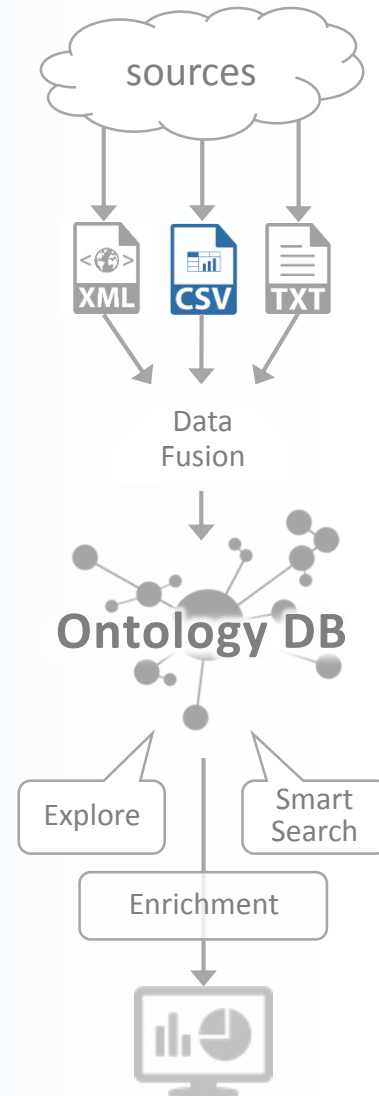






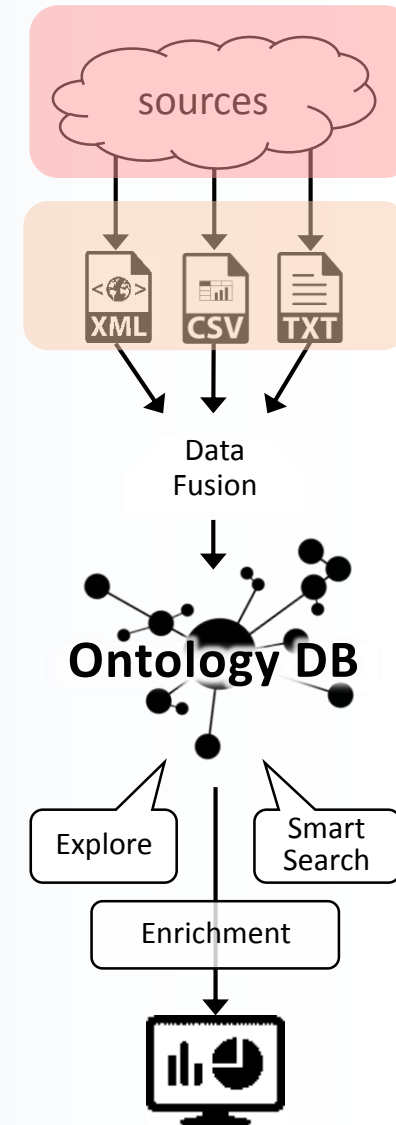
# The Semantic Web

title	budget	box_office
Army of Darkness	11000000	21500000
Barry Lyndon	11000000	20000000
Barry Lyndon	11000000	31500000
Batman & Robin (film)	125	238200000
Batman (1966 film)	1500000	1700000
Batman (1989 film)	48000000	411300000
Batman Forever	100000000	336500000
Batman Returns	65000000	266800000
Blade Runner	28000000	33800000
Blazing Saddles	2600000	119500000
Blue Velvet (film)	6000000	8600000
Braveheart	72000000	210400000
Buffy the Vampire Slayer (film)	7000000	16624456



# Team of Heroes – The data Extractor

Team Member	Role
Source finder	Locate relevant data sources, evaluate coverage and accuracy, ensure freedom to operate on the data
Data extractor	<b>Extract relevant data from the sources into files (scraping, semantic web)</b>





# Building *Our* Movie DB – Ontology

Source: DBpedia

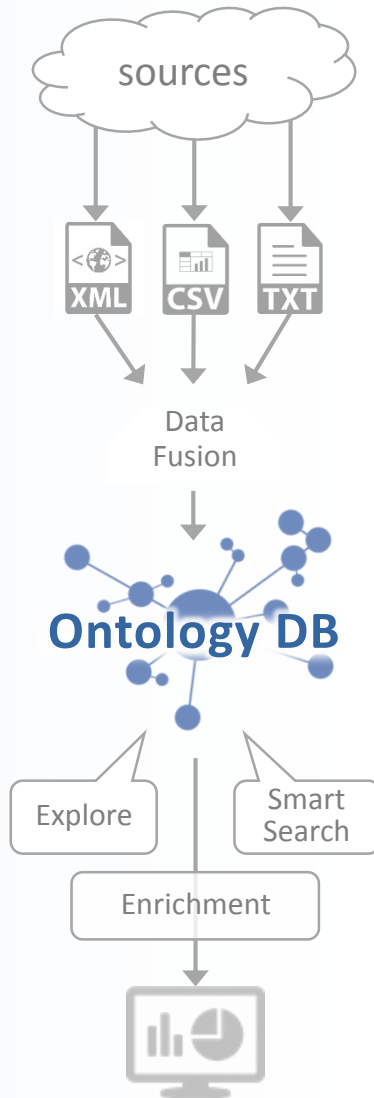
	A	B	C
1	<b>title</b>	<b>budget</b>	<b>box_office</b>
2	Army of Darkness	11,000,000	21,500,000
3	Barry Lyndon	11,000,000	20,000,000
4	Barry Lyndon	11,000,000	31,500,000
5	Batman & Robin (film)	125	238,200,000

Source: List of American superhero films

	A	B	C	D	E	F
1	<b>year</b>	<b>film</b>	<b>publisher</b>	<b>director</b>	<b>distributor</b>	<b>worldwide gross</b>
2	<a href="#">2017</a>	<a href="#">Logan</a>	<a href="#">Marvel Comics</a>	<a href="#">James Mangold</a>	<a href="#">20th Century Fox</a>	\$607,733,434
3						
4	<a href="#">2017</a>	<a href="#">Power Rangers</a>	Original	<a href="#">Dean Israelite</a>	<a href="#">Lionsgate</a>	\$134,162,300
5	<a href="#">2017</a>	<a href="#">Guardians of the Galaxy Vol. 2</a>	<a href="#">Marvel Comics</a>	<a href="#">James Gunn</a>	<a href="#">Walt Disney Studios Motion Pictures</a>	
6						
7	<a href="#">2017</a>	<a href="#">Wonder Woman</a>	<a href="#">DC Comics</a>	<a href="#">Patty Jenkins</a>	<a href="#">Warner Bros.</a>	
8						

Source: IMDb

	A	B	C	D	E	F
1	<b>Title</b>	<b>storyline</b>	<b>synopsis</b>	<b>URL</b>	<b>#ratings</b>	<b>User rating</b>
2	Alter Egos (2012) - IMDb	Brendan, a.k.a. the instant ice-forming superhero		<a href="http://www.imdb.com/title/tt1907614">http://www.imdb.com/title/tt1907614</a>	1,351	5.2
3	Avengers: Age of Ultron (2015) - IMDb	Tony Stark creates the	In the Eastern Europe	<a href="http://www.imdb.com/title/tt2395427">http://www.imdb.com/title/tt2395427</a>	513,723	7.4
4	Big Hero 6 (2014) - IMDb	When a devastating eve	In the city of San Frans	<a href="http://www.imdb.com/title/tt2245084">http://www.imdb.com/title/tt2245084</a>	307,626	7.8
5	Captain America: Civil War (2016) - IMDb	With many people fear	In 1991, Bucky Barnes	<a href="http://www.imdb.com/title/tt3498820">http://www.imdb.com/title/tt3498820</a>	406,826	7.9
6	Fantastic 4: Rise of the Silver Surfer (2007) - IMDb	Everything seems to be	The movie opens out i	<a href="http://www.imdb.com/title/tt0486576">http://www.imdb.com/title/tt0486576</a>	219,817	5.6
7	Fantastic Four (2005) - IMDb	Reed Richards, a brilliant but timid and bankrupt		<a href="http://www.imdb.com/title/tt0120667">http://www.imdb.com/title/tt0120667</a>	276,058	5.7
8	Fantastic Four (2015) - IMDb	FANTASTIC FOUR, a co	In the year 2007, youn	<a href="http://www.imdb.com/title/tt1502712">http://www.imdb.com/title/tt1502712</a>	121,344	4.3



