Web-Scale Data Integration in Life Sciences and Healthcare through Knowledge Graphs

Maulik R. Kamdar, PhD
Director of Clinical Informatics, Optum Health

Knowledge Graph Conference, May 2023

Disclosure

- Employee of Optum/UHG*

** This presentation is a review talk on the potential applications and challenges for knowledge graphs in healthcare and life sciences.*

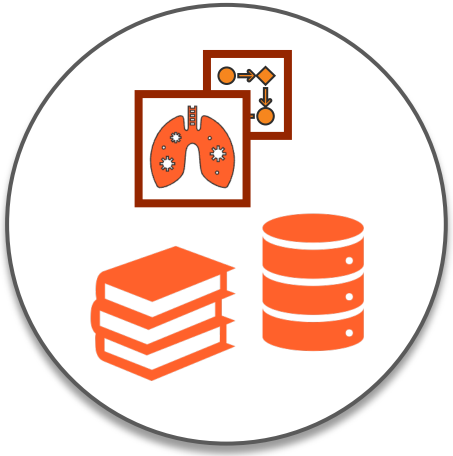
“

Doubling time of biomedical knowledge in 1950 was 50 years; in 1980, 7 years; in 2010, 3.5 years; in 2020, just 73 days; **and decreasing continuously after the COVID-19 pandemic**

Biomedical knowledge is expanding faster than the ability of professionals to aggregate, assimilate and apply it effectively during **education**, **patient care**, and **research**

1. Durack, DT. *The weight of medical knowledge."New England Journal of Medicine* (1978): 773-775.
2. Densen P. *Challenges and opportunities facing medical education.* Trans Am Clin Climatol Assoc. (2011):122:48-58.
3. Else, H. *How a torrent of COVID science changed research publishing--in seven charts.* Nature (2020): 553-554.
4. Brainard, J. *Scientists are drowning in COVID-19 papers. Can new tools keep them afloat. Science* (2020): 1126.

Explosion of Biomedical Knowledge and Data



- **Unstructured sources** (e.g., clinical trials, scientific publications, preprints, medical textbooks, flowcharts, guidelines, etc.)
- **Structured sources** (e.g., drug targets, biochemical pathways, epidemiology, etc.)

Kamdar, Maulik R., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.

Explosion of Biomedical Knowledge and Data

- **Unstructured sources** (e.g., clinical trials, scientific publications, preprints, medical textbooks, flowcharts, guidelines, etc.)
- **Structured sources** (e.g., drug targets, biochemical pathways, epidemiology, etc.)

- **Biological data** (e.g., genomics, proteomics, metabolomics, pharmacological data)
- **Patient data** (e.g., lab tests, comorbidities, diagnoses, administered drugs, image scans)
- **Sensor data** (e.g., physiological, wearable activity, environmental, social media)
- **Other sources (**e.g., chemical data, socio-economic determinants of health, etc.)

Kamdar, Maulik R., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.

Explosion of Biomedical Knowledge and Data

- **Unstructured sources** (e.g., clinical trials, scientific publications, preprints, medical textbooks, flowcharts, guidelines, etc.)
- **Structured sources** (e.g., drug targets, biochemical pathways, epidemiology, etc.)

- **Biological data** (e.g., genomics, proteomics, metabolomics, pharmacological data)
- **Patient data** (e.g., lab tests, comorbidities, diagnoses, administered drugs, image scans)
- **Sensor data** (e.g., physiological, wearable activity, environmental, social media)
- **Other sources (**e.g., chemical data, socio-economic determinants of health, etc.)

Kamdar, Maulik R., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.

Knowledge Graphs in Life Sciences and Healthcare

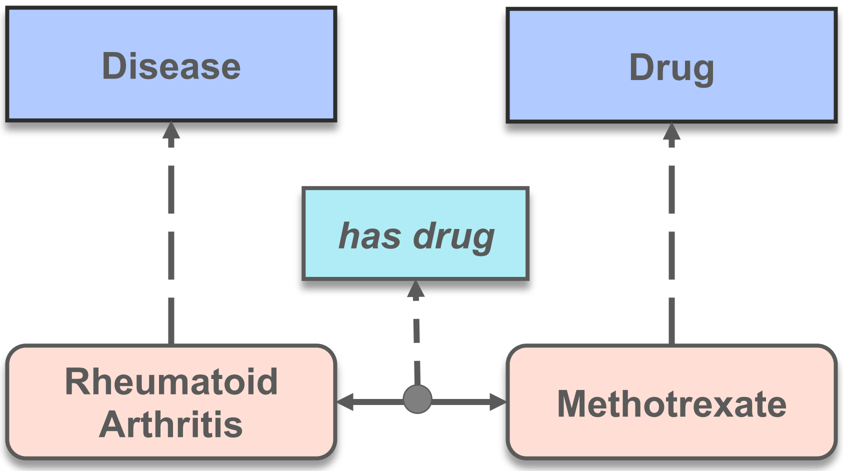
- ❑ The Hype and Historical Context behind Knowledge Graphs
- ❑ The Knowledge Graph Ecosystem – Tools, Techniques, and Methods
- ❑ Applications for Knowledge Graphs
- ❑ Challenges and Opportunities

The Hype and Historical Context behind Knowledge Graphs

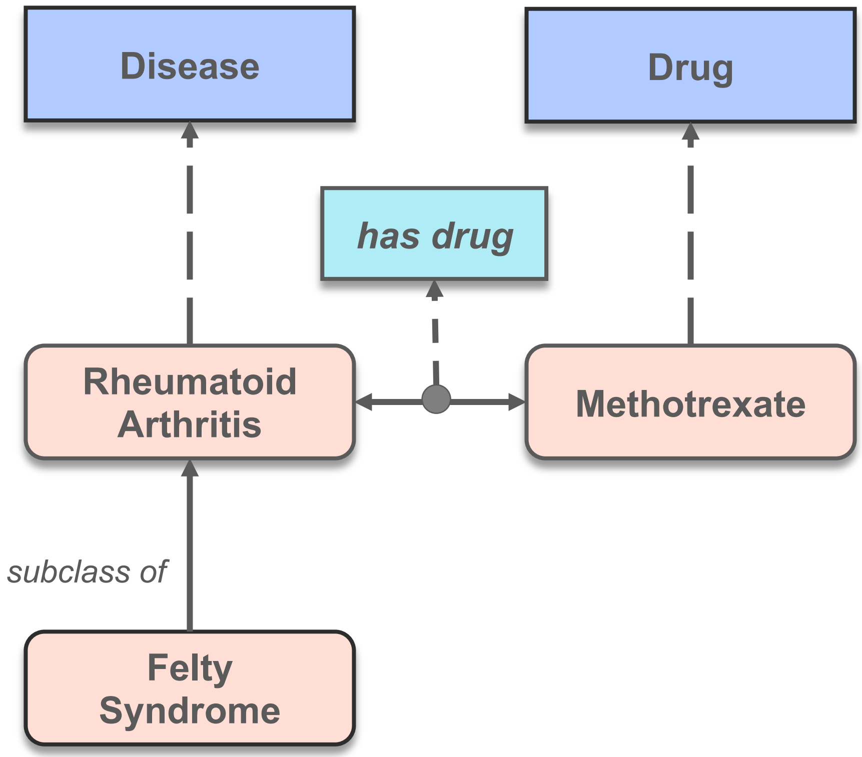
“Those who do not learn history are doomed to repeat it.

Biomedical Data and Knowledge as an Interconnected Graph

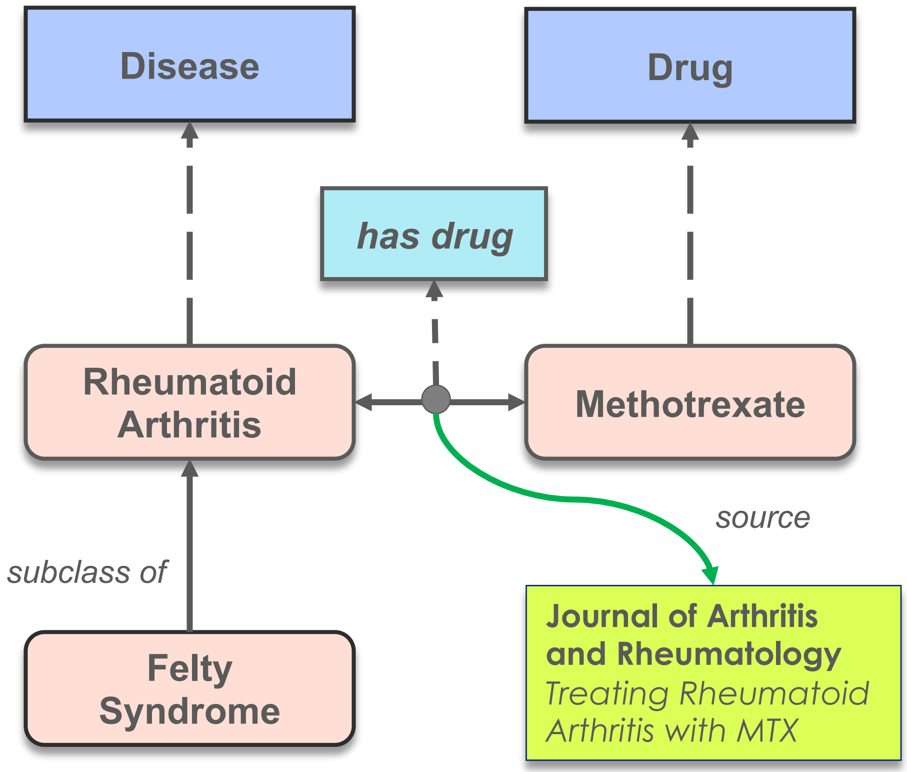
Biomedical Data and Knowledge as an Interconnected Graph



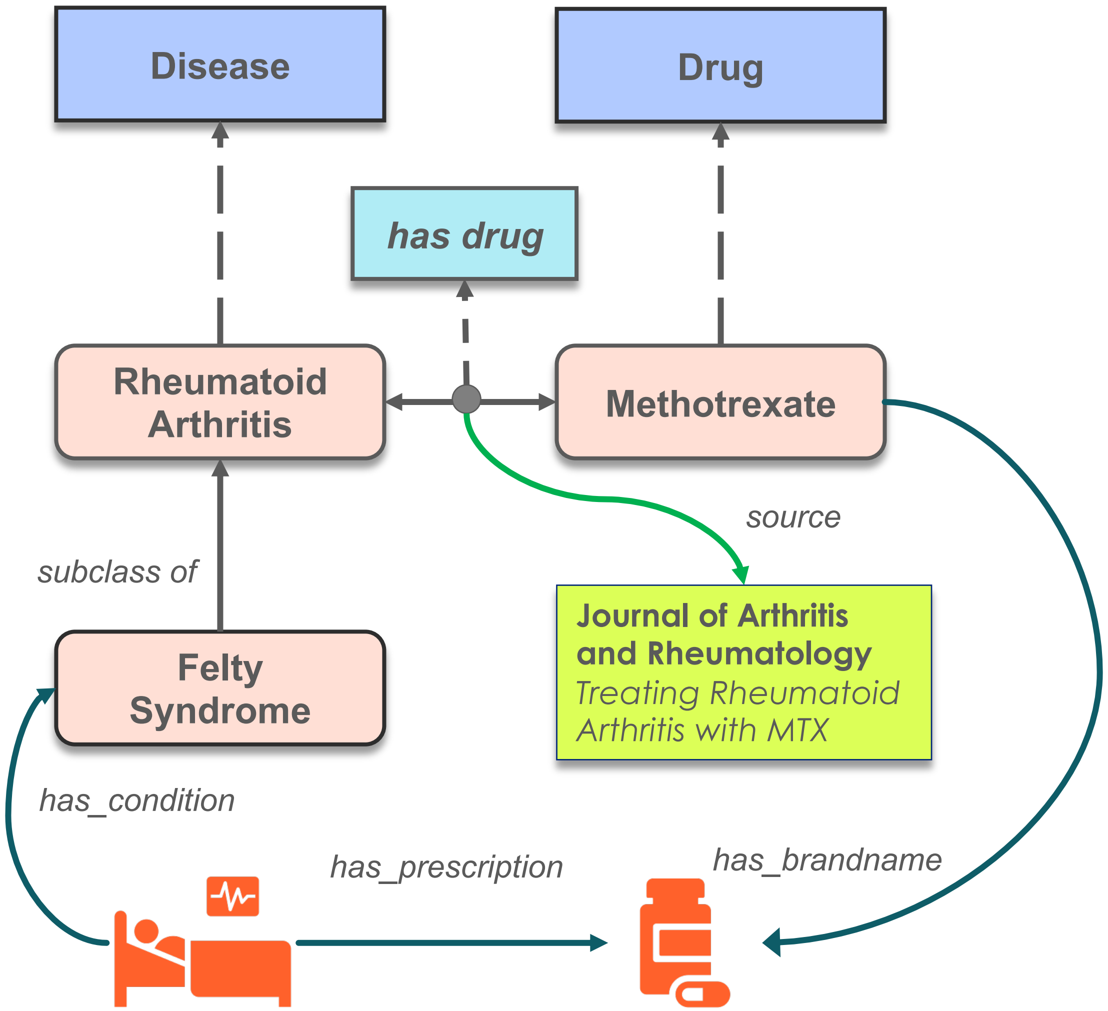
Biomedical Data and Knowledge as an Interconnected Graph



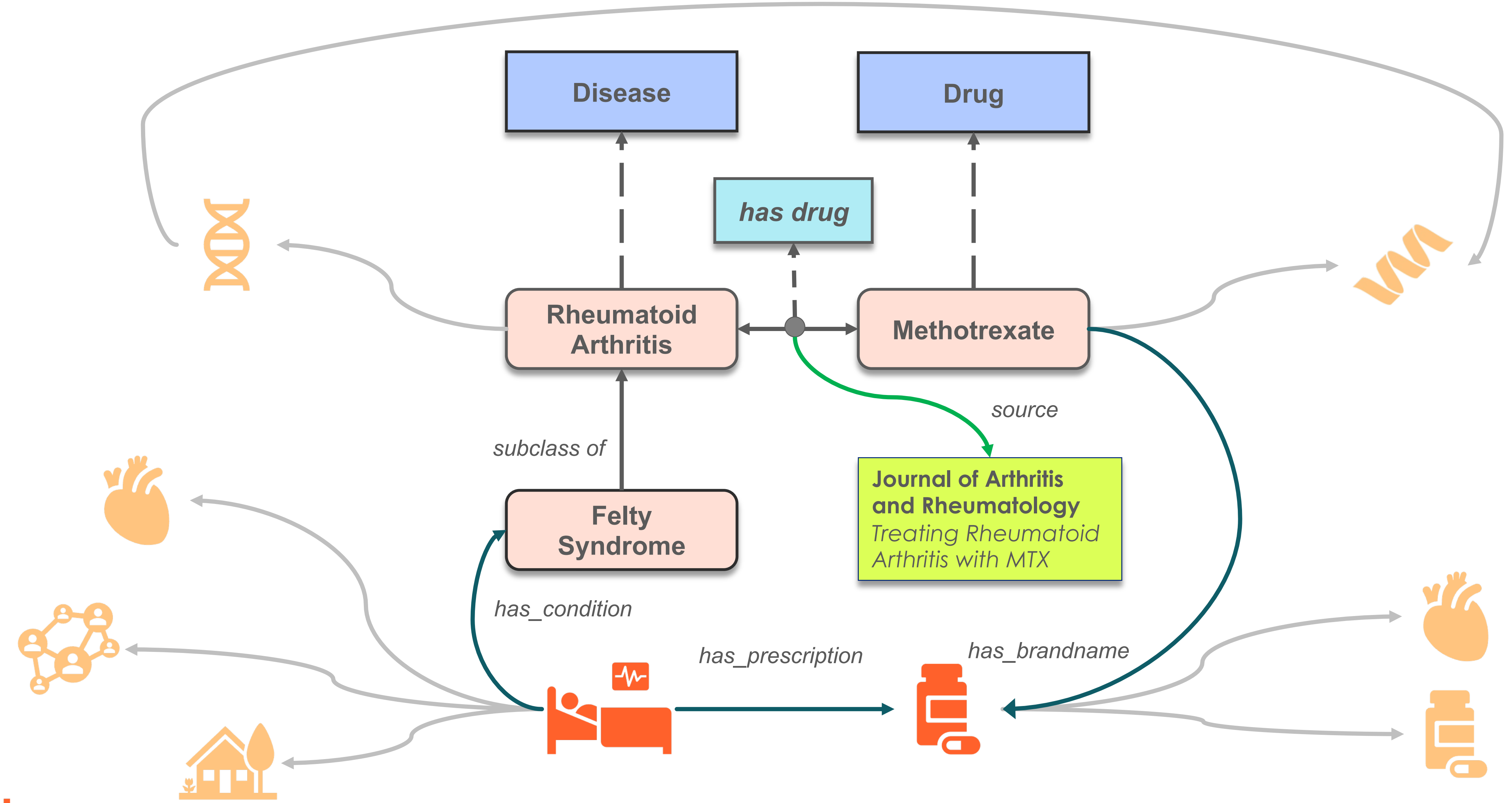
Biomedical Data and Knowledge as an Interconnected Graph