# Building *Our* Movie DB – Ontology

Source: DBpedia

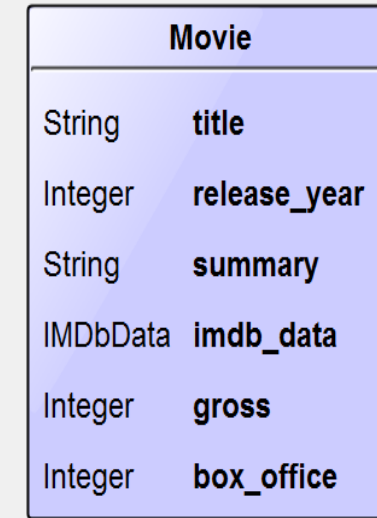
	A	B	C
1	<b>title</b>	<b>budget</b>	<b>box_office</b>
2	Army of Darkness	11,000,000	21,500,000
3	Barry Lyndon	11,000,000	20,000,000
4	Barry Lyndon	11,000,000	31,500,000
5	Batman & Robin (film)	125	238,200,000

Source: List of American superhero films

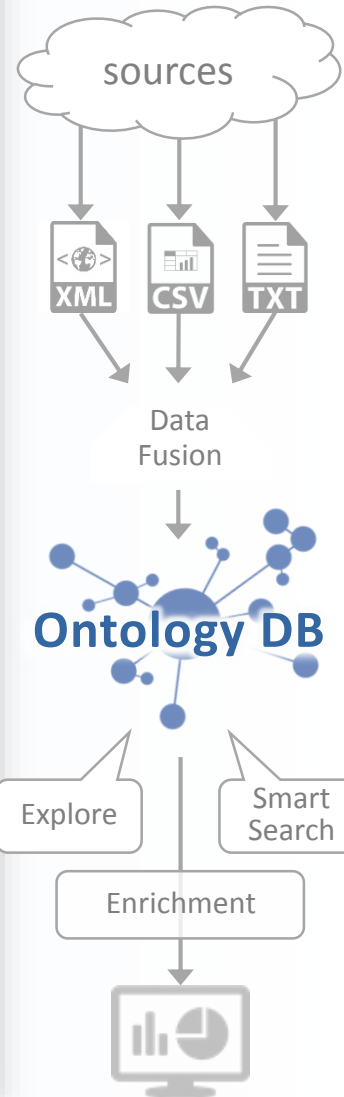
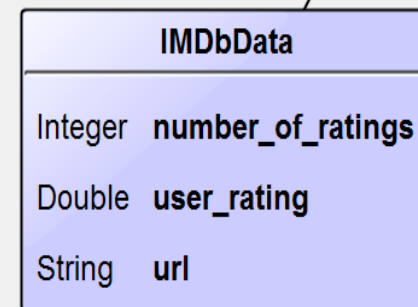
	A	B	C	D	E	F
1	<b>year</b>	<b>film</b>	<b>publisher</b>	<b>director</b>	<b>distributor</b>	<b>worldwide gross</b>
2	<a href="#">2017</a>	<a href="#">Logan</a>	<a href="#">Marvel Comics</a>	<a href="#">James Mangold</a>	<a href="#">20th Century Fox</a>	\$607,733,434
3						
4	<a href="#">2017</a>	<a href="#">Power Rangers</a>	Original	<a href="#">Dean Israelite</a>	<a href="#">Lionsgate</a>	\$134,162,300
5	<a href="#">2017</a>	<a href="#">Guardians of the Galaxy Vol. 2</a>	<a href="#">Marvel Comics</a>	<a href="#">James Gunn</a>	<a href="#">Walt Disney Studios Motion Pictures</a>	
6						
7	<a href="#">2017</a>	<a href="#">Wonder Woman</a>	<a href="#">DC Comics</a>	<a href="#">Patty Jenkins</a>	<a href="#">Warner Bros.</a>	
8						

Source: IMDb

	A	B	C	D	E	F
1	<b>Title</b>	<b>storyline</b>	<b>synopsis</b>	<b>URL</b>	<b>#ratings</b>	<b>User rating</b>
2	Alter Egos (2012) - IMDb	Brendan, a.k.a. the instant ice-forming superhero		<a href="http://www.imdb.com/title/tt1907614">http://www.imdb.com/title/tt1907614</a>	1,351	5.2
3	Avengers: Age of Ultron (2015) - IMDb	Tony Stark creates the	In the Eastern Europe	<a href="http://www.imdb.com/title/tt2395427">http://www.imdb.com/title/tt2395427</a>	513,723	7.4
4	Big Hero 6 (2014) - IMDb	When a devastating ev	In the city of San Frans	<a href="http://www.imdb.com/title/tt2245084">http://www.imdb.com/title/tt2245084</a>	307,626	7.8
5	Captain America: Civil War (2016) - IMDb	With many people fear	In 1991, Bucky Barnes	<a href="http://www.imdb.com/title/tt3498820">http://www.imdb.com/title/tt3498820</a>	406,826	7.9
6	Fantastic 4: Rise of the Silver Surfer (2007) - IMDb	Everything seems to be	The movie opens out i	<a href="http://www.imdb.com/title/tt0486576">http://www.imdb.com/title/tt0486576</a>	219,817	5.6
7	Fantastic Four (2005) - IMDb	Reed Richards, a brilliant but timid and bankrupt		<a href="http://www.imdb.com/title/tt0120667">http://www.imdb.com/title/tt0120667</a>	276,058	5.7
8	Fantastic Four (2015) - IMDb	FANTASTIC FOUR, a co	In the year 2007, youn	<a href="http://www.imdb.com/title/tt1502712">http://www.imdb.com/title/tt1502712</a>	121,344	4.3



imdb\_data



# Building *Our* Movie DB – Data Fusion

Source: DBpedia

	A	B	C
1	title	budget	box_office
2	Army of Darkness	11,000,000	21,500,000
3	Barry Lyndon	11,000,000	20,000,000
4	Barry Lyndon	11,000,000	31,500,000
5	Batman & Robin (film)	125	238,200,000

Source: List of American superhero films

	A	B	C	D	E	F
1	year	film	publisher	director	distributor	worldwide gross
2	2017	<a href="#">Logan</a>	<a href="#">Marvel Comics</a>	<a href="#">James Mangold</a>	<a href="#">20th Century Fox</a>	\$607,733,434
3						
4	2017	<a href="#">Power Rangers</a>	Original	<a href="#">Dean Israelite</a>	<a href="#">Lionsgate</a>	\$134,162,300
5	2017	<a href="#">Guardians of the Galaxy Vol. 2</a>	<a href="#">Marvel Comics</a>	<a href="#">James Gunn</a>	<a href="#">Walt Disney Studios Motion Pictures</a>	
6						
7						
8	2017	<a href="#">Wonder Woman</a>	<a href="#">DC Comics</a>	<a href="#">Patty Jenkins</a>	<a href="#">Warner Bros.</a>	

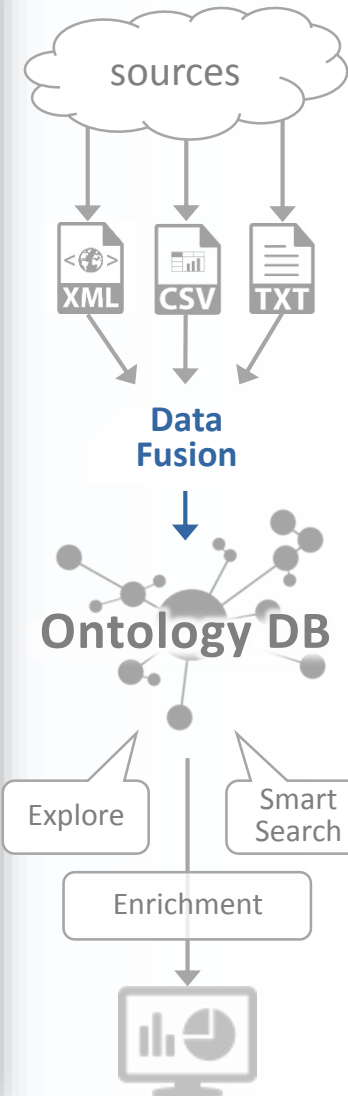
Source: IMDb

	A	B	C	D	E	F
1	Title	storyline	synopsis	URL	#ratings	User rating
2	Alter Egos (2012) - IMDb	Brendan, a.k.a. the instant ice-forming superhero		<a href="http://www.imdb.com/title/tt1907614">http://www.imdb.com/title/tt1907614</a>	1,351	5.2
3	Avengers: Age of Ultron (2015) - IMDb	Tony Stark creates the	In the Eastern Europe	<a href="http://www.imdb.com/title/tt2395427">http://www.imdb.com/title/tt2395427</a>	513,723	7.4
4	Big Hero 6 (2014) - IMDb	When a devastating ev	In the city of San Frans	<a href="http://www.imdb.com/title/tt2245084">http://www.imdb.com/title/tt2245084</a>	307,626	7.8
5	Captain America: Civil War (2016) - IMDb	With many people fear	In 1991, Bucky Barnes	<a href="http://www.imdb.com/title/tt3498820">http://www.imdb.com/title/tt3498820</a>	406,826	7.9
6	Fantastic 4: Rise of the Silver Surfer (2007) - IMDb	Everything seems to be	The movie opens out i	<a href="http://www.imdb.com/title/tt0486576">http://www.imdb.com/title/tt0486576</a>	219,817	5.6
7	Fantastic Four (2005) - IMDb	Reed Richards, a brilliant but timid and bankrupt		<a href="http://www.imdb.com/title/tt0120667">http://www.imdb.com/title/tt0120667</a>	276,058	5.7
8	Fantastic Four (2015) - IMDb	FANTASTIC FOUR, a co	In the year 2007, youn	<a href="http://www.imdb.com/title/tt1502712">http://www.imdb.com/title/tt1502712</a>	121,344	4.3

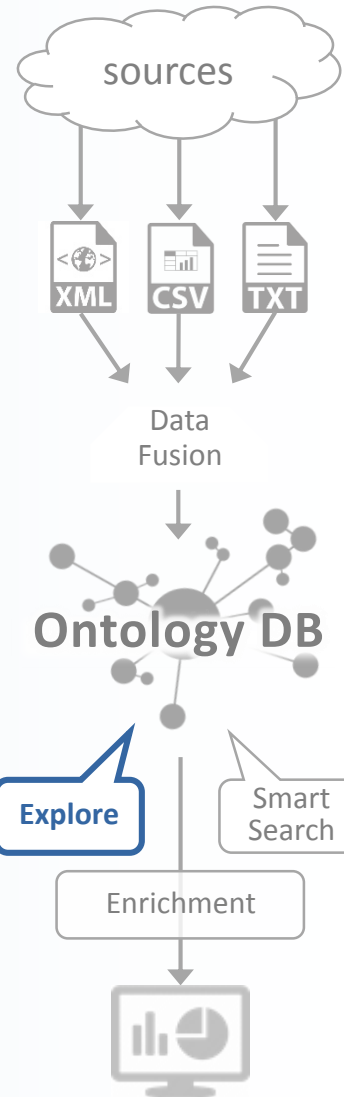
Movie	
String	title
Integer	release_year
String	summary
IMDbData	imdb_data
Integer	gross
Integer	box_office

imdb\_data

IMDbData	
Integer	number_of_ratings
Double	user_rating
String	url



# and Voilà!



### Movie Query

Search

Id

Title

Release Year

Box Office [\$]

Budget [\$]

IMDb Number of Ratings

IMDb User Rating

Showing all 137 results

#	Release Year	Title	Box Office [\$]	IMDb User Rating	IMDb Number of Ratings	Summary
1	2012	The Avengers	623,279,547	8	1042790	Nick F...
2	2008	The Dark Knight	533,316,061	9	1785100	Set w...
3	2015	Avengers: Age of Ultron	458,991,599	7	513723	Tony
4	2012	The Dark Knight Rises	448,130,642	8	1218258	Despi...
5	2013	Iron Man 3	408,992,272	7	588804	Marve...
6	2016	Captain America: Civil War	408,080,554	7	406826	With i...
7	2002	Spider-Man	403,706,375	7	572321	Basec...
8	2004	Spider-Man 2	373,377,893	7	437277	Peter
9	2016	Deadpool	363,024,263	8	621917	This i...
10	2007	Spider-Man 3	336,530,303	6	404815	Peter
11	2014	Guardians of the Galaxy	333,130,696	8	748309	After
12	2016	Batman v Superman: Dawn of Justice	330,249,062	6	469462	The g...
13	2016	Suicide Squad	325,021,779	6	386468	It feel...
14	2008	Iron Man	318,298,180	7	735007	Tony
15	2010	Iron Man 2	312,057,433	7	554455	With i...
16	2013	Man of Steel	291,021,565	7	575461	A you...
17	2012	The Amazing Spider-Man	262,030,663	7	472950	Peter
18	2004	The Incredibles	261,437,578	8	507047	Bob F...

### The Avengers (2012)

#### Poster



Box Office [\$] 623,279,547

Budget [\$] 220,000,000

#### IMDb Data

Url [http://www.imdb.com/title/tt0848228/?ref\\_=adv\\_li\\_tt](http://www.imdb.com/title/tt0848228/?ref_=adv_li_tt)

User Rating 8.1

Number of Ratings 1,042,790

#### Summary

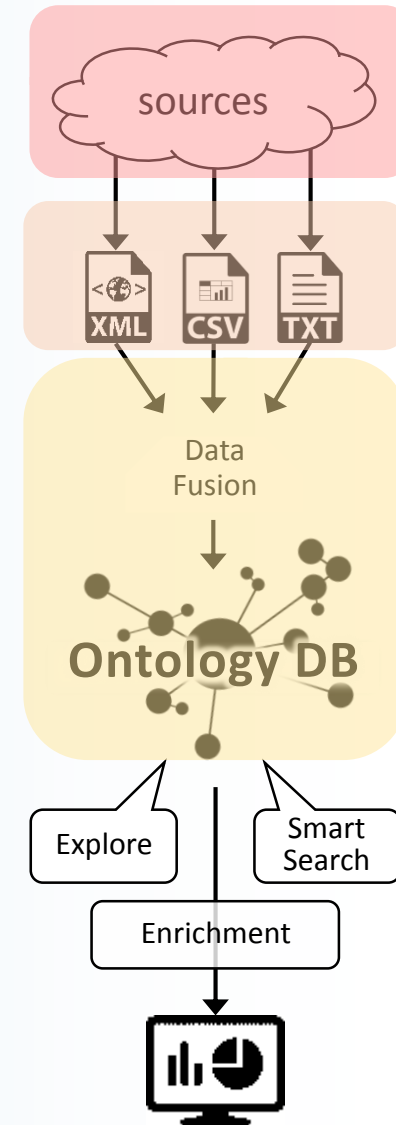
Nick Fury is the director of S.H.I.E.L.D., an international peace-keeping agency. The agency is a who's who of Marvel Super Heroes, with Iron Man, The Incredible Hulk, Thor, Captain America, Hawkeye and Black Widow. When global security is threatened by Loki and his cohorts, Nick Fury and his team will need all their powers to save the world from disaster.