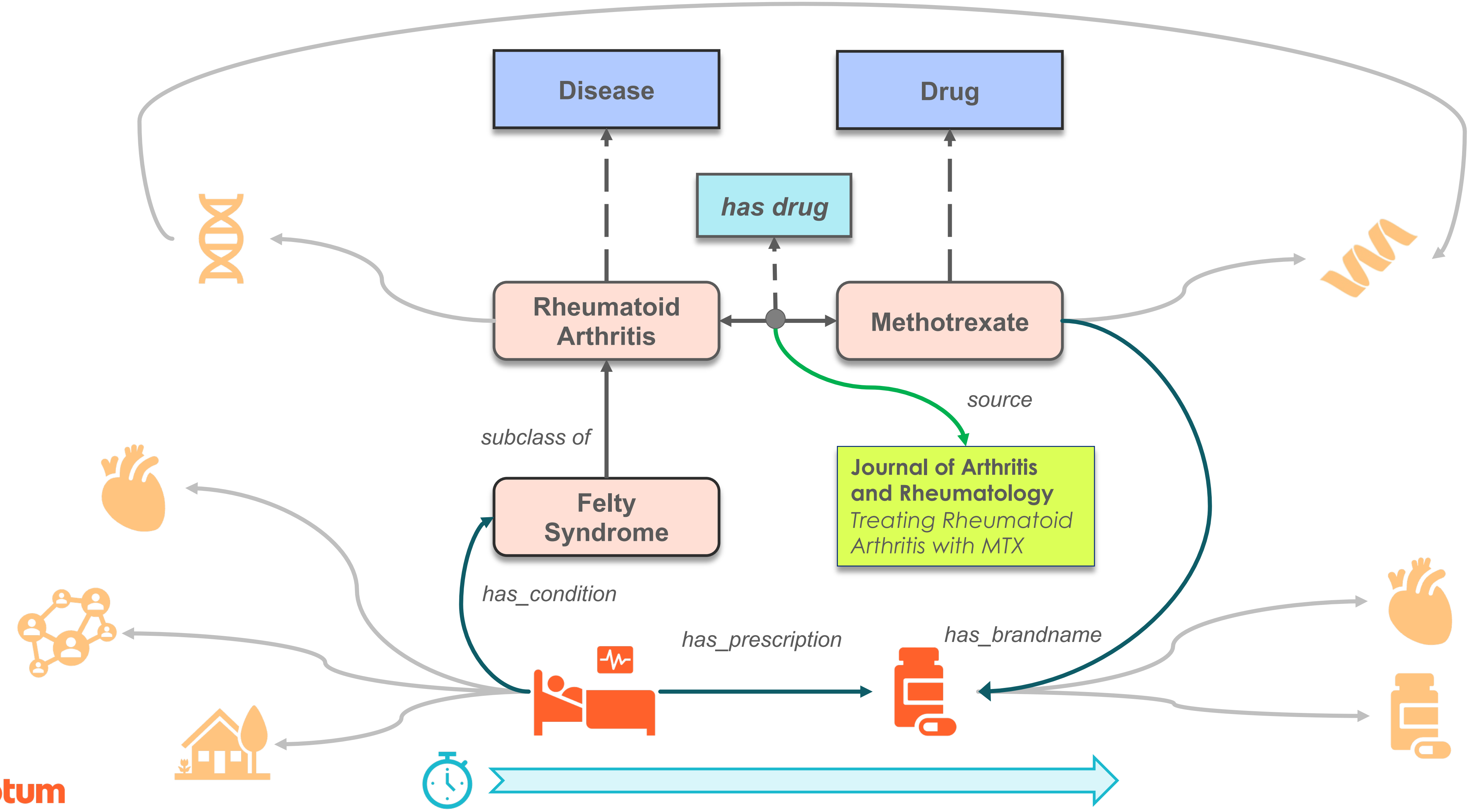
Biomedical Data and Knowledge as an Interconnected Graph



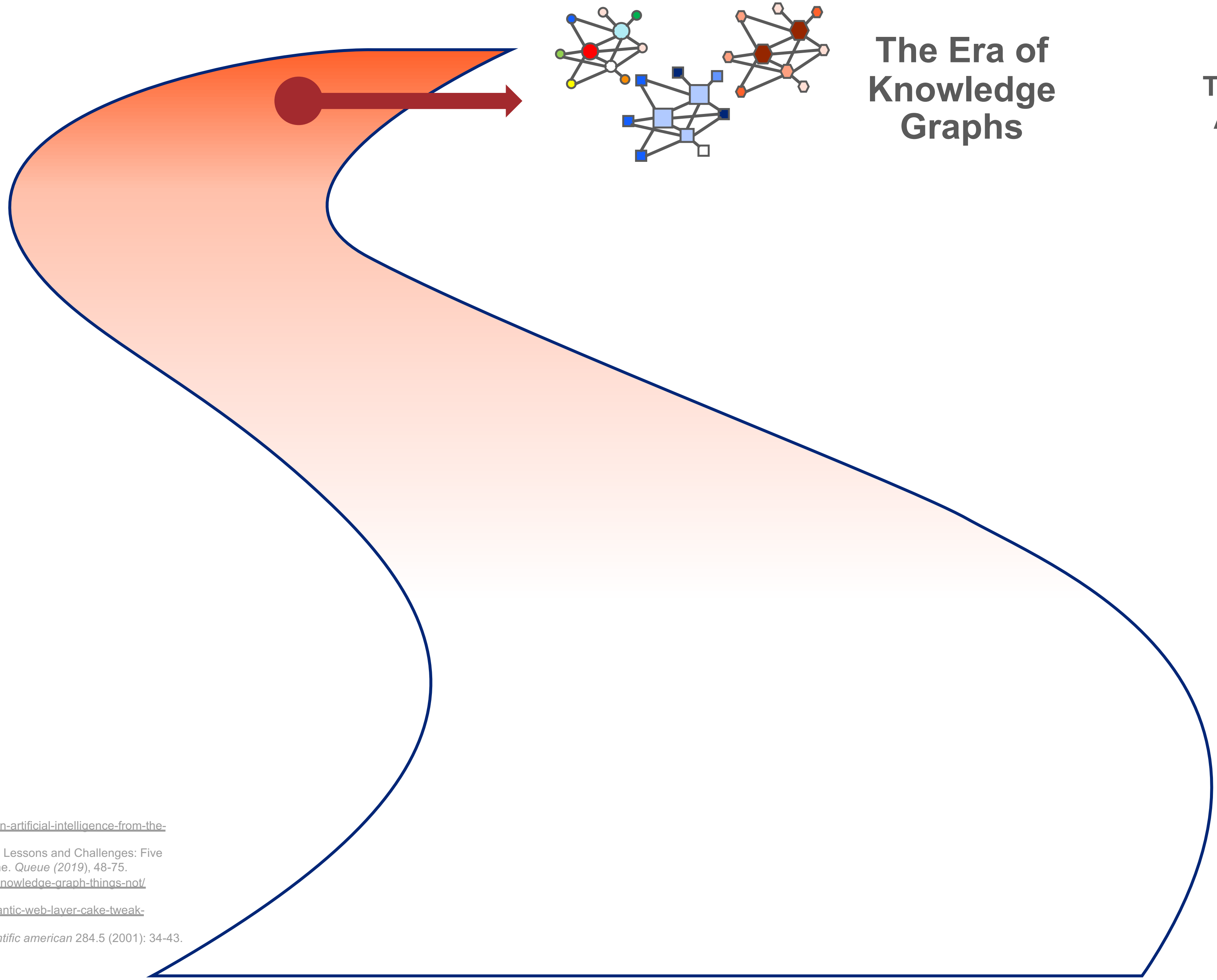
Biomedical Data and Knowledge as an Interconnected Graph



Biomedical Data and Knowledge as an Interconnected Graph



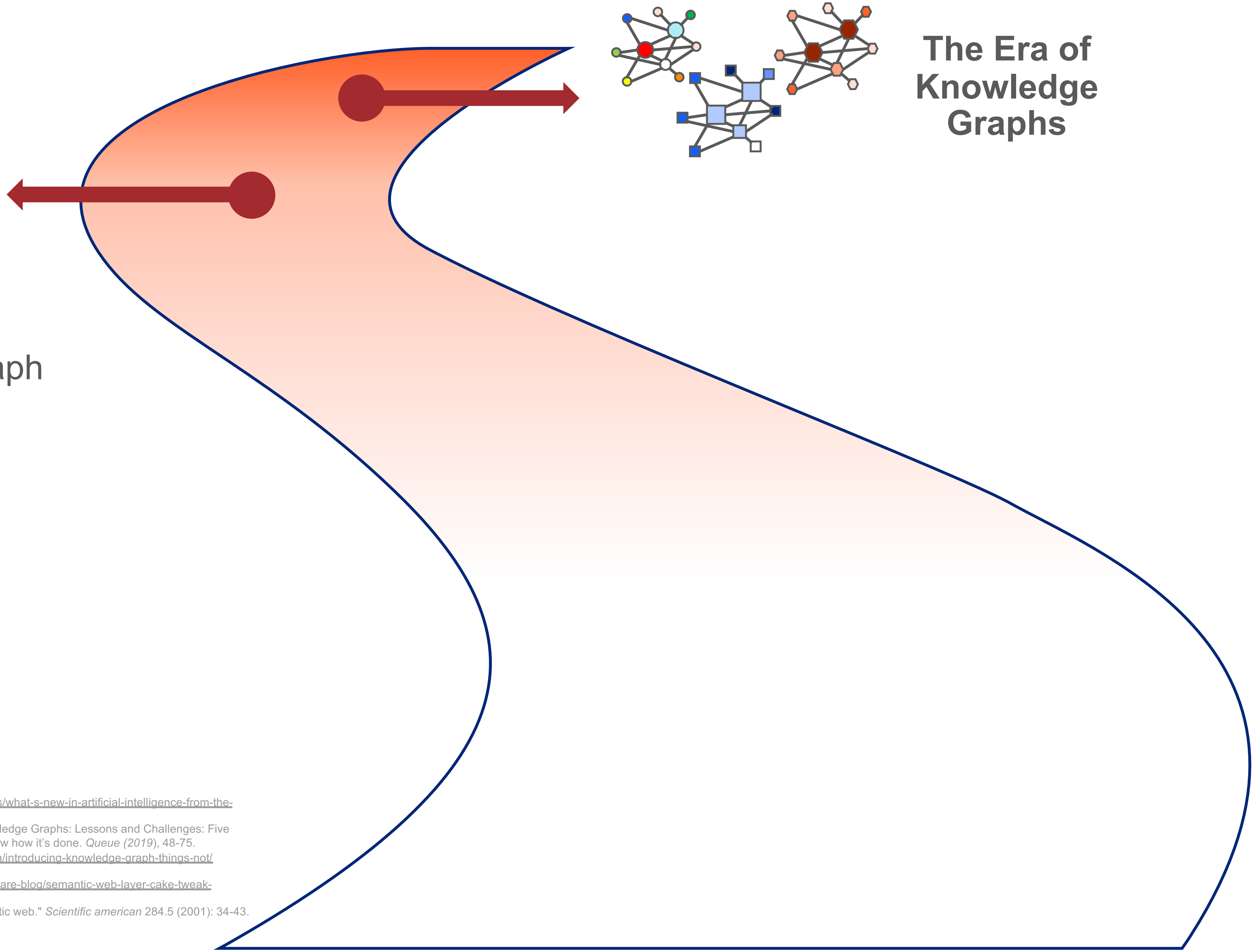
Optum



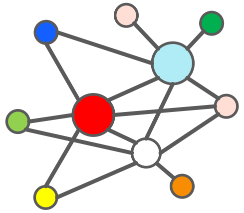
The Era of Knowledge Graphs

Emerging Technology for AI Innovation

1. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.
3. https://blog.google/products/search/introducing-knowledge-graph-things-not/
4. https://lod-cloud.net
5. https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab
6. Berners-Lee, T., et al. "The semantic web." *Scientific american* 284.5 (2001): 34-43.



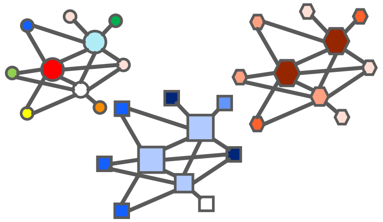
The Era of Knowledge Graphs



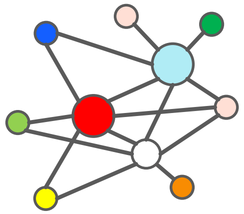
2012 –
Google
Knowledge Graph

1. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.
3. https://blog.google/products/search/introducing-knowledge-graph-things-not/
4. https://lod-cloud.net
5. https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab
6. Berners-Lee, T., et al. "The semantic web." *Scientific american* 284.5 (2001): 34-43.

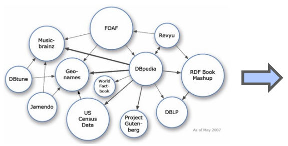
Optum

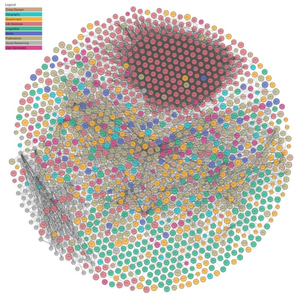


The Era of Knowledge Graphs



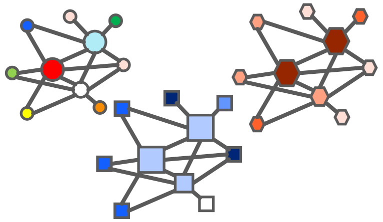
2012 –
Google
Knowledge Graph



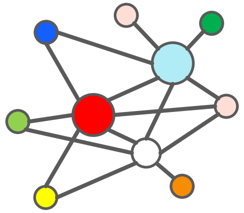


2006 –
Linked Open
Data Cloud

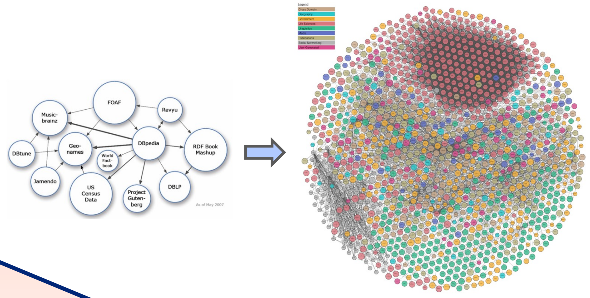
1. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.
3. https://blog.google/products/search/introducing-knowledge-graph-things-not/
4. https://lod-cloud.net
5. https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab
6. Berners-Lee, T., et al. "The semantic web." *Scientific american* 284.5 (2001): 34-43.



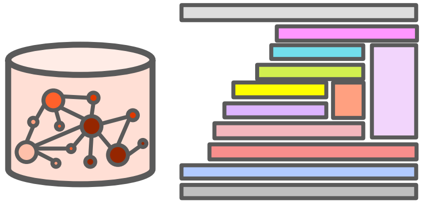
The Era of Knowledge Graphs



2012 –
Google Knowledge Graph

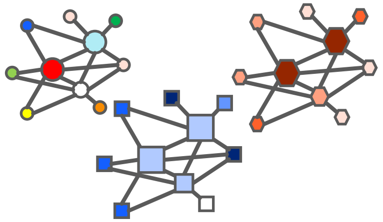


2006 –
Linked Open Data Cloud

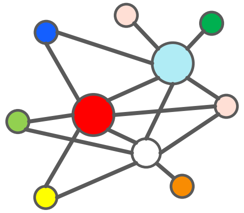


Graph and Semantic Web Technologies and Standards

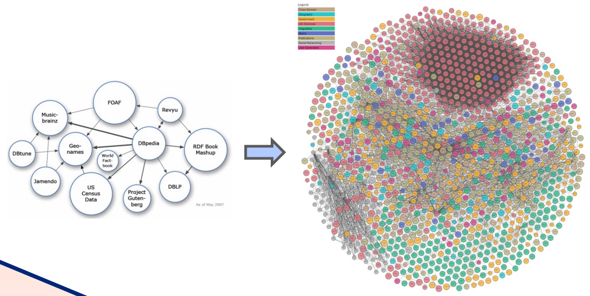
1. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.
3. https://blog.google/products/search/introducing-knowledge-graph-things-not/
4. https://lod-cloud.net
5. https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab
6. Berners-Lee, T., et al. "The semantic web." *Scientific american* 284.5 (2001): 34-43.