Nick Fury (), director of S.H.I.E.L.D. (Strategic Homeland Intervention, Enforcement, and Logistics Division), arrives at S.H.I.E.L.D. headquarters outside of Santa Fe, New Mexico, during



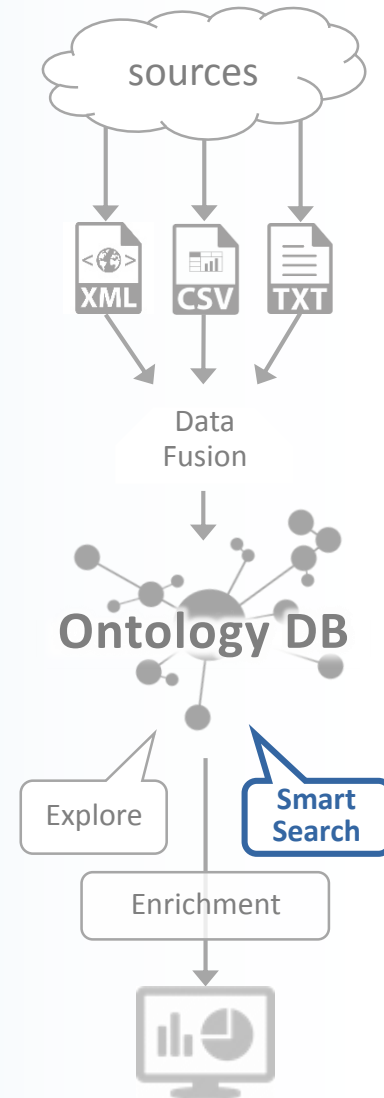
# Team of Heroes – The Modeler

Team Member	Role
Source finder	Locate relevant data sources, evaluate coverage and accuracy, ensure freedom to operate on the data
Data extractor	Extract relevant data from the sources into files (scraping, semantic web)
Modeler	<b>Build the ontology, load the data into the DB, handle data cleansing, alignment and transformation. Resolve conflicts between sources</b>



# Textual Analysis – Smart Search

- How could we classify a movie as **single hero** vs. **team of heroes**?
- The answer lies within the **movie summary**
- Natural Language Processing (NLP) is often too complex
- “**Smart Search**” is a method for identifying relevant document out of a large set of documents



# Textual Analysis – Smart Search

Article - Superhero team movies **Harvest** Showing all 69 results **I'm feeling lucky** **Refine**

Query Terms (4)

#	Term
1	team
2	superheroes
3	band
4	heroes

Suggested Query Terms

#	Term
---	------

Ignored Query Terms

Movie

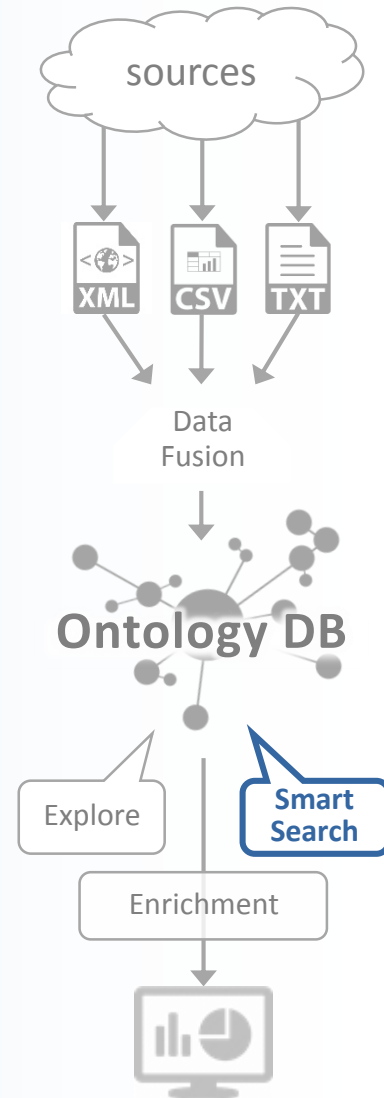
### Justice Squad

**Summary**

In the lead up to the first ever Australian superhero audition, a documentary crew spend a week with South Australia's premiere crime-fighting **team**, the Justice Squad. The micro-managing boss, Captain Justice, leads a less-than-super squad of **heroes**, including the matricidal Dark Avenger, the womanizing Mr Quick, the short-circuiting Electrode and the newest addition to the **team**, the elusive Mistress Mystery. The **heroes** might just get some work done if not for the distractions of the interfering camera crew, a less than ambitious work experience kid, Dark Avenger's meddling mother, and a group of nefarious super-villains who work just down the hall. As tensions between the **heroes** threaten to tear the squad apart, will they be able to pull together in time to impress the judges and win the holy grail of super-heroism... the comic book deal?

Mention the words 'Australian' and 'Superhero' in the same sentence and people will laugh in your face. But they do exist... and there are a lot more of them than you think. In the lead up to the first ever Australian superhero audition, a documentary crew spend a week with South Australia's premiere crime-fighting **team**, the Justice Squad. The micro-managing boss, Captain Justice, leads a less-than-super squad of **heroes**, including the matricidal Dark Avenger, the womanizing Mr Quick, the short-circuiting Electrode and the newest addition to the **team**, the elusive Mistress Mystery. The **heroes** might just get some work done if not for the distractions of the interfering camera crew, a less than ambitious work experience kid, Dark Avenger's meddling mother, and a group of nefarious super-villains who work just down the hall. As tensions between the **heroes** threaten to tear the squad apart, will they be able to pull together in time to impress the judges and win the holy grail of super-heroism... the comic book deal?

#	Score	Relevancy Score	Relevancy	Document	Tags
27				Howard the Duck (1986)	
28				Zoom (2006)	
29				Ultraman X the Movie (2016)	



# Textual Analysis – Smart Search

Article - Superhero team movies **Harvest**

Showing all 69 results **I'm feeling lucky** **Refine**

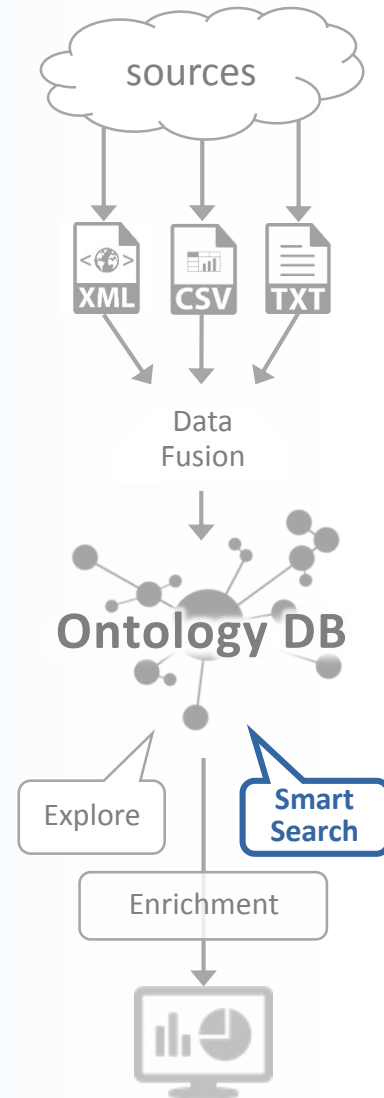
Query Terms (4)

#	Term
1	team
2	superheroes
3	band
4	heroes

Suggested Query Terms (20)

#	Term
1	together
2	join
3	group
4	crew
5	squad
6	they
7	them
8	their
9	elusive
10	interfering
11	micro
12	tensions
13	experience
14	ed
15	judges
16	torture
17	audition
18	packed
19	debate

#	Score	Relevancy Score	Relevancy	Document	Tags
1	<div></div>	<div></div>	<div></div>	Justice Squad (2012)	
2	<div></div>	<div></div>	<div></div>	Super Capers: The Origins of Ed and the Missing Bullion (2009)	
3	<div></div>	<div></div>	<div></div>	American Rescue Squad (2015)	
4	<div></div>	<div></div>		Suicide Squad (2016)	
5	<div></div>	<div></div>		Watchmen (2009)	
6	<div></div>	<div></div>		Surge of Power: The Stuff of Heroes (2004)	
7	<div></div>	<div></div>		Super (2010)	
8	<div></div>	<div></div>		Kick-Ass (2010)	
9	<div></div>	<div></div>		Big Hero 6 (2014)	
10	<div></div>	<div></div>		X-Men Origins: Wolverine (2009)	
11	<div></div>	<div></div>		Night-Wolf (2005)	
12	<div></div>	<div></div>		Megamind (2010)	
13	<div></div>	<div></div>		Unsung Heroes (2003)	
14	<div></div>	<div></div>		The Avengers (2012)	
15	<div></div>	<div></div>		Kick-Ass 2 (2013)	
16	<div></div>	<div></div>		Unbreakable (2000)	
17	<div></div>	<div></div>		Batman: The Movie (1966)	
18	<div></div>	<div></div>		Zoom (2006)	
19	<div></div>	<div></div>		Sky High (2005)	
20	<div></div>	<div></div>		Thor (2011)	
21	<div></div>	<div></div>		The Incredibles (2004)	
22	<div></div>	<div></div>		Defective Man! (2009)	
23	<div></div>	<div></div>		Superman Returns (2006)	
24	<div></div>	<div></div>		Captain America: Civil War (2016)	
25	<div></div>	<div></div>		TMNT (2007)	
26	<div></div>	<div></div>		The Marvel Experience (2014)	
27	<div></div>	<div></div>		Heroes Crossing (2010)	
28	<div></div>	<div></div>		Howard the Duck (1986)	
29	<div></div>	<div></div>		The Specials (2000)	





# Textual Analysis – Smart Search

Article - Superhero team movies **Harvest** Showing all 199 results **I'm feeling lucky** **Refine**

Query Terms (12)

#	Term
1	band
2	crew
3	group
4	heroes
5	join
6	squad
7	superheroes
8	team
9	their
10	them
11	they
12	together

Suggested Query Terms (20)

#	Term
1	taxpayer
2	capers
3	gruberman
4	australian
5	bullion
6	ambitious
7	mistress
8	numbers
9	midst
10	man2010
11	heroism
12	gold

**Movie**

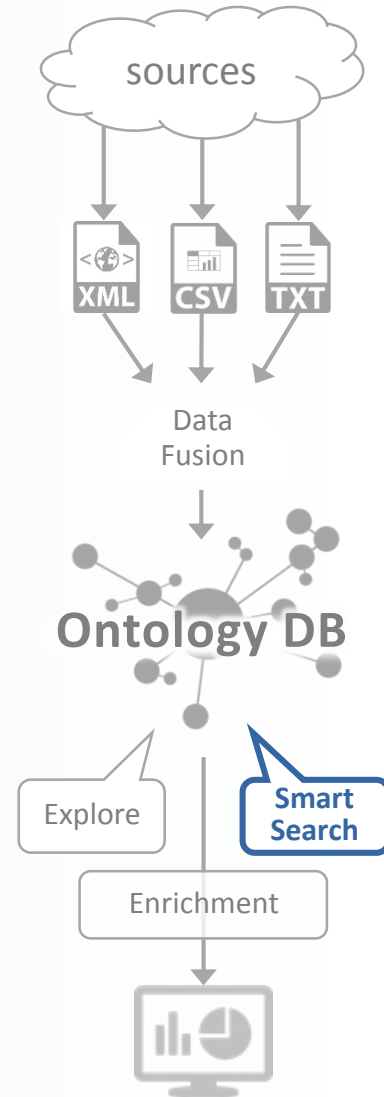
### Avengers: Age of Ultron

**Summary**

Tony Stark creates the Ultron Program to protect the world, but when the peacekeeping program becomes hostile, The Avengers go into action to try and defeat a virtually impossible enemy **together**. Earth's mightiest **heroes** must come **together** once again to protect the world from global extinction.