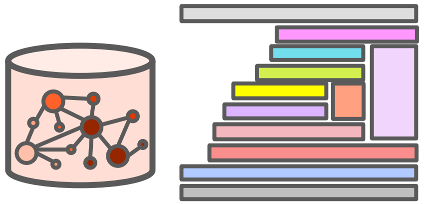
The Era of Knowledge Graphs



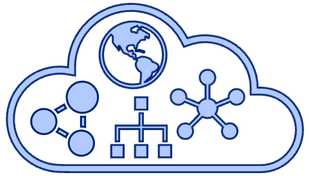
2012 –
Google Knowledge Graph



2006 –
Linked Open Data Cloud



Graph and Semantic Web Technologies and Standards

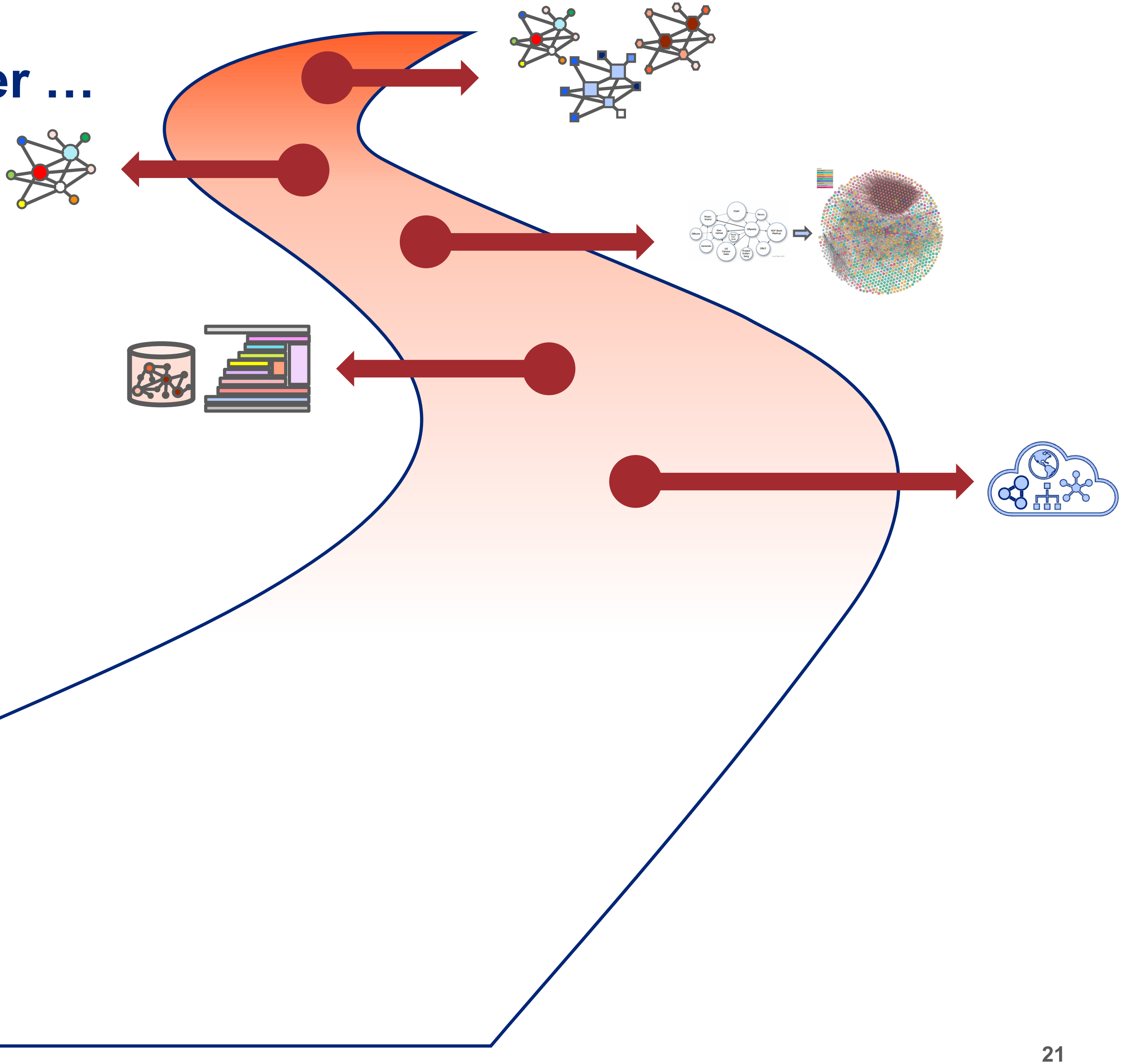


2001 –
The Semantic Web Vision

1. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.
3. https://blog.google/products/search/introducing-knowledge-graph-things-not/
4. https://lod-cloud.net
5. https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab
6. Berners-Lee, T., et al. "The semantic web." *Scientific american* 284.5 (2001): 34-43.

Biomedical AI has been here earlier …

- Since 1970s, investment cycles drove the development of **semantic networks** and **knowledge-based systems**



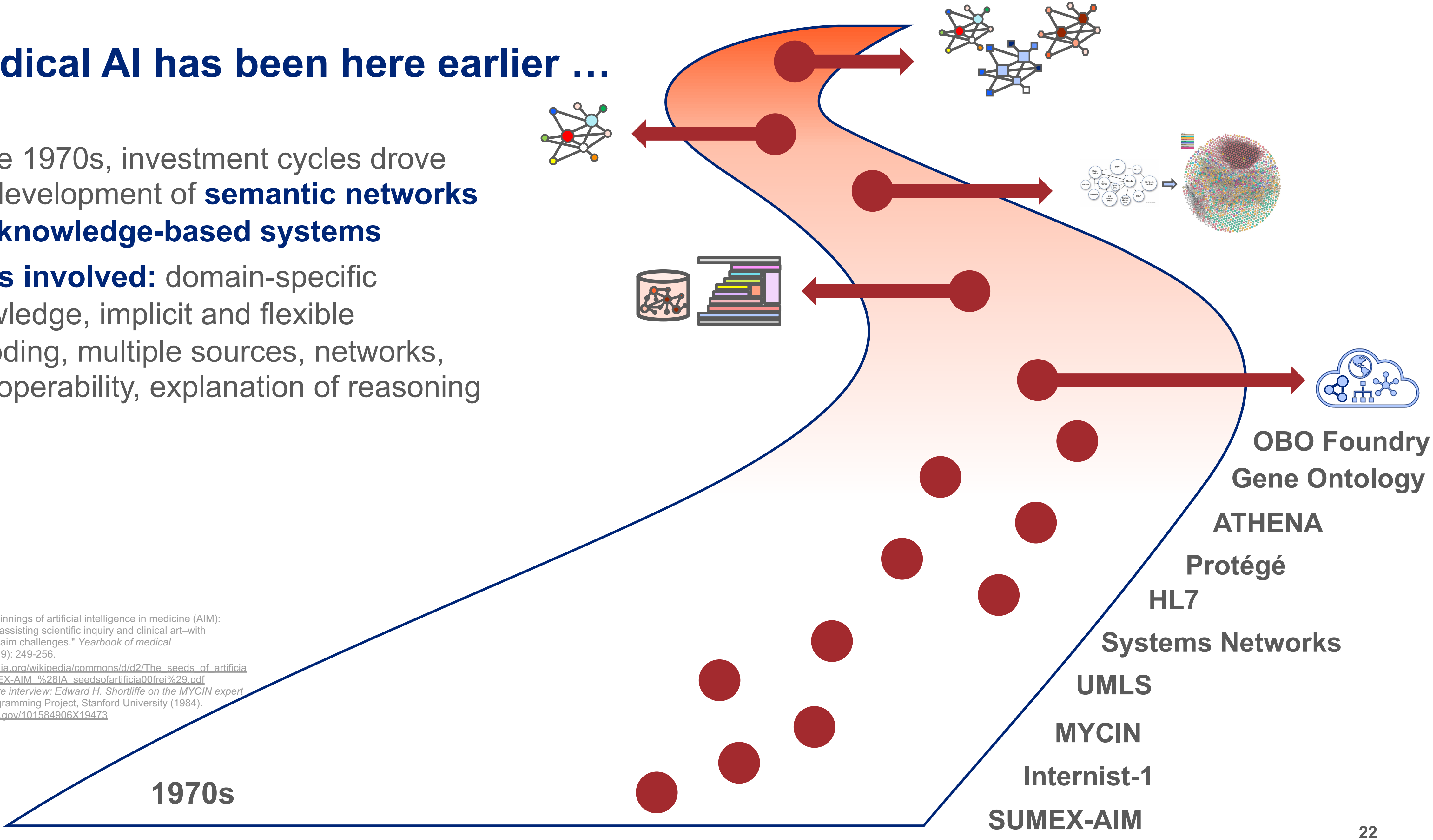
1970s

1. Kulikowski, C. A. "Beginnings of artificial intelligence in medicine (AIM): computational artifice assisting scientific inquiry and clinical art–with reflections on present aim challenges." *Yearbook of medical informatics* 28.01 (2019): 249-256.
2. https://upload.wikimedia.org/wikipedia/commons/d/d2/The_seeds_of_artificial_intelligence_-_SUMEX-AIM_%28IA_seedsofartificia00frei%29.pdf
3. Shortliffe, E. H. *Feature interview: Edward H. Shortliffe on the MYCIN expert system.* Heuristic Programming Project, Stanford University (1984).
4. https://profiles.nlm.nih.gov/101584906X19473

Optum

Biomedical AI has been here earlier …

- Since 1970s, investment cycles drove the development of **semantic networks** and **knowledge-based systems**
- **Ideas involved:** domain-specific knowledge, implicit and flexible encoding, multiple sources, networks, interoperability, explanation of reasoning

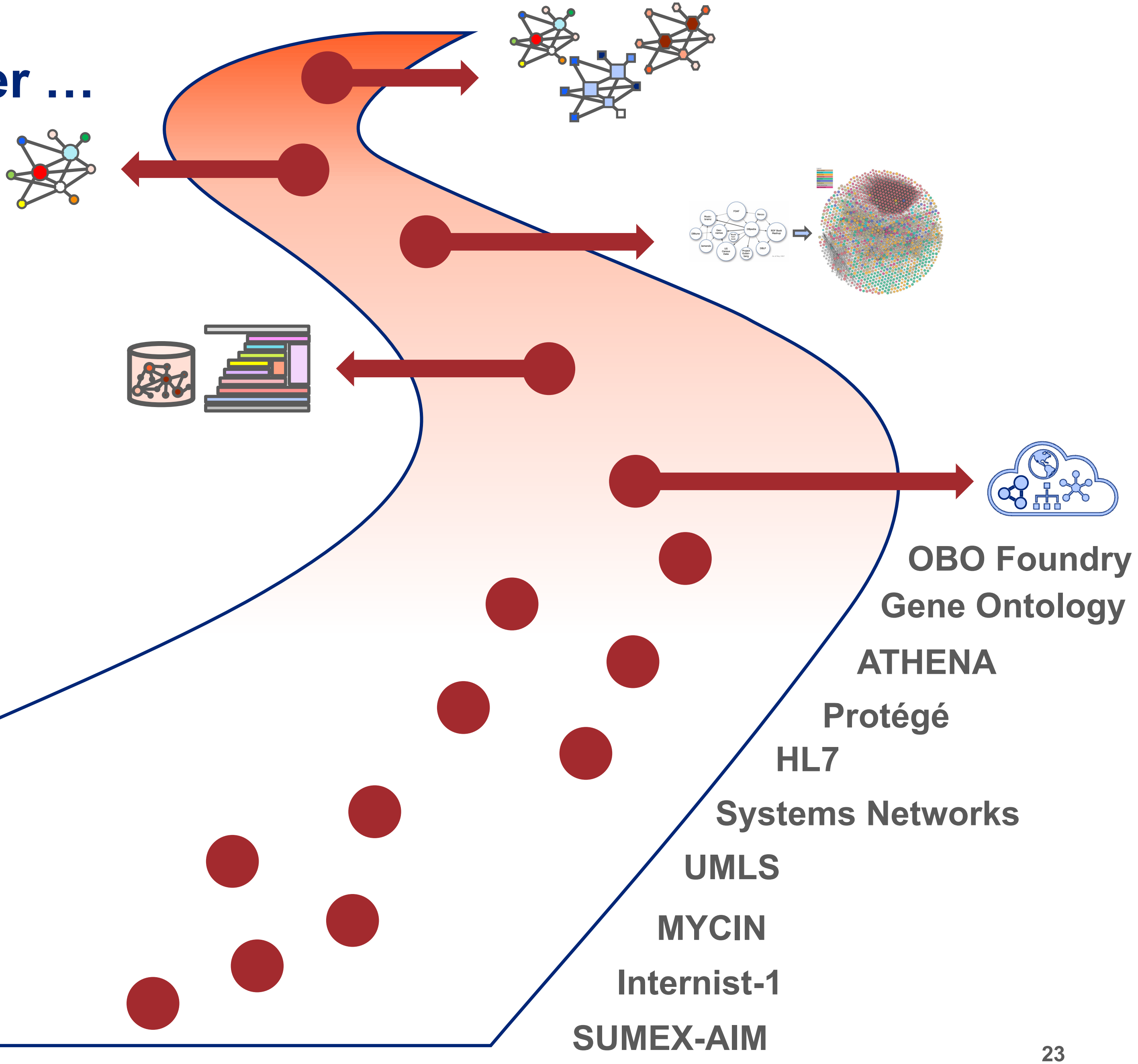


1. Kulikowski, C. A. "Beginnings of artificial intelligence in medicine (AIM): computational artifice assisting scientific inquiry and clinical art–with reflections on present aim challenges." *Yearbook of medical informatics* 28.01 (2019): 249-256.
2. https://upload.wikimedia.org/wikipedia/commons/d/d2/The_seeds_of_artificial_intelligence_-_SUMEX-AIM_%28IA_seedsofartificia00frei%29.pdf
3. Shortliffe, E. H. *Feature interview: Edward H. Shortliffe on the MYCIN expert system*. Heuristic Programming Project, Stanford University (1984).
4. https://profiles.nlm.nih.gov/101584906X19473

Optum

Biomedical AI has been here earlier …

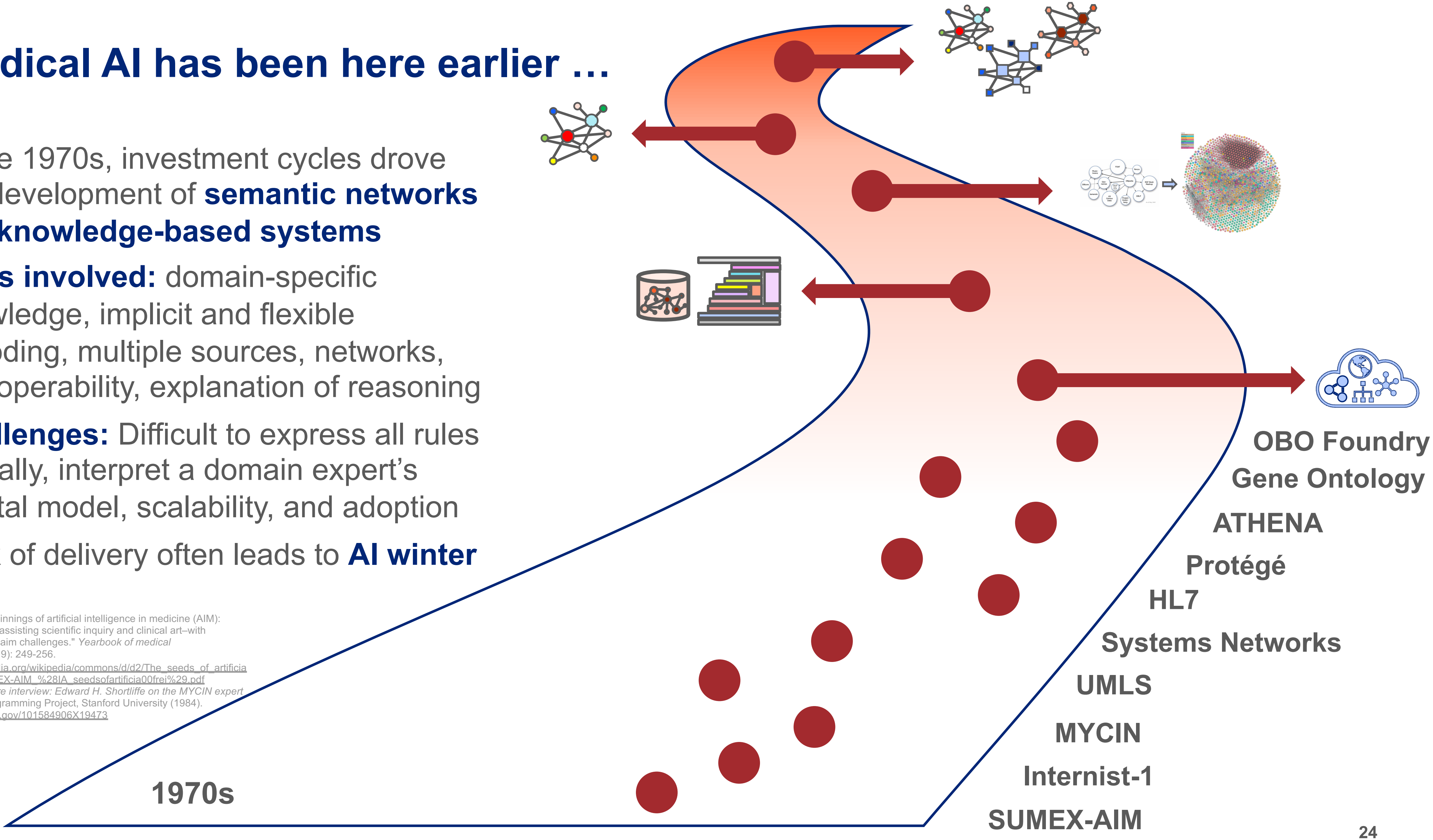
- Since 1970s, investment cycles drove the development of **semantic networks** and **knowledge-based systems**
- **Ideas involved:** domain-specific knowledge, implicit and flexible encoding, multiple sources, networks, interoperability, explanation of reasoning
- **Challenges:** Difficult to express all rules formally, interpret a domain expert's mental model, scalability, and adoption