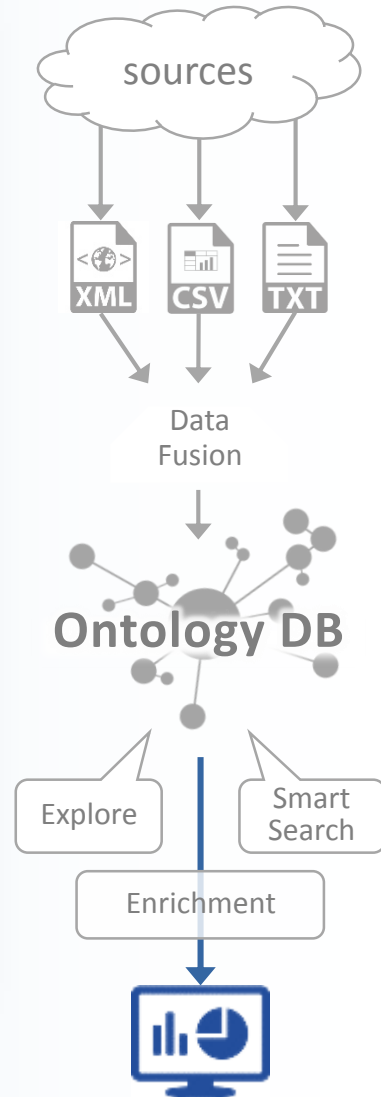
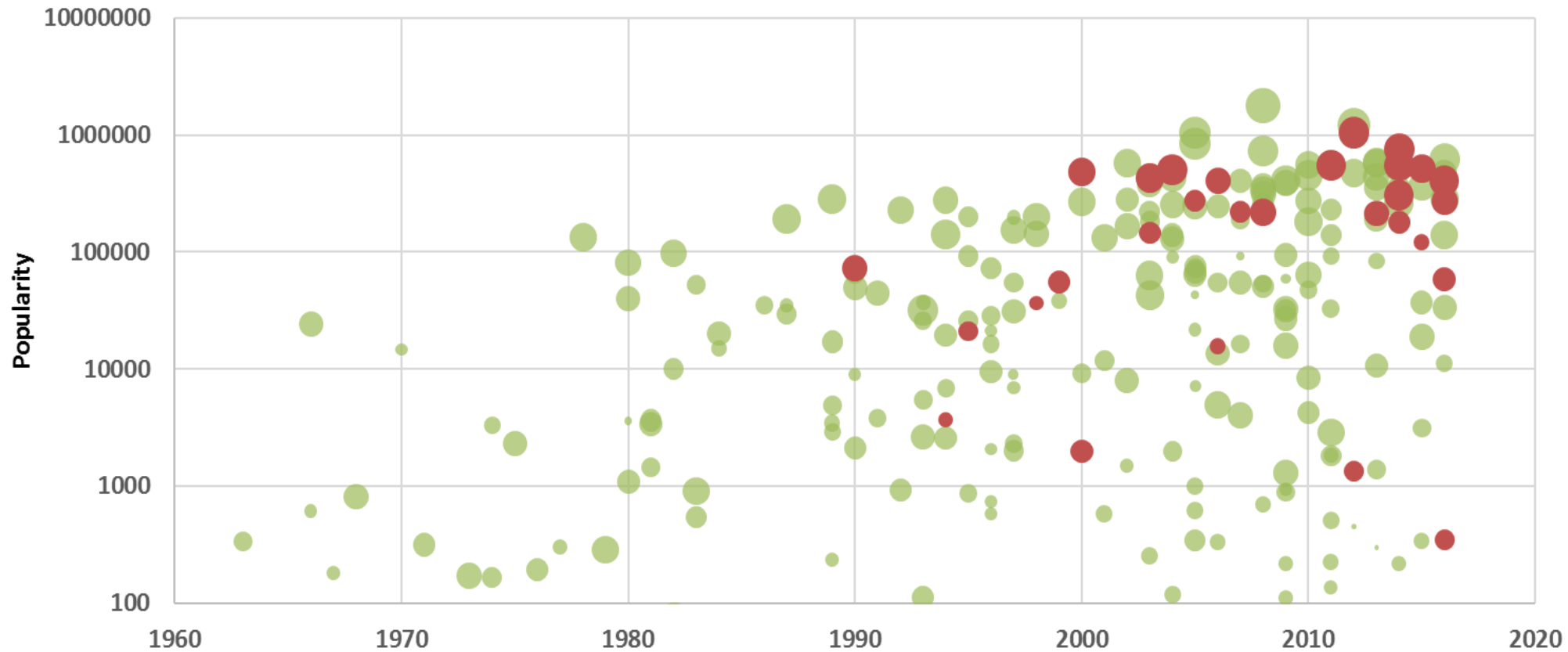
In the Eastern European country of Sokovia, the Avengers - Tony Stark / Iron Man (Robert Downey Jr), Steve Rogers / Captain America (Chris Evans), Thor (Chris Hemsworth), Bruce Banner / Hulk (Mark Ruffalo), Natasha Romanoff / Black Widow (Scarlett Johansson), and Clint Barton / Hawkeye (Jeremy Renner) - raid a Hydra outpost commanded by Baron Wolfgang von Strucker (Thomas Kretschmann), who has been experimenting on humans using the sceptre previously wielded by Loki. **They** encounter two of Strucker's experiments: the Maximoff twins Pietro (Aaron Taylor-Johnson), who has superhuman speed, and Wanda (Elizabeth Olsen), who has telekinetic abilities, and can manipulate minds and project energy. Rogers captures Strucker, while Stark retrieves the sceptre after Wanda implants a fear-filled vision of a bleak possible future in his mind. Back at the **team's** headquarters - Avengers Tower (formerly Stark Tower) - Stark and Banner discover an artificial intelligence within the sceptre's gem, and secretly plan to use it to complete Stark's "Ultron" global defence program. After many failed attempts, **they** leave for a victory party to celebrate the recapture of the sceptre. Unbeknownst to **them**, the program comes to life... The unexpectedly sentient Ultron (James Spader) awakens, believing that he must eradicate humanity to save the Earth, and overwhelms Stark's AI, JARVIS (Paul Bettany). In the guise of one of Stark's Iron Legion robot drones, Ultron attacks the Avengers as **they** relax **together** after **their** party. Escaping with the sceptre, Ultron uses the resources in Strucker's Sokovia base to upgrade his rudimentary body and build an army of robot drones. After killing Strucker, he then meets and recruits the Maximoff twins, who hold Stark responsible for **their** parents' deaths

#	Score	Relevancy Score	Relevancy	Document	Tags
27				Defective Man! (2009)	
28				Teenage Mutant Ninja Turtles (2014)	
29				Forest Warrior (1996)	



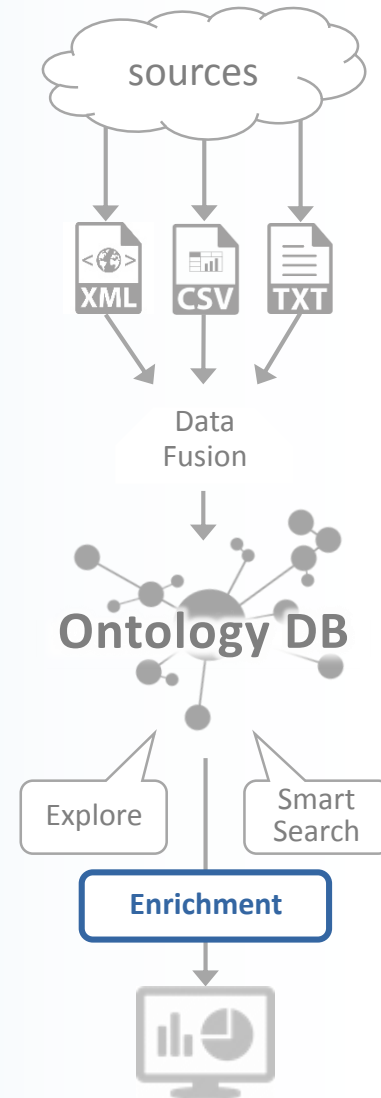
# Infographics

## Hero-Teams vs. Superhero Movies



# Enrichment Tasks

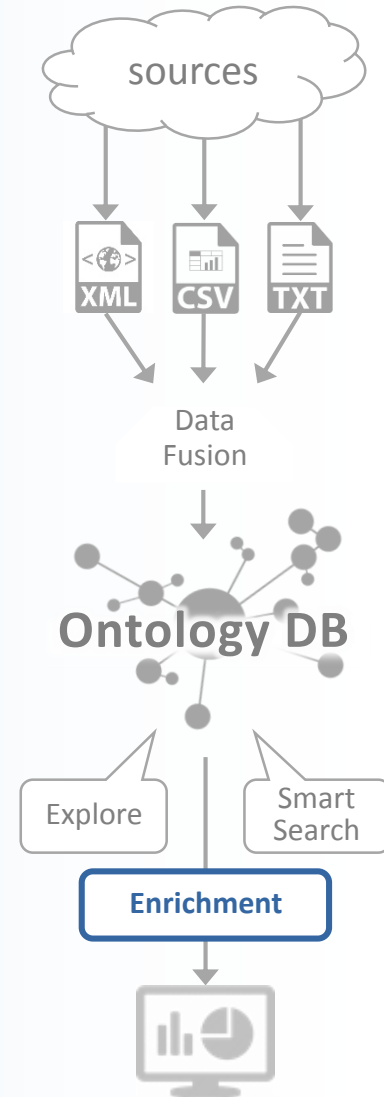
- Other interesting parameters:
  - Male hero vs. female hero
  - Animated vs. non-animated
  - comics based vs. non-comics based
- Hard to classify algorithmically
- Use humans: “artificial artificial intelligence”
  - Generate tasks
  - Employ users (technicians) to accomplish these tasks
  - Monitor progress and manage the project