1. Kulikowski, C. A. "Beginnings of artificial intelligence in medicine (AIM): computational artifice assisting scientific inquiry and clinical art–with reflections on present aim challenges." *Yearbook of medical informatics* 28.01 (2019): 249-256.
2. https://upload.wikimedia.org/wikipedia/commons/d/d2/The_seeds_of_artificial_intelligence_-_SUMEX-AIM_%28IA_seedsofartificia00frei%29.pdf
3. Shortliffe, E. H. *Feature interview: Edward H. Shortliffe on the MYCIN expert system.* Heuristic Programming Project, Stanford University (1984).
4. https://profiles.nlm.nih.gov/101584906X19473

Biomedical AI has been here earlier …

- Since 1970s, investment cycles drove the development of **semantic networks** and **knowledge-based systems**
- **Ideas involved:** domain-specific knowledge, implicit and flexible encoding, multiple sources, networks, interoperability, explanation of reasoning
- **Challenges:** Difficult to express all rules formally, interpret a domain expert's mental model, scalability, and adoption
- Lack of delivery often leads to **AI winter**



1. Kulikowski, C. A. "Beginnings of artificial intelligence in medicine (AIM): computational artifice assisting scientific inquiry and clinical art–with reflections on present aim challenges." *Yearbook of medical informatics* 28.01 (2019): 249-256.
2. https://upload.wikimedia.org/wikipedia/commons/d/d2/The_seeds_of_artificial_intelligence_-_SUMEX-AIM_%28IA_seedsofartificia00frei%29.pdf
3. Shortliffe, E. H. *Feature interview: Edward H. Shortliffe on the MYCIN expert system.* Heuristic Programming Project, Stanford University (1984).
4. https://profiles.nlm.nih.gov/101584906X19473

Optum

The Knowledge Graph Ecosystem – Tools, Techniques, and Methods

“ *It is not just about a graph database, but a graph ecosystem.*

Identify new sources with a constant eye out on the end use cases

Identify new sources

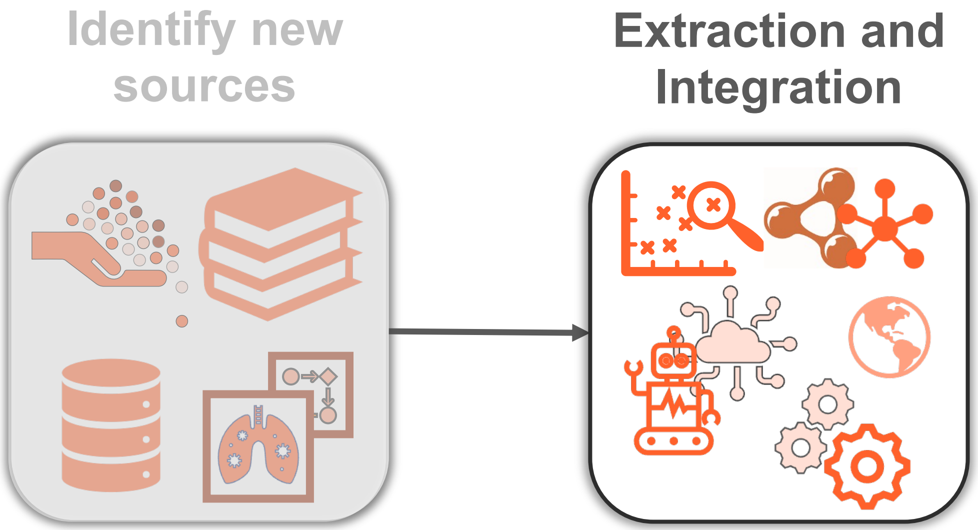
Applications

- **Popular sources** (e.g., SNOMEDCT, ICD-10, UMLS, UniProt, DrugBank)
- **Public/enterprise unstructured sources** (e.g., scientific papers)
- **Public/enterprise structured sources** (e.g., gene expression datasets)
- **Private secure sources** (e.g., patient records, search and browsing history)
- **Other sources** (e.g., common vernacular, slang terms, flowcharts etc.)

Optum

1. **Kamdar, Maulik R.**, et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.**, et al. "An empirical meta-analysis of the life sciences linked open data on the web." *Scientific data* 8.1 (2021): 24.

Identify the extraction and integration steps and processes



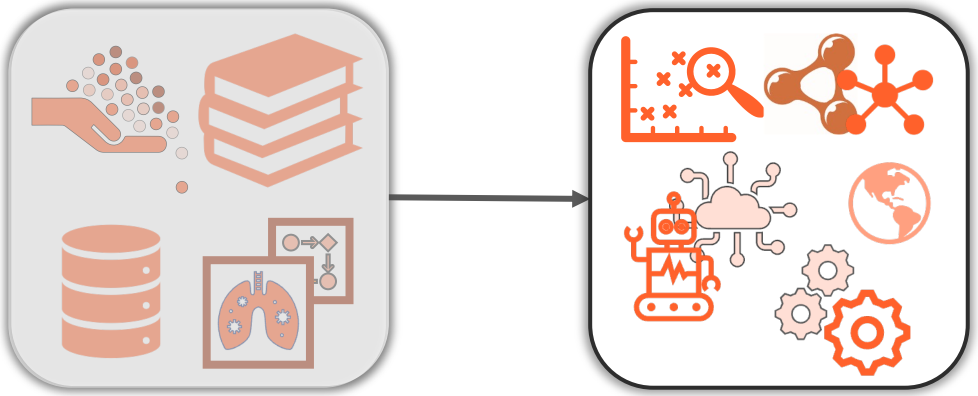
- **Classical integration approaches** (e.g., ETL pipelines)
- **Virtual integration** approaches (e.g., data federation, OMOP, FHIR)

1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. Saleem, M., **Kamdar M. R.**, et al. "Big linked cancer data: Integrating linked TCGA and Pubmed." *Journal of web semantics* 27 (2014): 34-41.
3. **Kamdar, Maulik R.**, et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* (2020): 288.

Identify the extraction and integration steps and processes

Identify new sources

Extraction and Integration



Diagnosis of Common Eye Diseases
… Conjunctivitis is the most common cause of red eye …

- **Classical integration approaches** (e.g., ETL pipelines)
- **Virtual integration** approaches (e.g., data federation, OMOP, FHIR)
- **Automated extraction** (e.g., ML/NLP approaches)

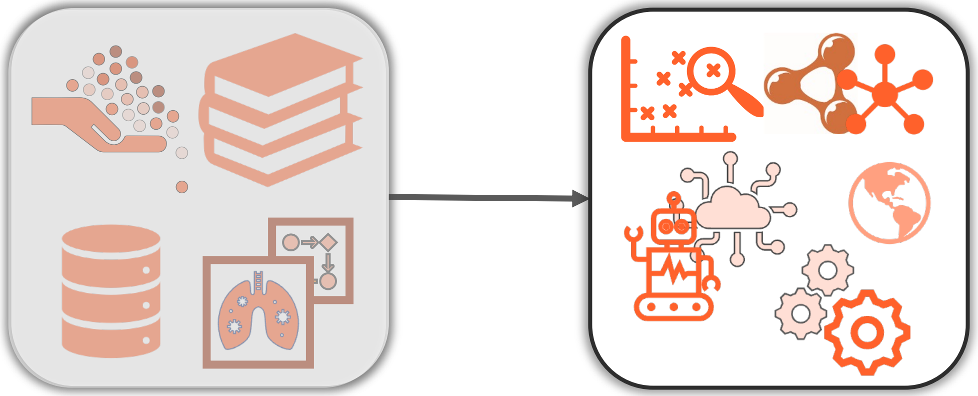
Optum

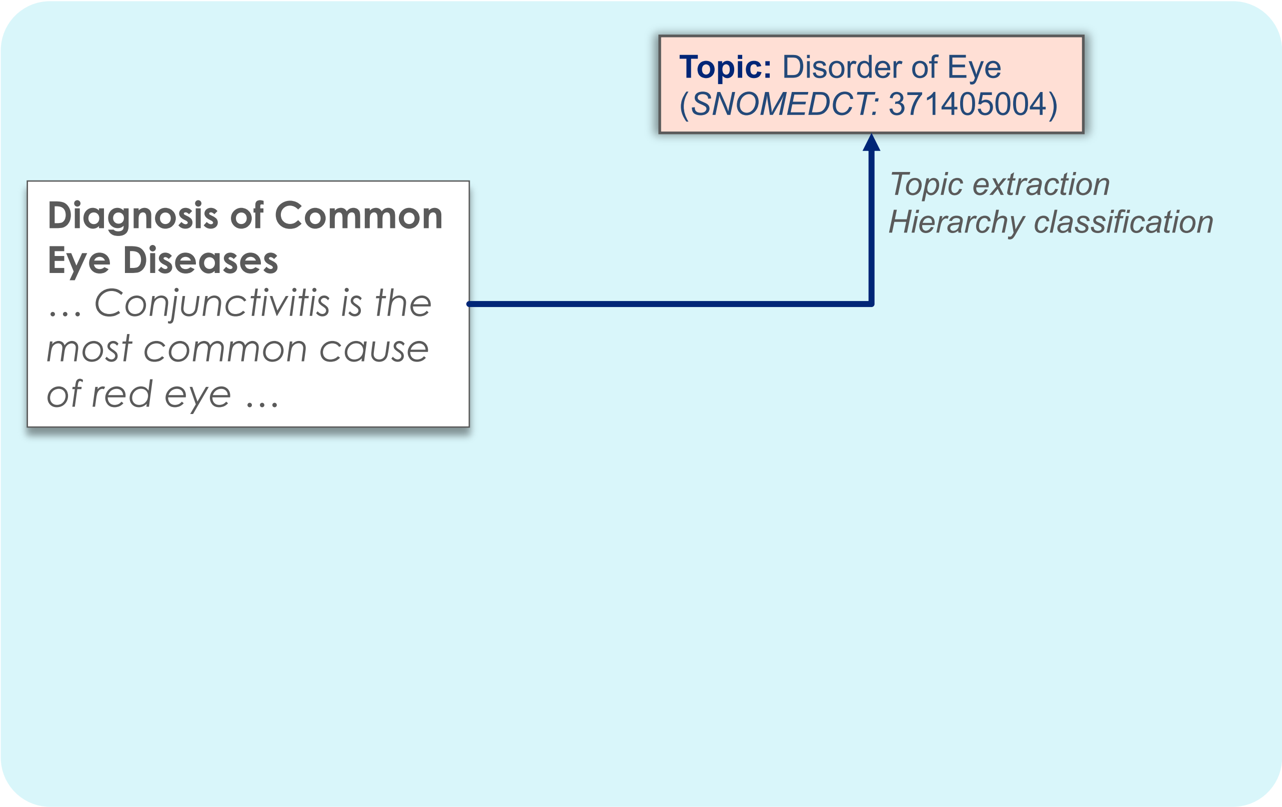
1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. Saleem, M., **Kamdar M. R.**, et al. "Big linked cancer data: Integrating linked TCGA and Pubmed." *Journal of web semantics* 27 (2014): 34-41.
3. **Kamdar, Maulik R.,** et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* (2020): 288.

Identify the extraction and integration steps and processes

Identify new sources

Extraction and Integration



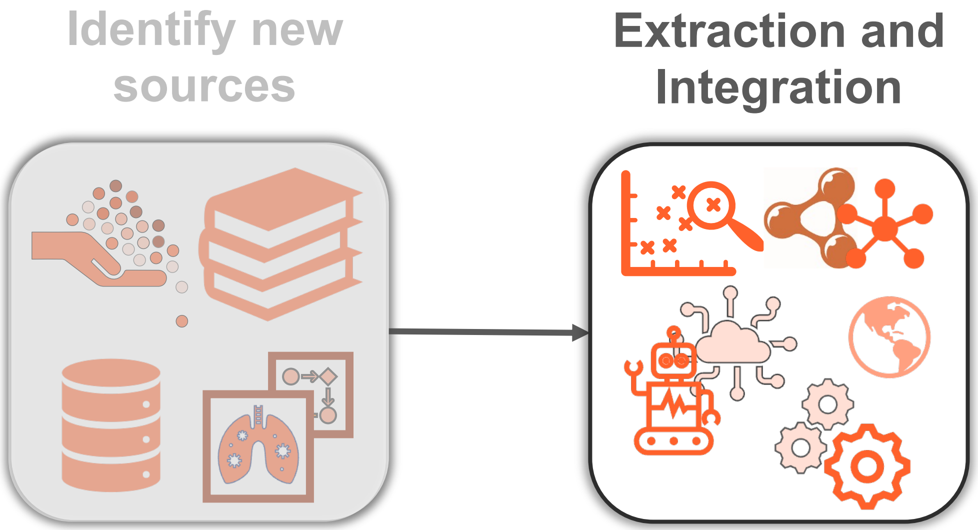


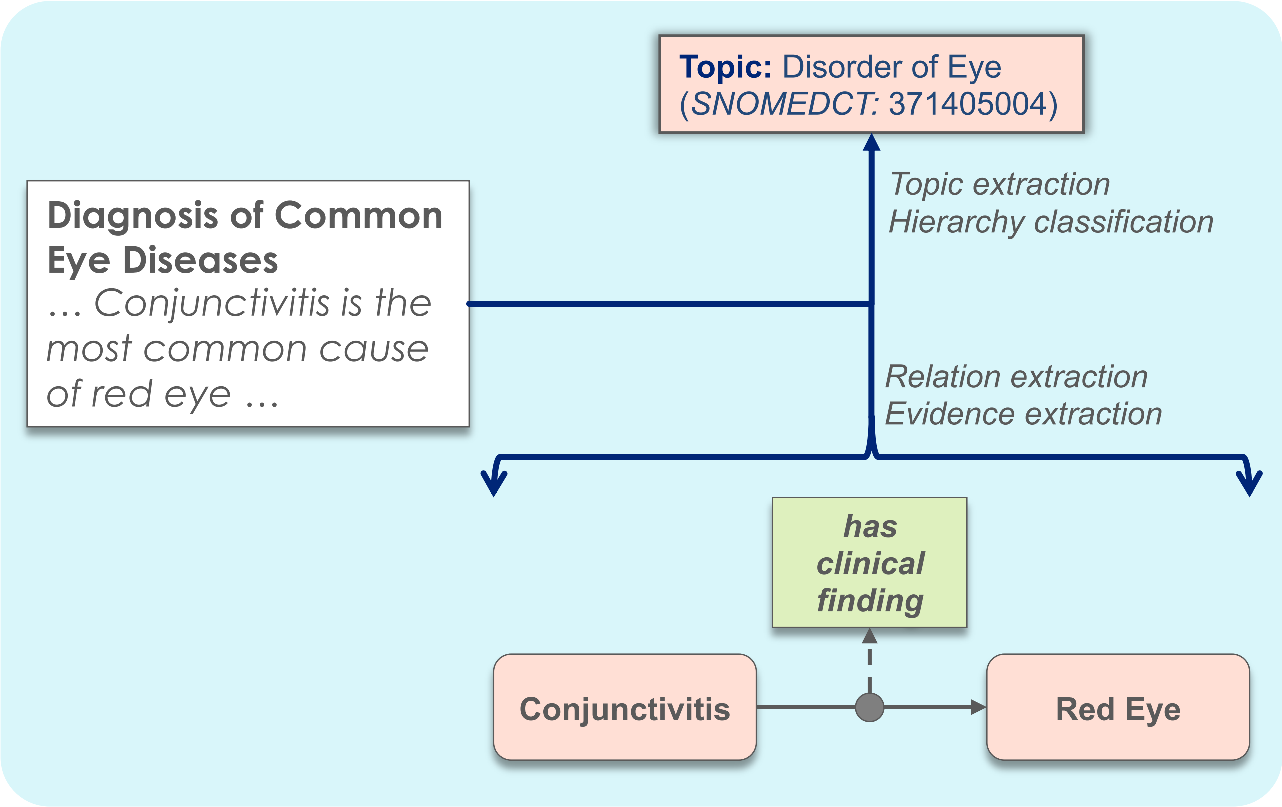
- **Classical integration approaches** (e.g., ETL pipelines)
- **Virtual integration** approaches (e.g., data federation, OMOP, FHIR)
- **Automated extraction** (e.g., ML/NLP approaches)

Optum

1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. Saleem, M., **Kamdar M. R.**, et al. "Big linked cancer data: Integrating linked TCGA and Pubmed." *Journal of web semantics* 27 (2014): 34-41.
3. **Kamdar, Maulik R.**, et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* (2020): 288.

Identify the extraction and integration steps and processes



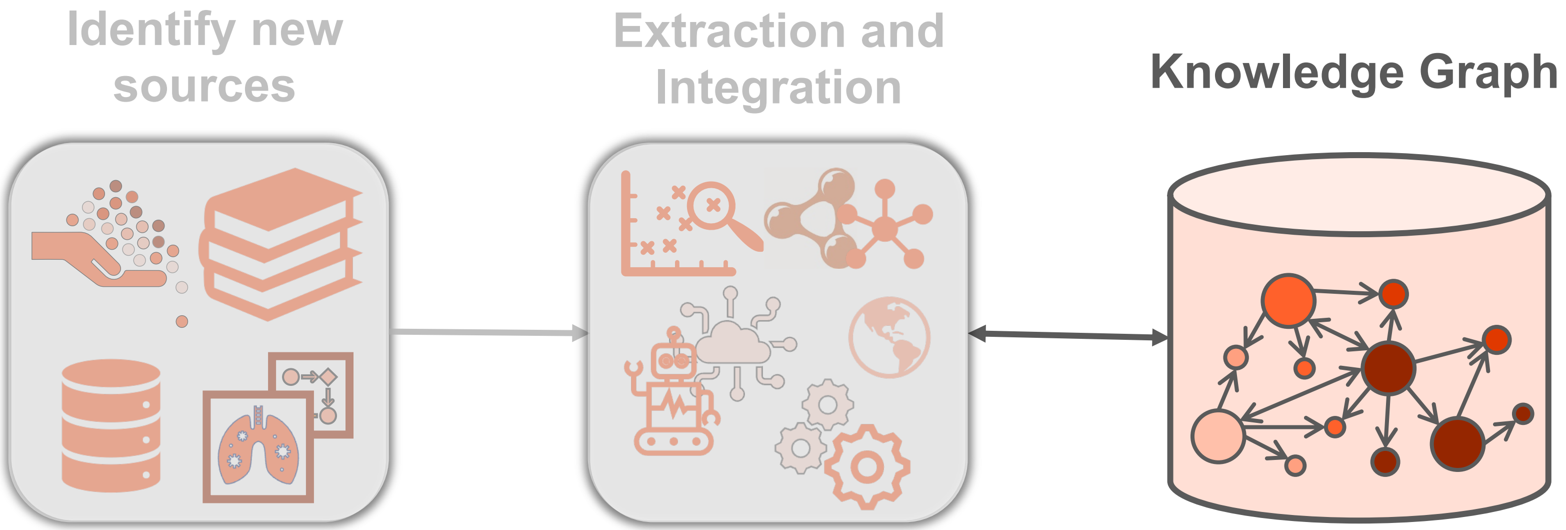


- **Classical integration approaches** (e.g., ETL pipelines)
- **Virtual integration** approaches (e.g., data federation, OMOP, FHIR)
- **Automated extraction** (e.g., ML/NLP approaches)

Optum

1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. Saleem, M., **Kamdar M. R.**, et al. "Big linked cancer data: Integrating linked TCGA and Pubmed." *Journal of web semantics* 27 (2014): 34-41.
3. **Kamdar, Maulik R.**, et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* (2020): 288.

Decide on the graph database and the model

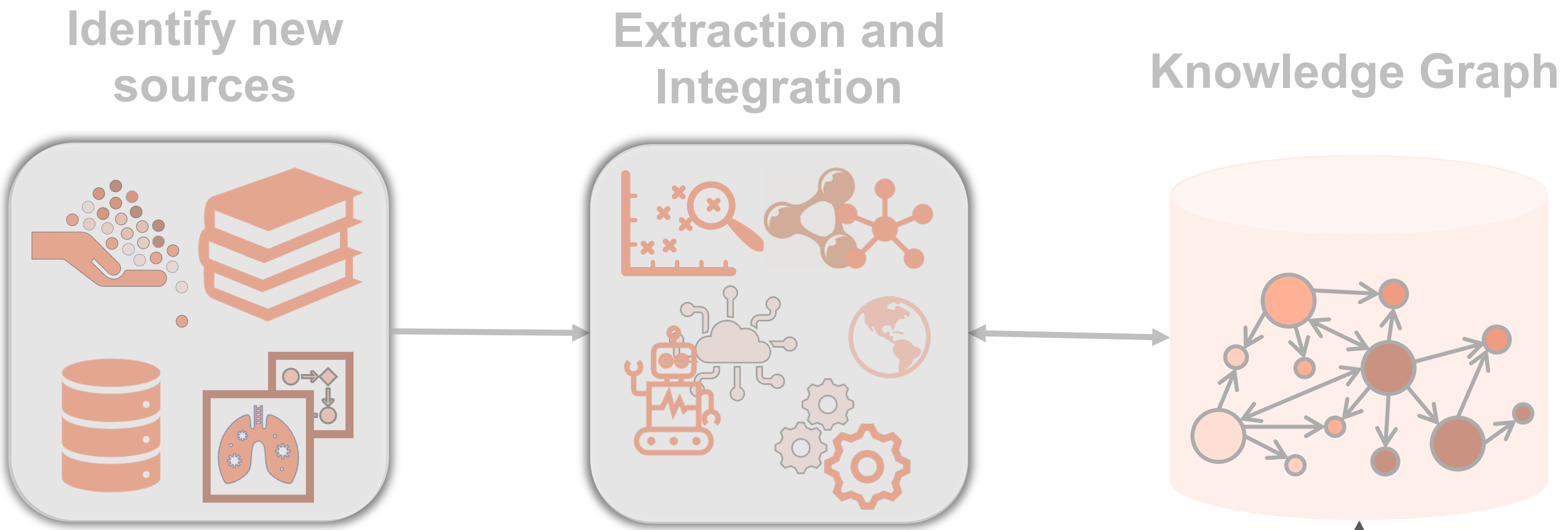


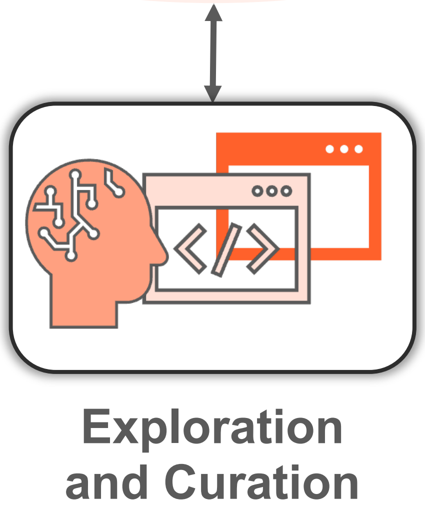
Several options for **scalable graph database** vendors:

- Labelled property graphs
- RDF triple stores

Optum

Identify how domain experts will explore and curate the graph

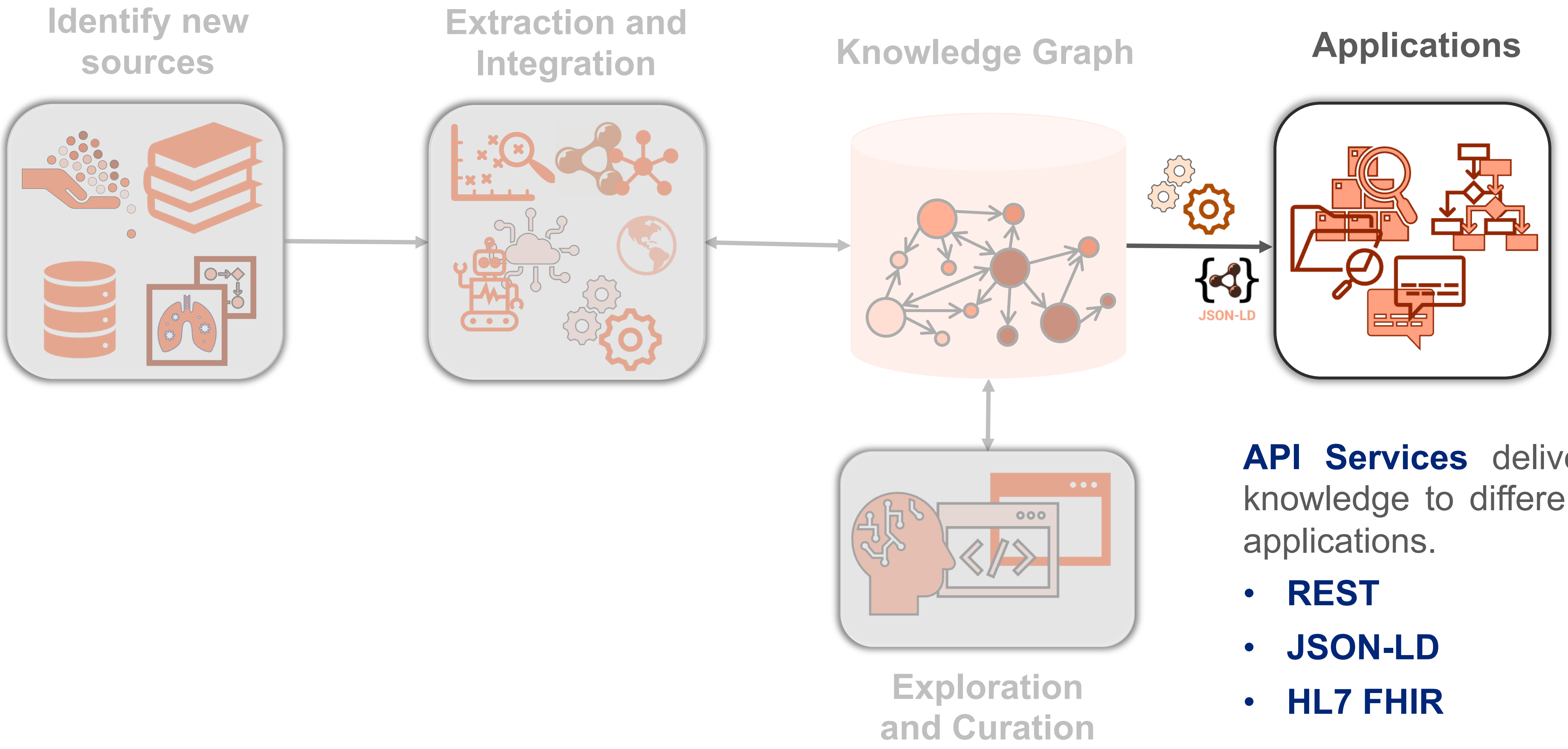




- **Terminology and ontology editors**
- **Visualization platforms**
- **Visual query systems**

1. **Kamdar, Maulik R.**, et al. "Interactive Exploration and Collaborative Curation of an Industry-Scale Healthcare Knowledge Graph Using the WebProtégé Cloud-Based Editor." Preprint (2022).
2. Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., **Kamdar, M.R.** et al. "The Reactome pathway knowledgebase." *Nucleic acids research* 42.D1 (2014): D472-D477.
3. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

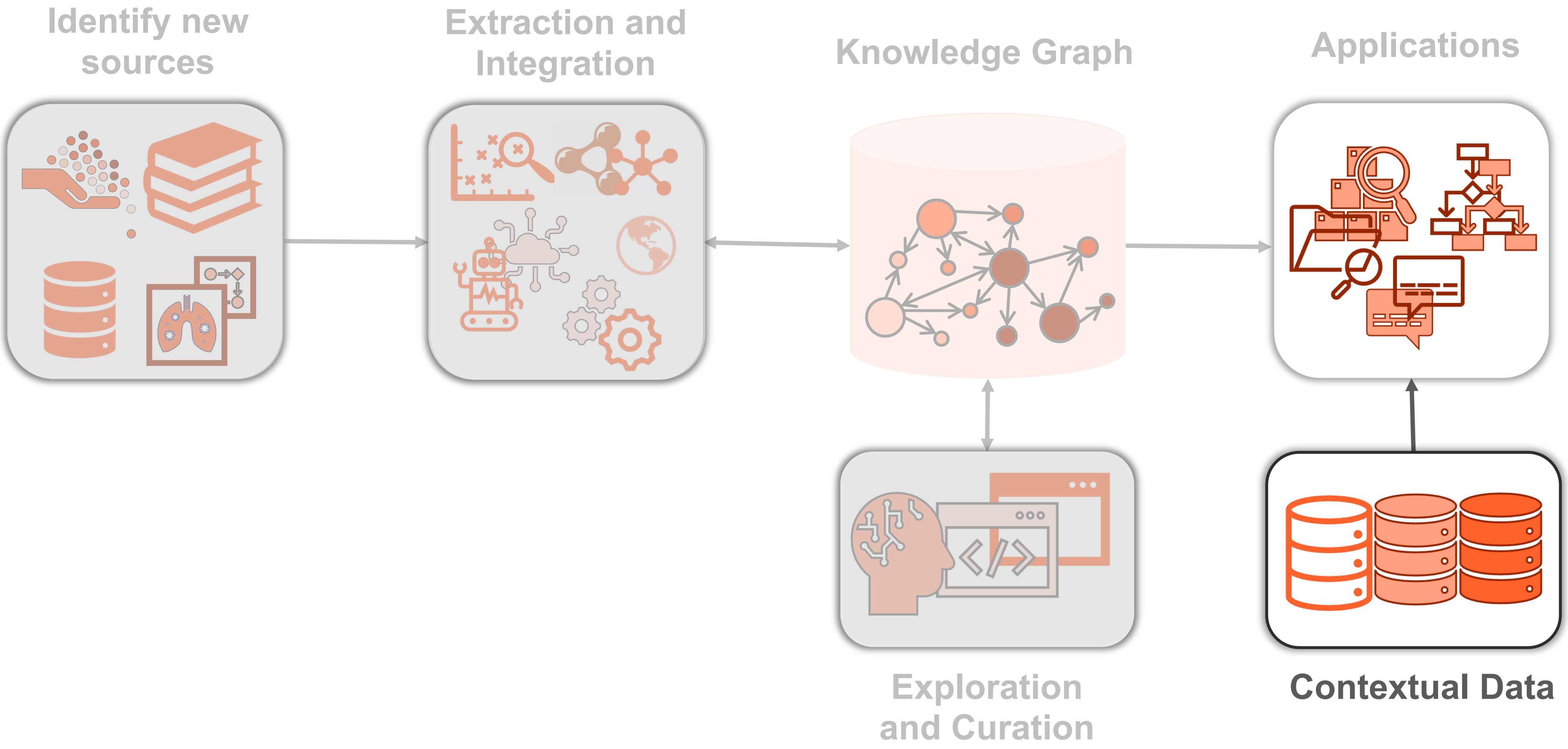
Identify how developers will consume the graph for applications



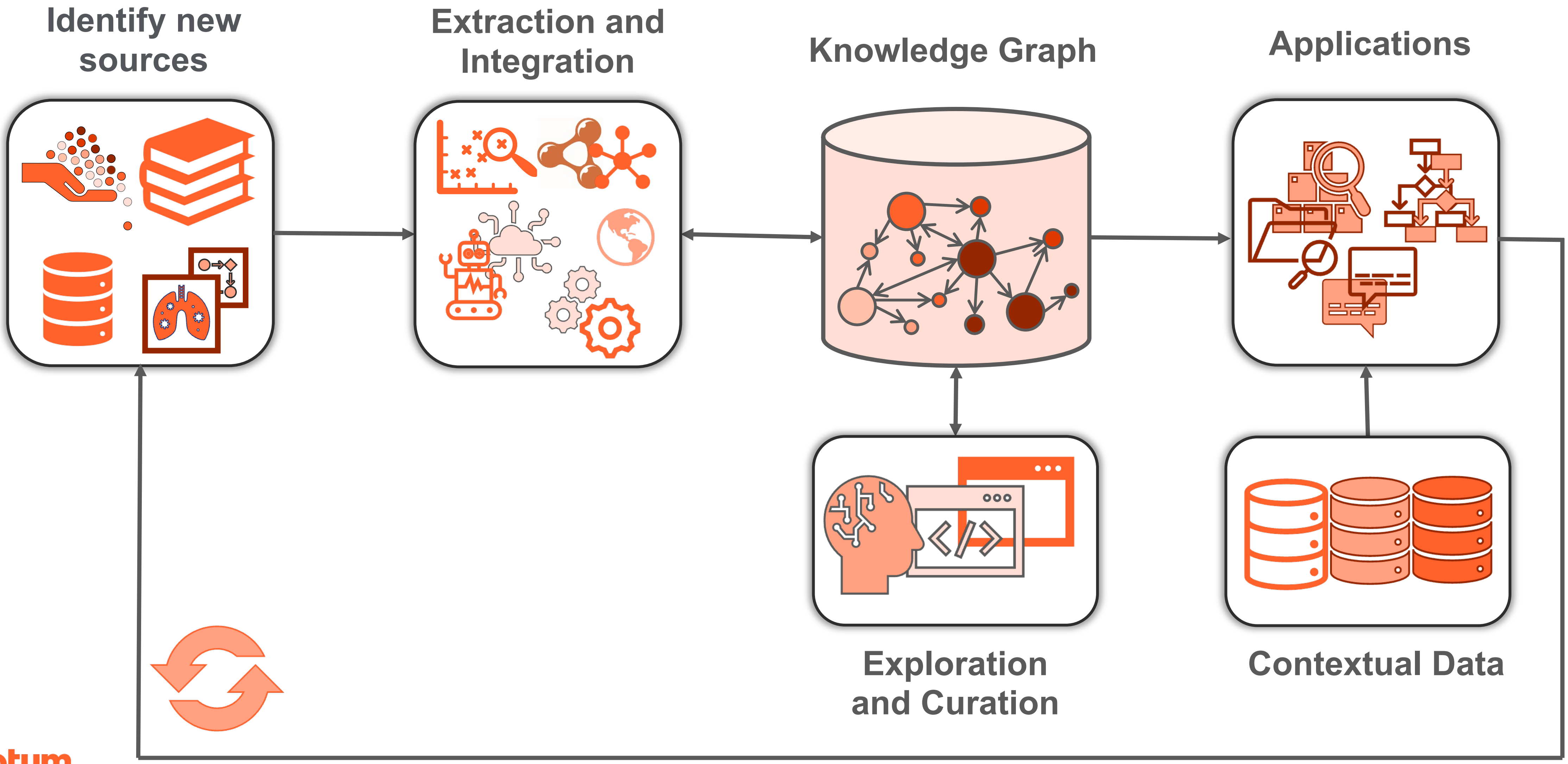
API Services deliver knowledge to different applications.