# Crowd Sourcing

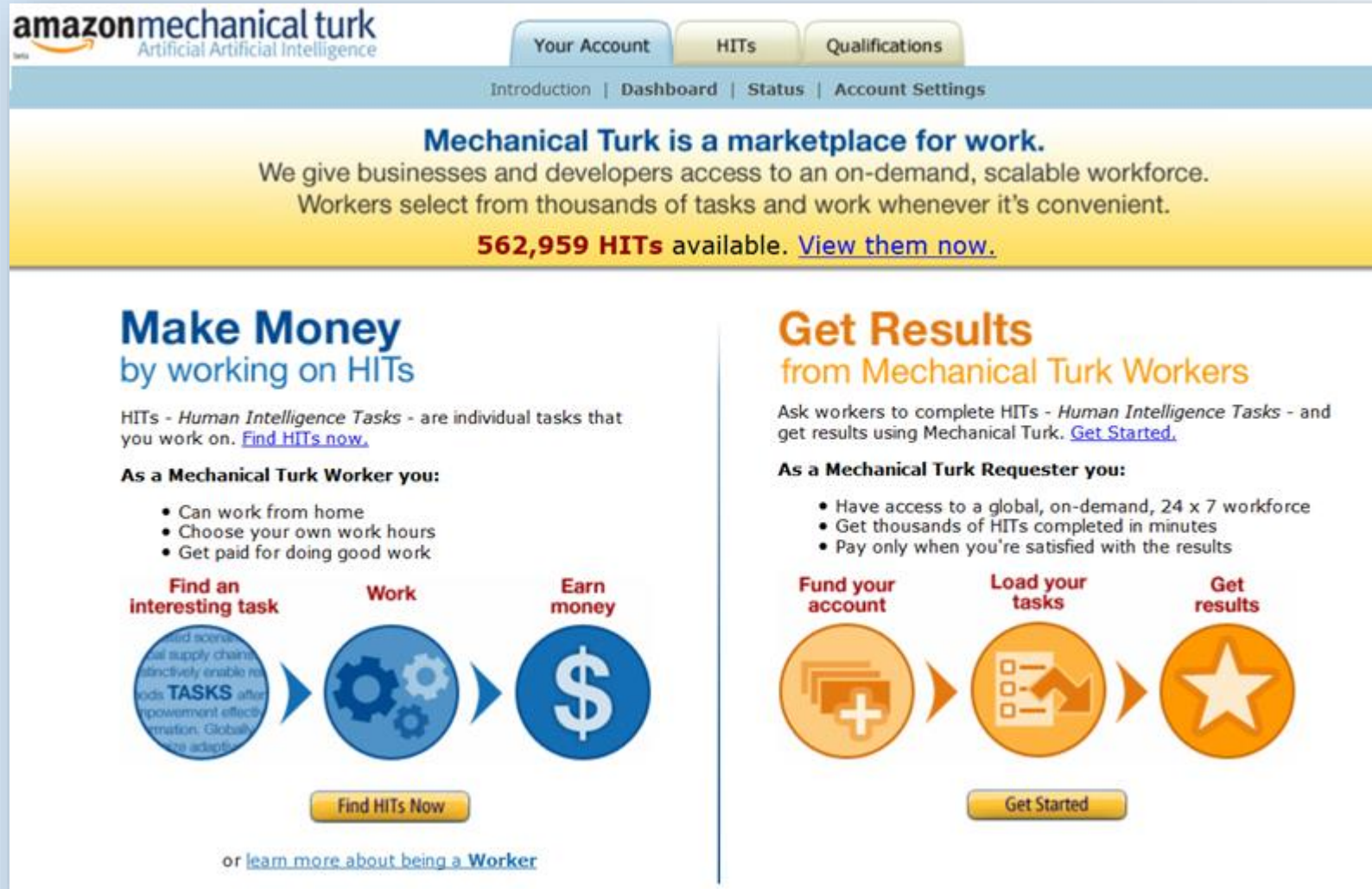
What if there are more tasks than we can handle?

- Ask the crowd for help!
- People (“workers”) are willing to accomplish simple tasks (“micro-tasks”) for a small fee (~1 cent)
- More than a million workers are enrolled in crowd sourcing platforms





# Amazon Mechanical Turk



The screenshot shows the Amazon Mechanical Turk homepage. At the top, there's a navigation bar with 'Your Account', 'HITs', and 'Qualifications' tabs. Below this is a banner stating 'Mechanical Turk is a marketplace for work.' and 'We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 562,959 HITs available. View them now.' The main content is divided into two columns. The left column, 'Make Money by working on HITs', describes HITs as Human Intelligence Tasks and lists benefits for workers: working from home, choosing work hours, and getting paid for good work. It includes a flowchart: 'Find an interesting task' (TASKS icon) -> 'Work' (gears icon) -> 'Earn money' (\$ icon), with a 'Find HITs Now' button. The right column, 'Get Results from Mechanical Turk Workers', asks workers to complete HITs and get results using Mechanical Turk. It lists benefits for requesters: access to a global workforce, fast completion, and pay only when satisfied. It includes a flowchart: 'Fund your account' (+ icon) -> 'Load your tasks' (tasks icon) -> 'Get results' (star icon), with a 'Get Started' button. At the bottom, there's a link to 'learn more about being a Worker'.

**amazonmechanicalturk**  
Artificial Intelligence

Your Account | HITs | Qualifications

Introduction | Dashboard | Status | Account Settings

**Mechanical Turk is a marketplace for work.**  
We give businesses and developers access to an on-demand, scalable workforce.  
Workers select from thousands of tasks and work whenever it's convenient.  
**562,959 HITs** available. [View them now.](#)

### Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

**As a Mechanical Turk Worker you:**

- Can work from home
- Choose your own work hours
- Get paid for doing good work

**Find an interesting task**   **Work**   **Earn money**

TASKS   Gears   \$

[Find HITs Now](#)

or [learn more about being a Worker](#)

### Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

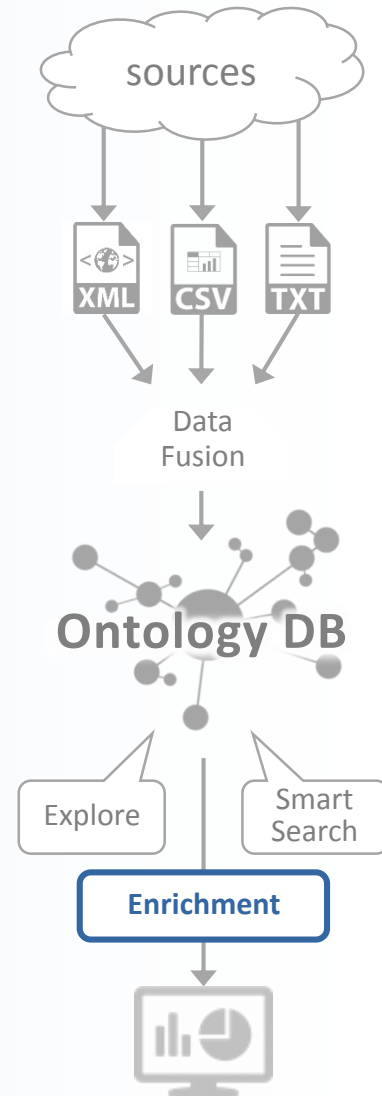
**As a Mechanical Turk Requester you:**

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

**Fund your account**   **Load your tasks**   **Get results**

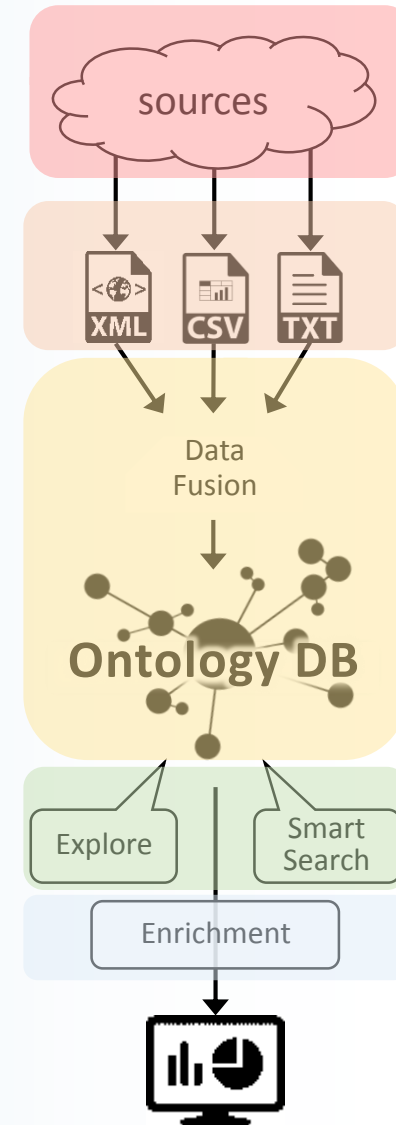
+   Tasks   ★

[Get Started](#)



# Team of Heroes - Technicians

Team Member	Role
Source finder	Locate relevant data sources, evaluate coverage and accuracy, ensure freedom to operate on the data
Data extractor	Extract relevant data from the sources into files (scraping, semantic web).
Modeler	Build the ontology, load the data into the DB, handle data cleansing, alignment and transformation. Resolve conflicts between sources
Analyst	Explore and process the data (run smart search, generate enrichment tasks)
Technicians	Perform enrichment tasks

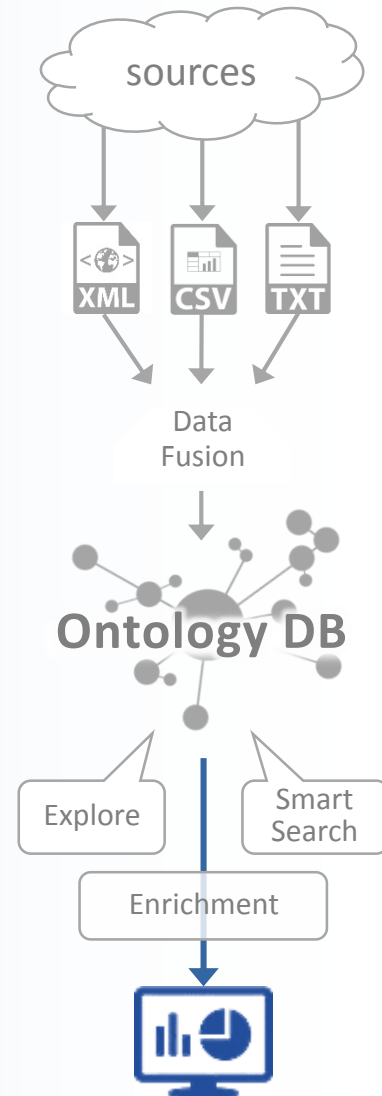


# Conveying Results

- Infographics help convey the investigation's findings in an effective and pleasant way
- There are many available tools for generating professional looking visualizations

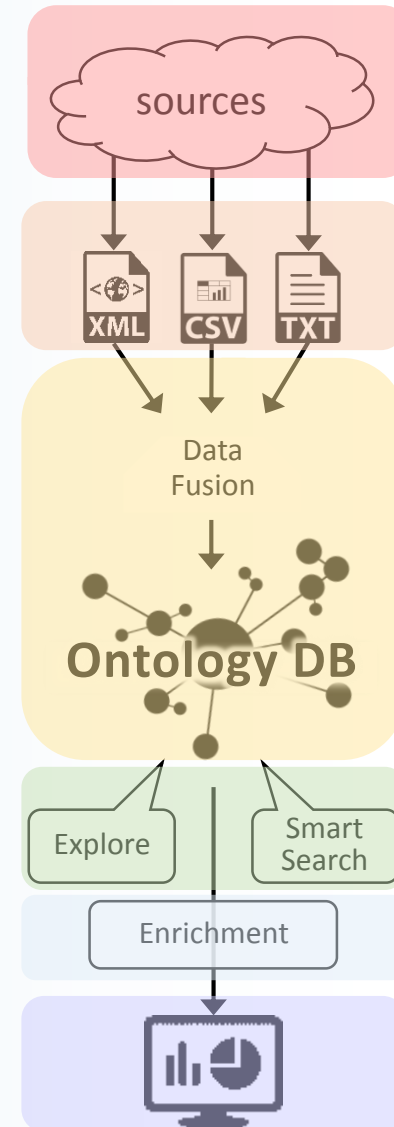


- Relevant data must be selected and exported in the format required by the visualization tool



# Team of Heroes

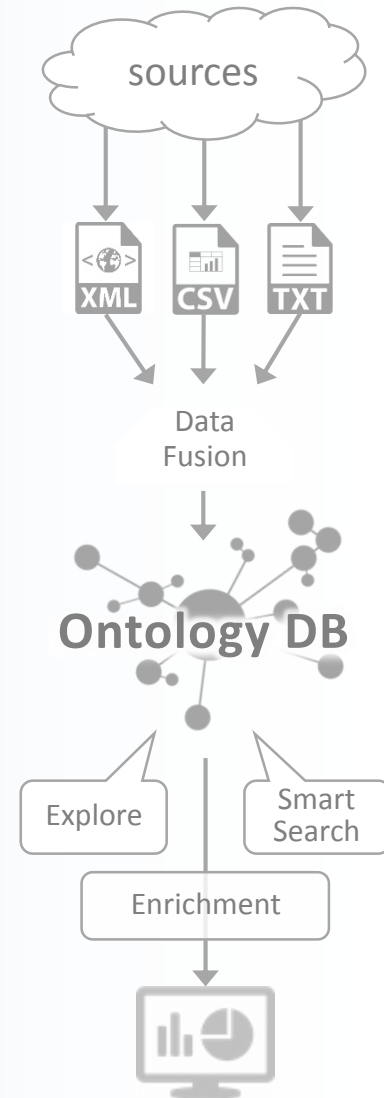
Team Member	Role
Source finder	Locate relevant data sources, evaluate coverage and accuracy, ensure freedom to operate on the data
Data extractor	Extract relevant data from the sources into files (scraping, semantic web).
Modeler	Build the ontology, load the data into the DB, handle data cleansing, alignment and transformation. Resolve conflicts between sources
Analyst	Explore and process the data (run smart search, generate enrichment tasks)
Technicians	Accomplish enrichment tasks
Visualization specialist	Suggest visualization, choose tools, export data, generate visualizations
Manager	<b>Orchestrate, handle crowd sourcing, monitor tasks, generate report</b>





# Conclusions

- New powerful tools and technologies transform the world of information science
- Generating knowledge vs. finding knowledge
- The need for professionalism
- Use of ad-hoc, decentralized team of professionals
- The information specialist becomes a team leader!





# Thank You