- **REST**
- **JSON-LD**
- **HL7 FHIR**

Identify any external contextual data sources which are needed



Continuously iterate as new applications and requirements emerge



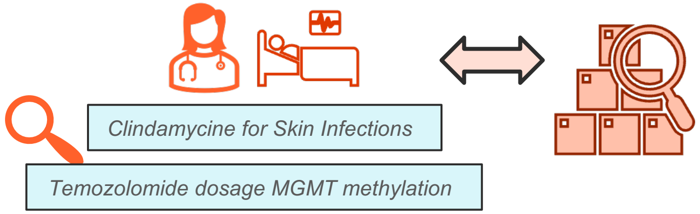
Optum

Applications for Knowledge Graphs in Life Sciences and Healthcare

“ *The world is your oyster, but … to a man with a hammer, everything looks like a nail*

Personalized Search and Recommendations for Patients, Providers, and Researchers

- **Document search:** Retrieving precise, updated, trustworthy information from documents for short queries



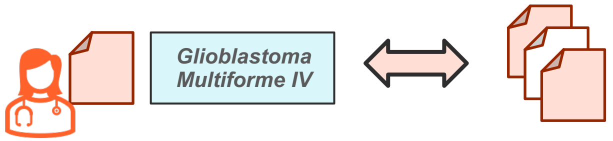
Optum

1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.,** et al. "A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search." *ISWC Industry Track* (2021).
3. **Kamdar, Maulik R.** "Elsevier's Healthcare Knowledge Graph: An Actionable Medical Knowledge Platform to Power Diverse Applications", *Knowledge Graph Conference* (2021).
4. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

Personalized Search and Recommendations for Patients, Providers, and Researchers

- **Document search:** Retrieving precise, updated, trustworthy information from documents for short queries

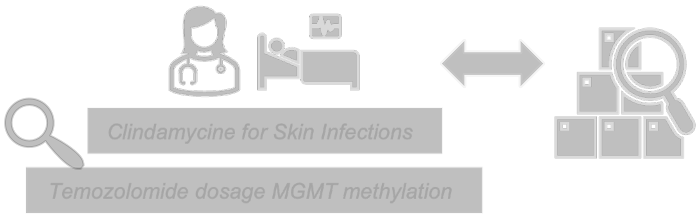
- **Document recommendation:** Browsing for information exploration based on related documents



1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.**, et al. "A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search." *ISWC Industry Track* (2021).
3. **Kamdar, Maulik R.** "Elsevier's Healthcare Knowledge Graph: An Actionable Medical Knowledge Platform to Power Diverse Applications", *Knowledge Graph Conference* (2021).
4. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

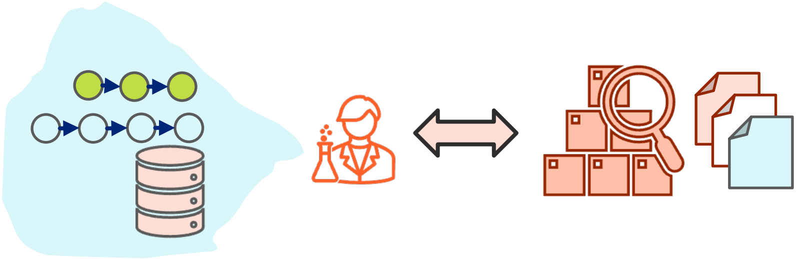
Personalized Search and Recommendations for Patients, Providers, and Researchers

- **Document search:** Retrieving precise, updated, trustworthy information from documents for short queries



- **Document recommendation:** Browsing for information exploration based on related documents

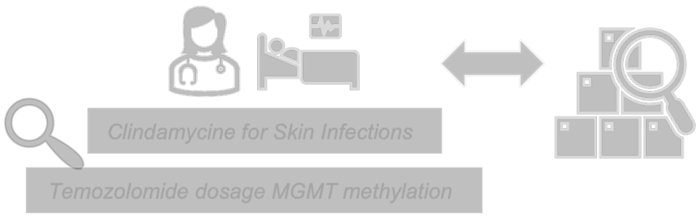
- **Personalized search and recommendations:** Use contextual information in a privacy-first approach



1. **Kamdar, Maulik R**., et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.**, et al. "A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search." *ISWC Industry Track* (2021).
3. **Kamdar, Maulik R.** "Elsevier's Healthcare Knowledge Graph: An Actionable Medical Knowledge Platform to Power Diverse Applications", *Knowledge Graph Conference* (2021).
4. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

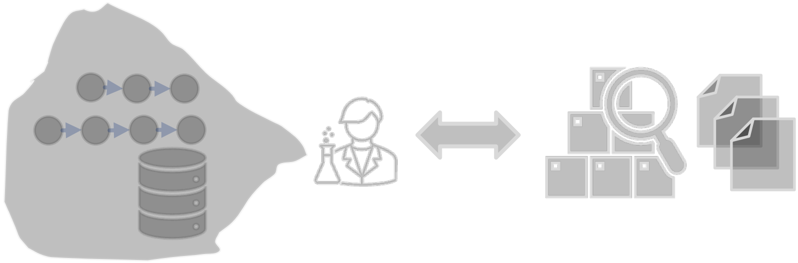
Personalized Search and Recommendations for Patients, Providers, and Researchers

- **Document search:** Retrieving precise, updated, trustworthy information from documents for short queries

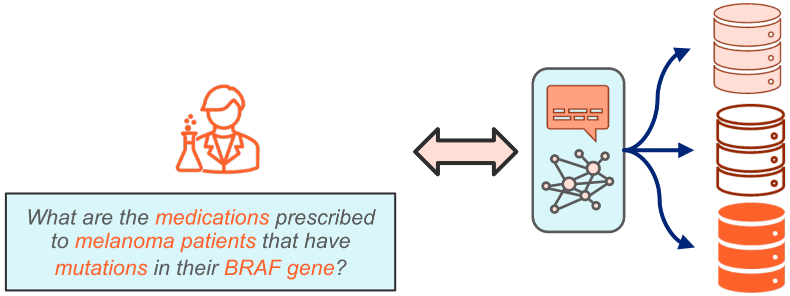


- **Document recommendation:** Browsing for information exploration based on related documents

- **Personalized search and recommendations:** Use contextual information in a privacy-first approach



- **Complex search across multiple databases:** Retrieving precise information with provenance (often datasets with specific entity identifiers and attributes for further analysis)

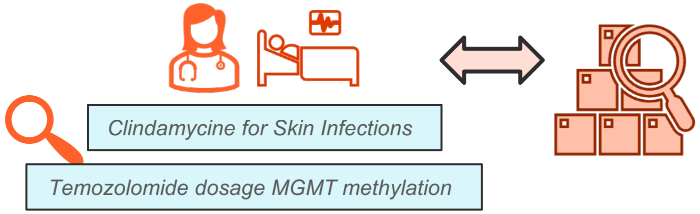


Optum

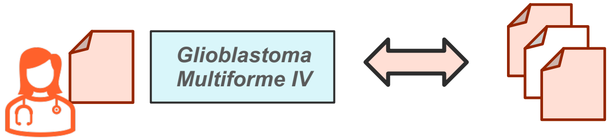
1. **Kamdar, Maulik R.**, et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.**, et al. "A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search." *ISWC Industry Track* (2021).
3. **Kamdar, Maulik R.** "Elsevier's Healthcare Knowledge Graph: An Actionable Medical Knowledge Platform to Power Diverse Applications", *Knowledge Graph Conference* (2021).
4. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

Personalized Search and Recommendations for Patients, Providers, and Researchers

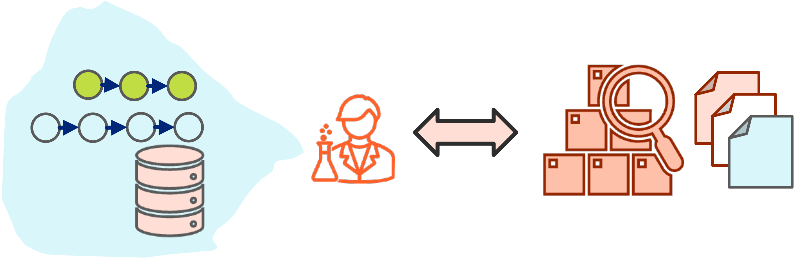
- **Document search:** Retrieving precise, updated, trustworthy information from documents for short queries



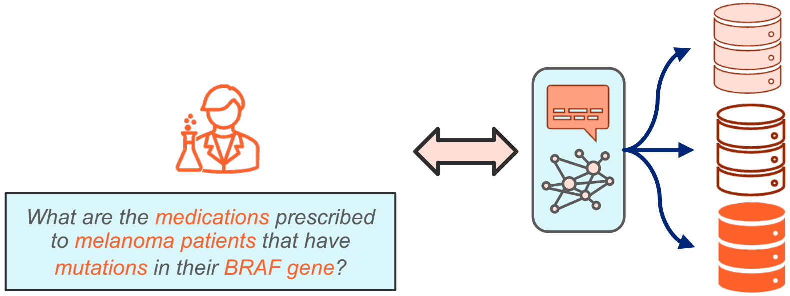
- **Document recommendation:** Browsing for information exploration based on related documents



- **Personalized search and recommendations:** Use contextual information in a privacy-first approach



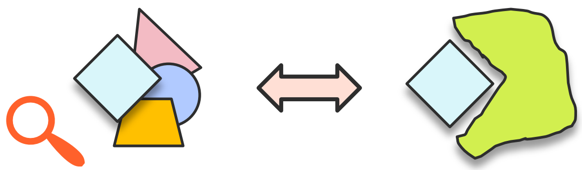
- **Complex search across multiple databases:** Retrieving precise information with provenance (often datasets with specific entity identifiers and attributes for further analysis)



1. **Kamdar, Maulik R.**, et al. "Enabling web-scale data integration in biomedicine through linked open data." *NPJ digital medicine* 2.1 (2019): 1-14.
2. **Kamdar, Maulik R.**, et al. "A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search." *ISWC Industry Track* (2021).
3. **Kamdar, Maulik R.** "Elsevier's Healthcare Knowledge Graph: An Actionable Medical Knowledge Platform to Power Diverse Applications", *Knowledge Graph Conference* (2021).
4. **Kamdar, Maulik R.**, et al. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research." *Journal of biomedical informatics* 47 (2014): 112-130.

Pathways and Predictions

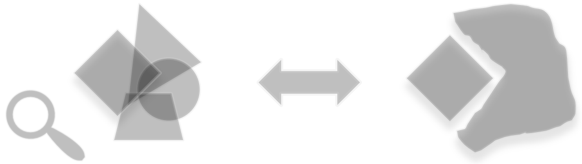
- **Drug discovery and drug repurposing:** Searching the space of potential ligands and drugs for prediction and docking models



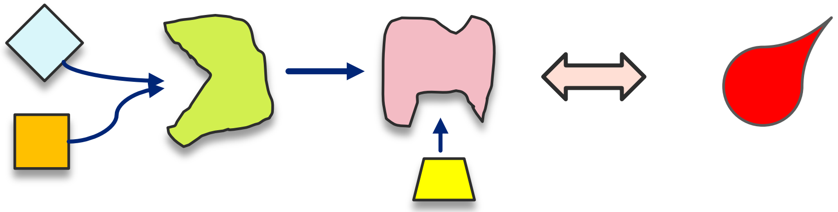
1. Himmelstein, Daniel Scott, et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *Elife* 6 (2017): e26726.
2. **Kamdar, Maulik R.**, et al. "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data." *26th World Wide Web Conference* (2017).
3. Zitnik, Marinka et al. "Modeling polypharmacy side effects with graph convolutional networks." *Bioinformatics* 34.13 (2018): i457-i466.
4. **Kamdar, Maulik R.**, et al. "An Ebola virus-centered knowledge base." *Database* (2015).
5. **Kamdar, Maulik R.**, et al. "A knowledge graph-based approach for exploring the US opioid epidemic." AI4SG Workshop, ICLR Conference (2019).

Pathways and Predictions

- **Drug discovery and drug repurposing:** Searching the space of potential ligands and drugs for prediction and docking models



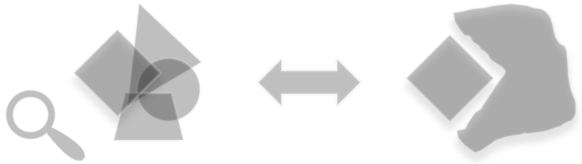
- **Pharmacovigilance and drug safety:** Integrating drug and protein interaction networks to predict adverse reactions



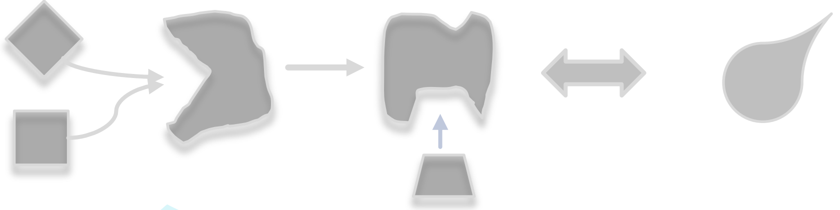
1. Himmelstein, Daniel Scott, et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *Elife* 6 (2017): e26726.
2. **Kamdar, Maulik R.**, et al. "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data." *26th World Wide Web Conference* (2017).
3. Zitnik, Marinka et al. "Modeling polypharmacy side effects with graph convolutional networks." *Bioinformatics* 34.13 (2018): i457-i466.
4. **Kamdar, Maulik R.**, et al. "An Ebola virus-centered knowledge base." *Database* (2015).
5. **Kamdar, Maulik R.**, et al. "A knowledge graph-based approach for exploring the US opioid epidemic." AI4SG Workshop, ICLR Conference (2019).

Pathways and Predictions

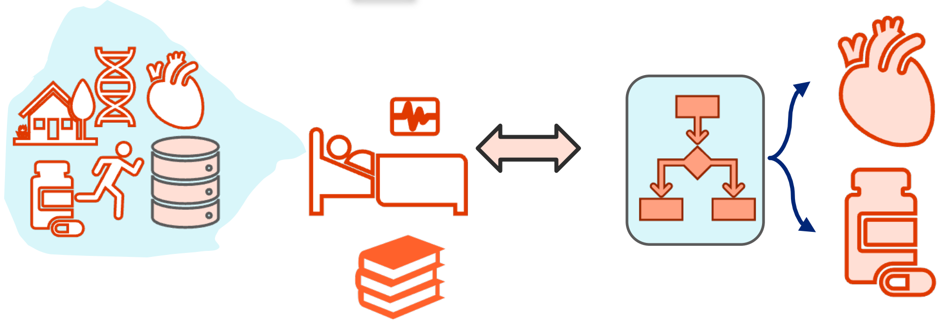
- **Drug discovery and drug repurposing:** Searching the space of potential ligands and drugs for prediction and docking models



- **Pharmacovigilance and drug safety:** Integrating drug and protein interaction networks to predict adverse reactions



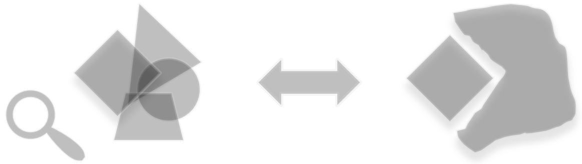
- **Clinical decision support:** Knowledge-driven clinical pathways for next steps (e.g., differential diagnosis, prescriptions)



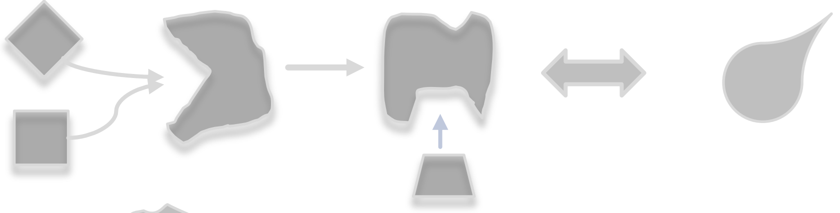
1. Himmelstein, Daniel Scott, et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *Elife* 6 (2017): e26726.
2. **Kamdar, Maulik R.**, et al. "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data." *26th World Wide Web Conference* (2017).
3. Zitnik, Marinka et al. "Modeling polypharmacy side effects with graph convolutional networks." *Bioinformatics* 34.13 (2018): i457-i466.
4. **Kamdar, Maulik R.**, et al. "An Ebola virus-centered knowledge base." *Database* (2015).
5. **Kamdar, Maulik R.**, et al. "A knowledge graph-based approach for exploring the US opioid epidemic." AI4SG Workshop, ICLR Conference (2019).

Pathways and Predictions

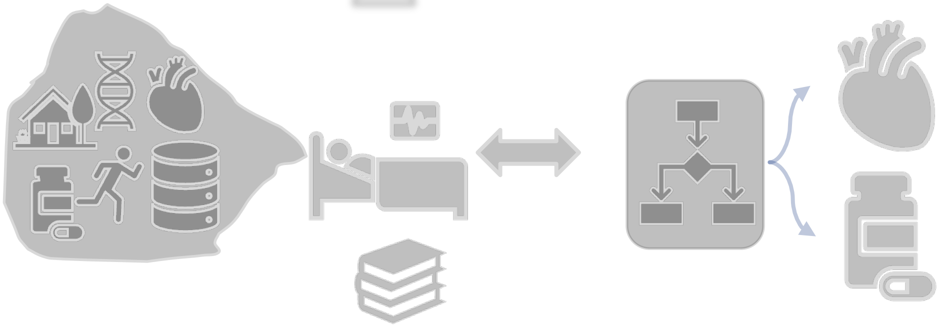
- **Drug discovery and drug repurposing:** Searching the space of potential ligands and drugs for prediction and docking models



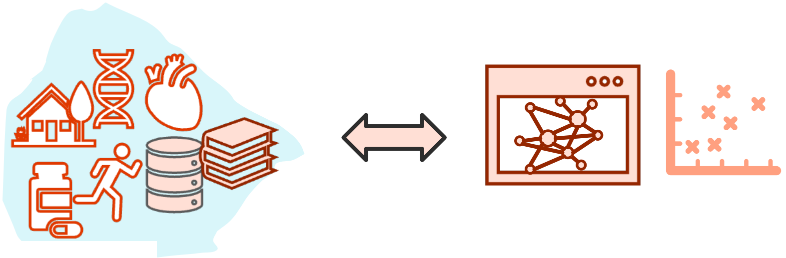
- **Pharmacovigilance and drug safety:** Integrating drug and protein interaction networks to predict adverse reactions



- **Clinical decision support:** Knowledge-driven clinical pathways for next steps (e.g., differential diagnosis, prescriptions)



- **Summarization and other prediction models:** Aggregating information for robust prediction models in epidemiology or outcomes research (e.g., Opioid epidemic, COVID-19, EVD)

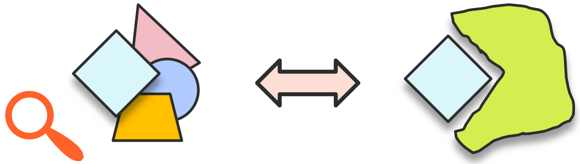


Optum

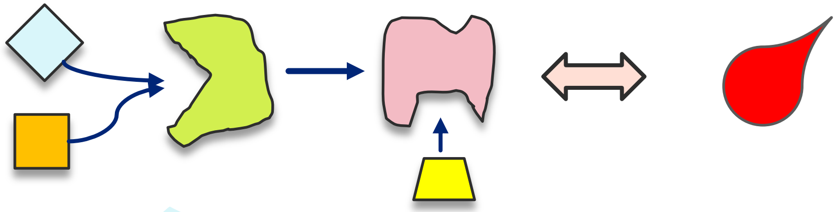
1. Himmelstein, Daniel Scott, et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *Elife* 6 (2017): e26726.
2. **Kamdar, Maulik R.**, et al. "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data." *26th World Wide Web Conference* (2017).
3. Zitnik, Marinka et al. "Modeling polypharmacy side effects with graph convolutional networks." *Bioinformatics* 34.13 (2018): i457-i466.
4. **Kamdar, Maulik R.**, et al. "An Ebola virus-centered knowledge base." *Database* (2015).
5. **Kamdar, Maulik R.**, et al. "A knowledge graph-based approach for exploring the US opioid epidemic." AI4SG Workshop, ICLR Conference (2019).

Pathways and Predictions

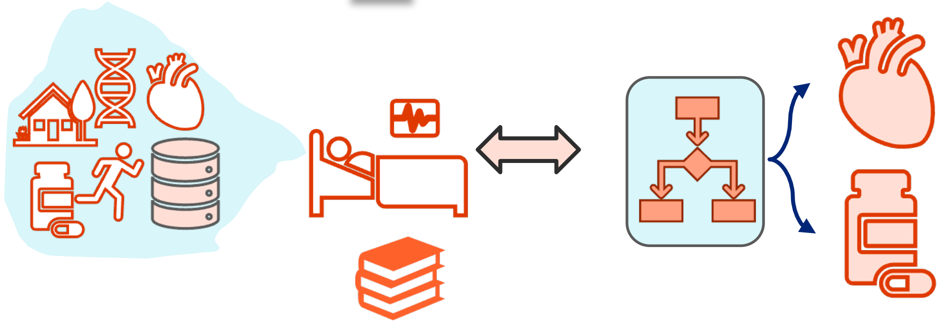
- **Drug discovery and drug repurposing:** Searching the space of potential ligands and drugs for prediction and docking models



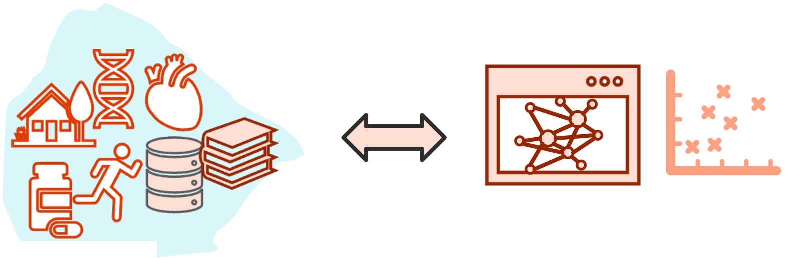
- **Pharmacovigilance and drug safety:** Integrating drug and protein interaction networks to predict adverse reactions



- **Clinical decision support:** Knowledge-driven clinical pathways for next steps (e.g., differential diagnosis, prescriptions)



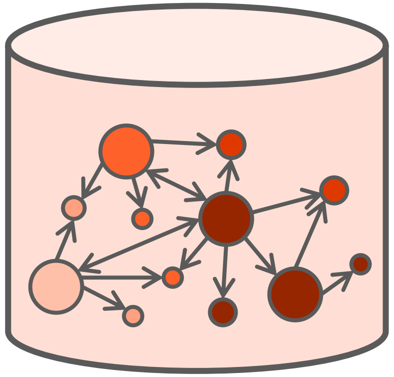
- **Summarization and other prediction models:** Aggregating information for robust prediction models in epidemiology or outcomes research (e.g., Opioid epidemic, COVID-19, EVD)



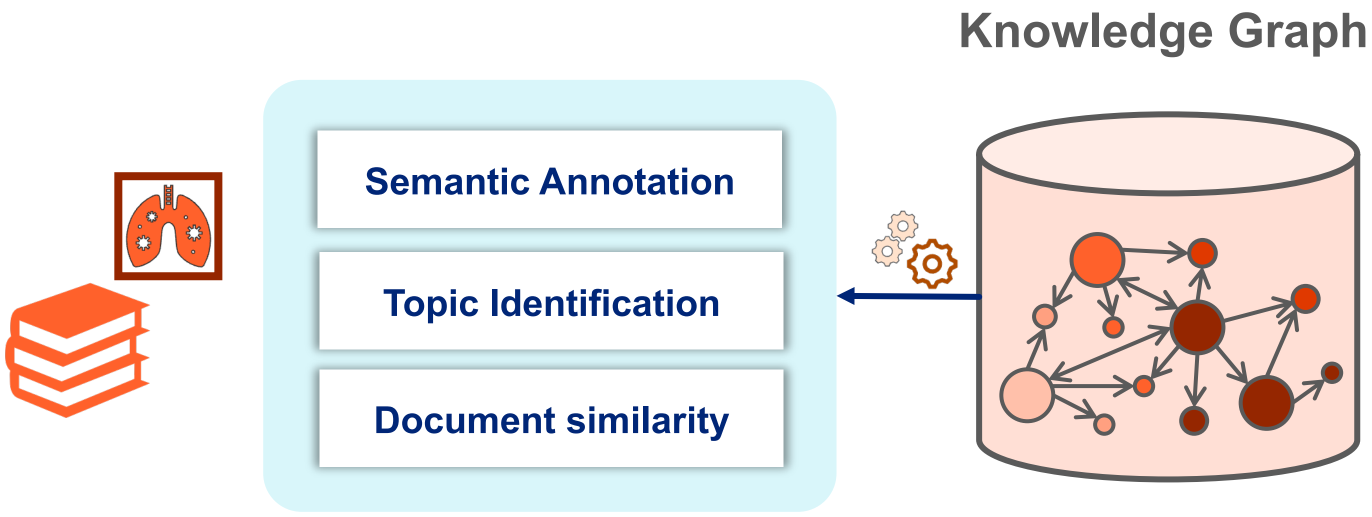
1. Himmelstein, Daniel Scott, et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *Elife* 6 (2017): e26726.
2. **Kamdar, Maulik R.**, et al. "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data." *26th World Wide Web Conference* (2017).
3. Zitnik, Marinka et al. "Modeling polypharmacy side effects with graph convolutional networks." *Bioinformatics* 34.13 (2018): i457-i466.
4. **Kamdar, Maulik R.**, et al. "An Ebola virus-centered knowledge base." *Database* (2015).
5. **Kamdar, Maulik R.**, et al. "A knowledge graph-based approach for exploring the US opioid epidemic." AI4SG Workshop, ICLR Conference (2019).

Graph-powered API services for different applications

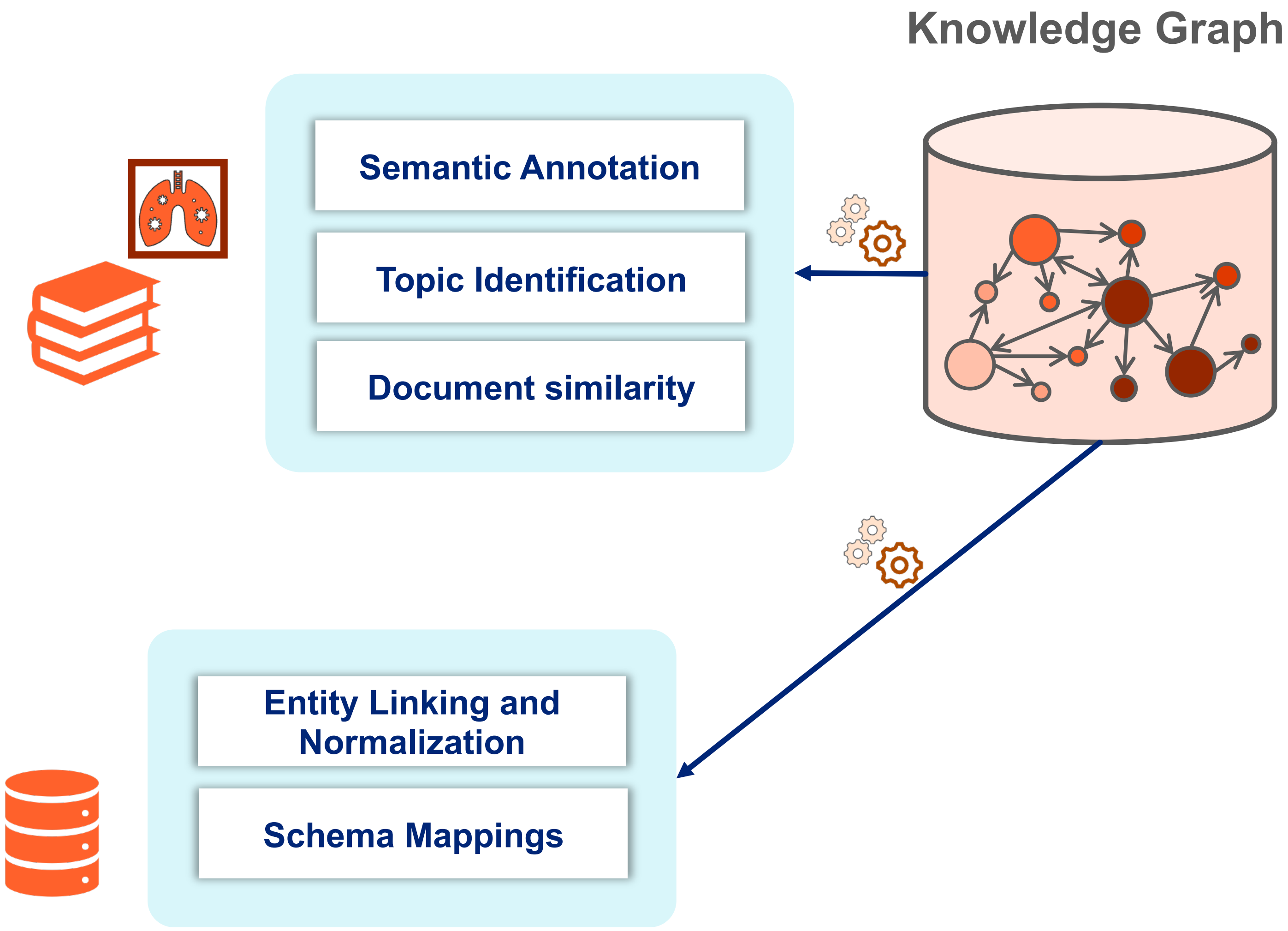
Knowledge Graph



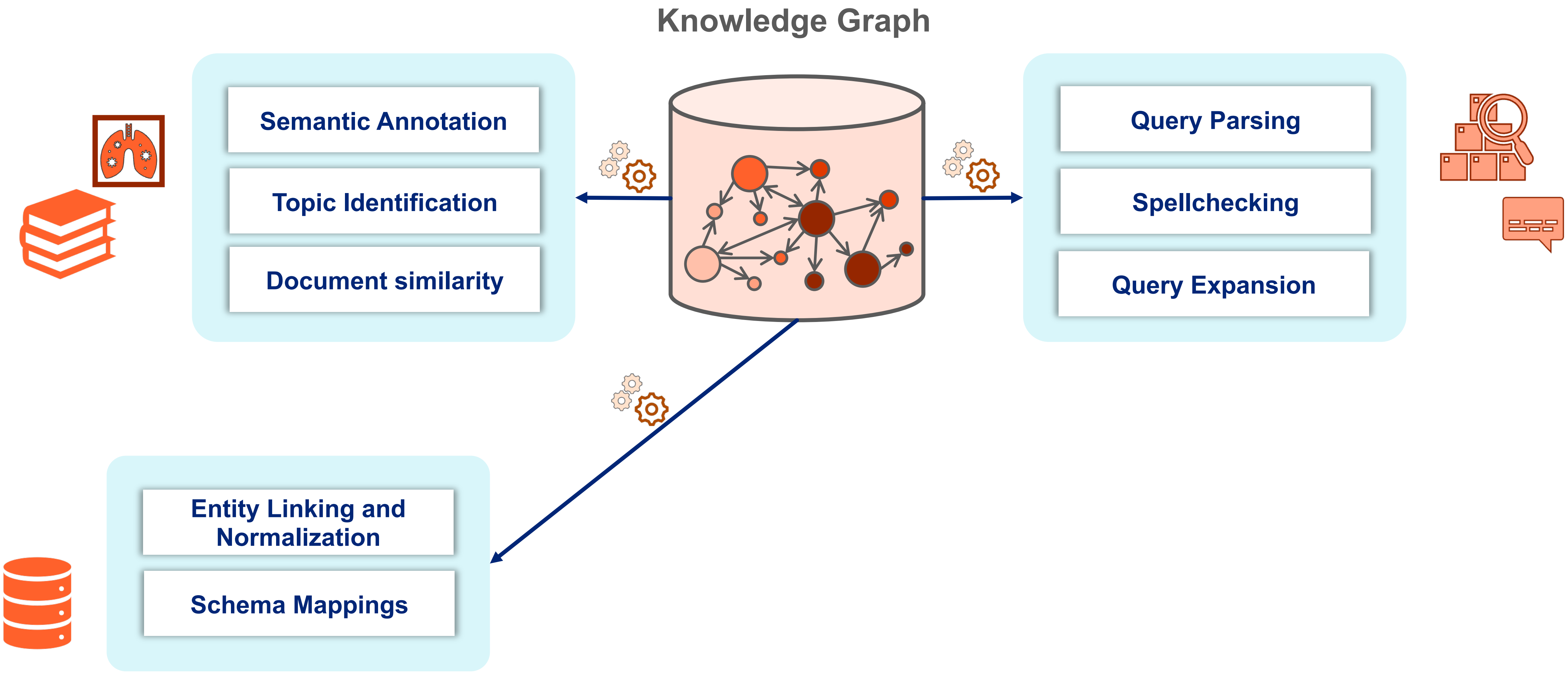
Graph-powered API services for different applications



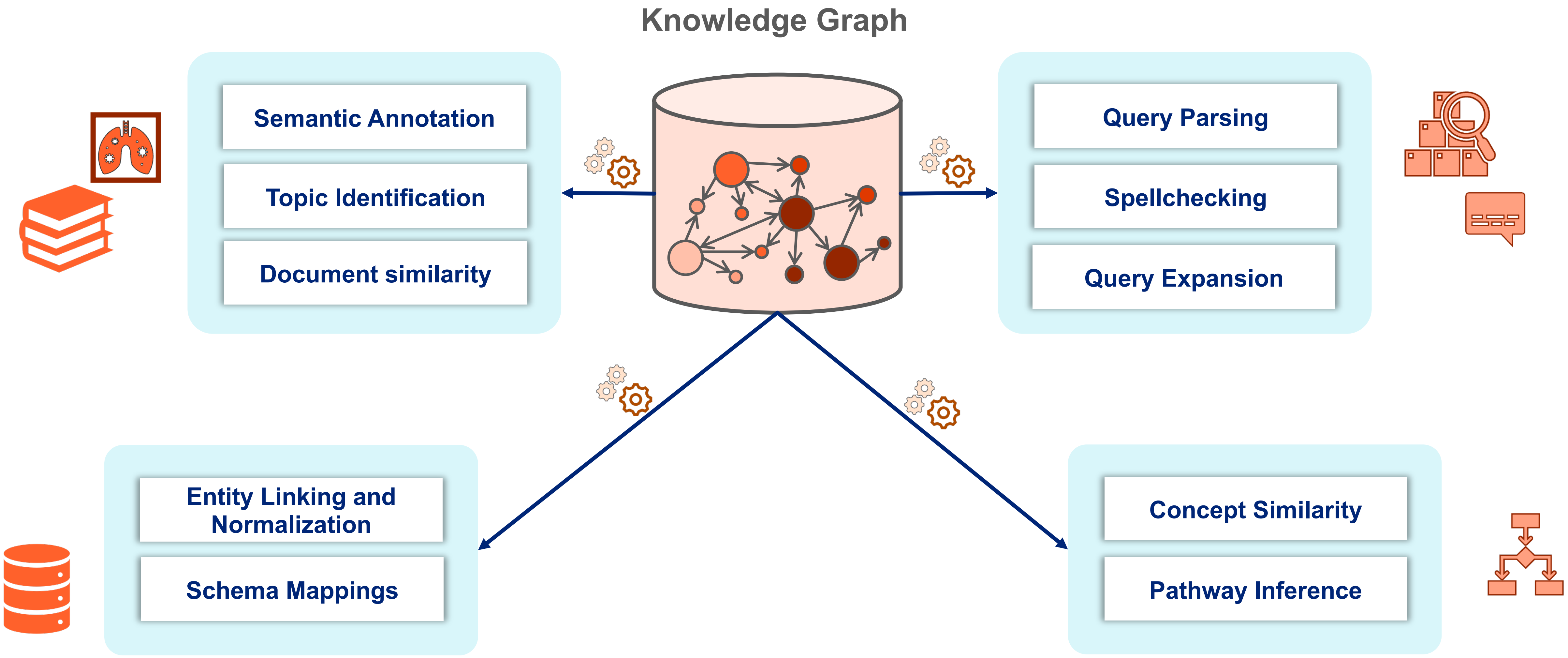
Graph-powered API services for different applications



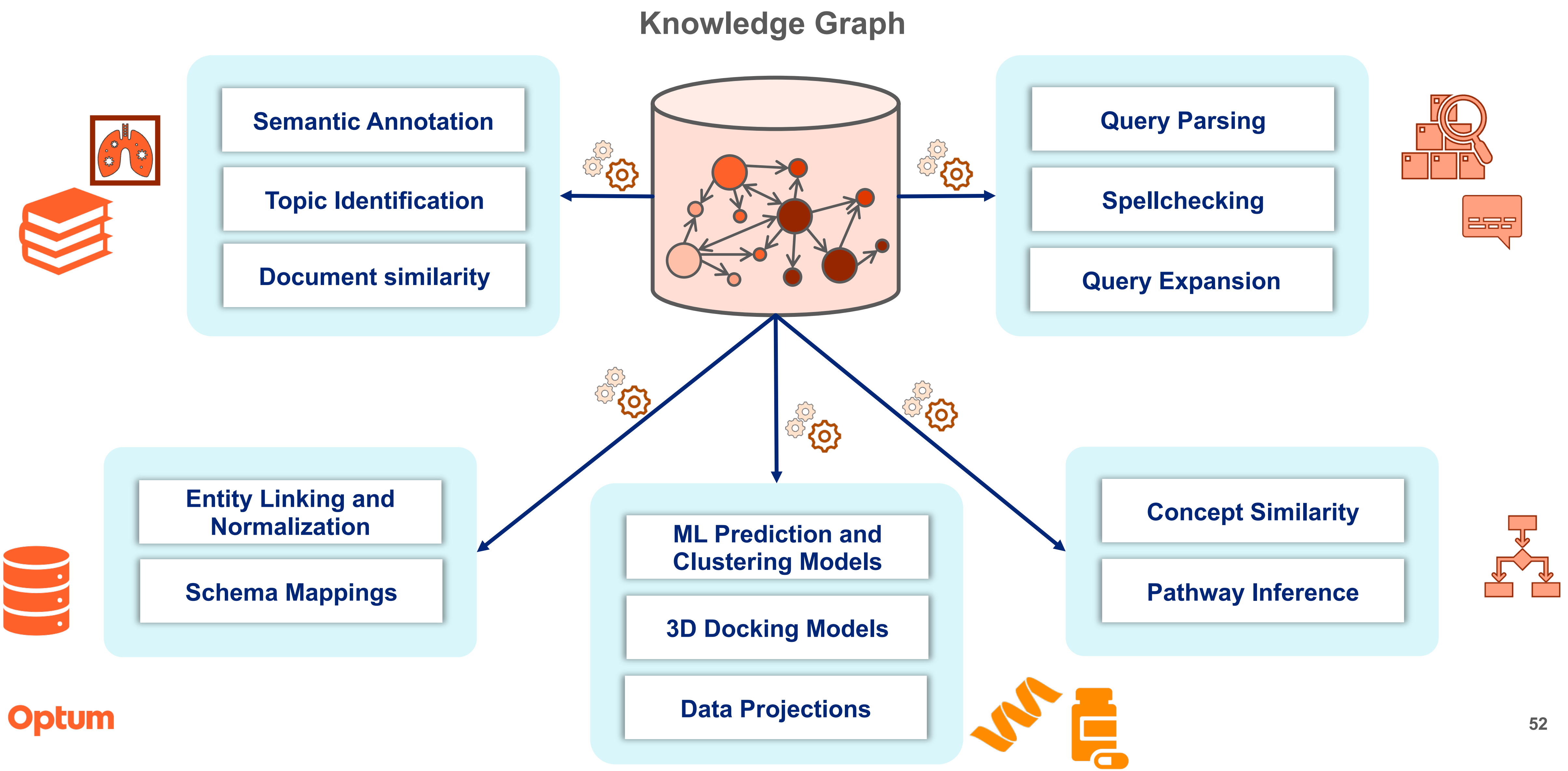
Graph-powered API services for different applications



Graph-powered API services for different applications



Graph-powered API services for different applications

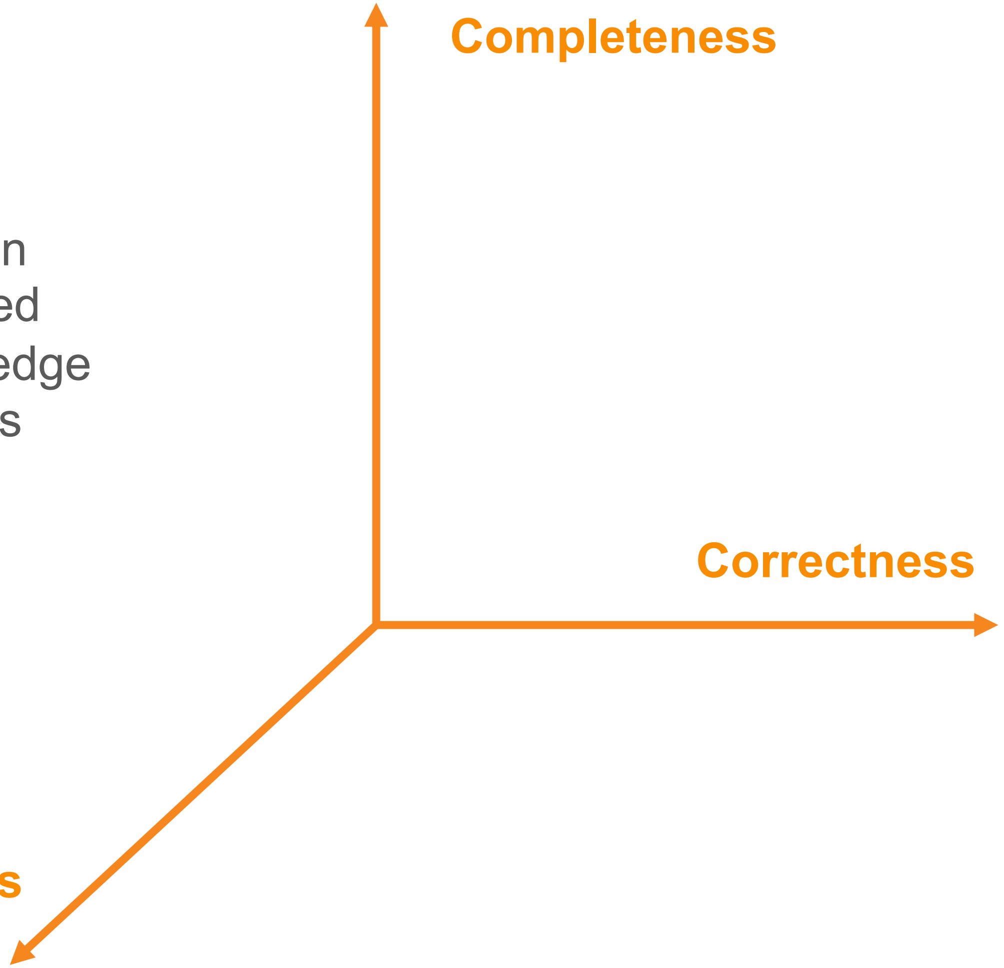


Challenges and Opportunities

“ *There is no such thing as a free lunch.*

Completeness, Correctness, and Freshness of Biomedical Knowledge

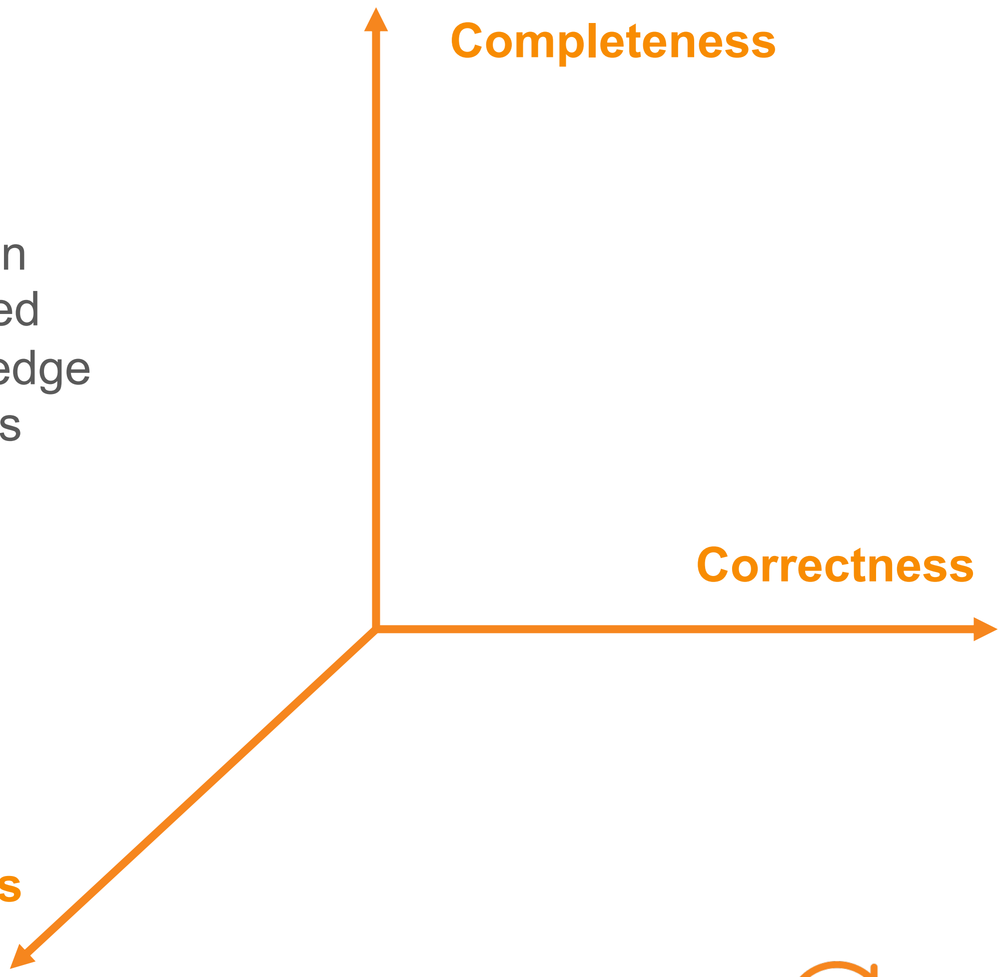
3 metrics are in conflict for scaled biomedical knowledge graph platforms



1. http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/index.html
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.

Completeness, Correctness, and Freshness of Biomedical Knowledge

3 metrics are in conflict for scaled biomedical knowledge graph platforms

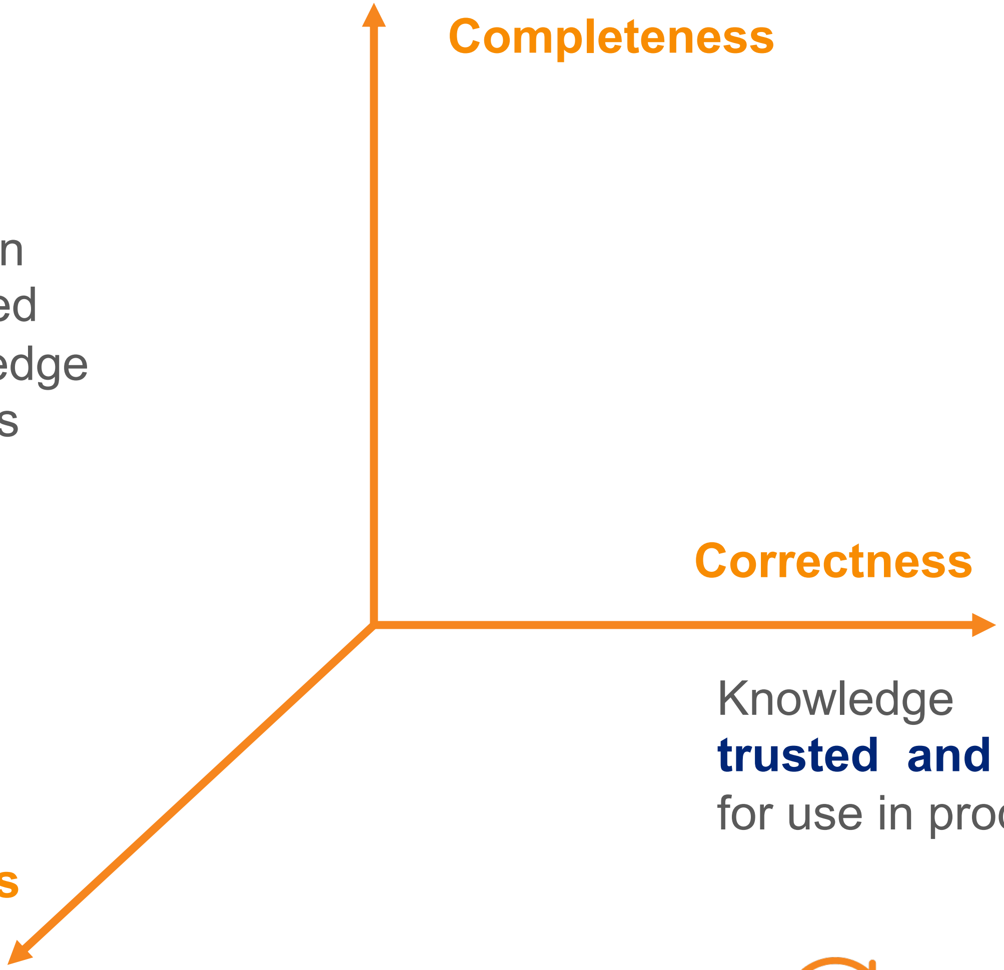


1. http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/index.html
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.

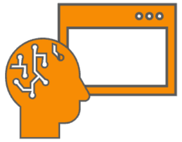
Knowledge is **continuously evolving and expanding** with new research and discoveries

Completeness, Correctness, and Freshness of Biomedical Knowledge

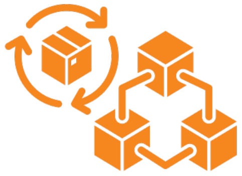
3 metrics are in conflict for scaled biomedical knowledge graph platforms



Knowledge must be **trusted and accurate** for use in products



Knowledge is **continuously evolving and expanding** with new research and discoveries



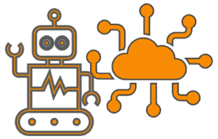
1. http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/index.html
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.

Completeness, Correctness, and Freshness of Biomedical Knowledge

3 metrics are in conflict for scaled biomedical knowledge graph platforms

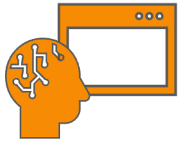
Completeness

Knowledge can arise from several different **heterogeneous** sources



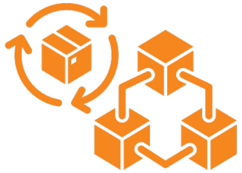
Correctness

Knowledge must be **trusted and accurate** for use in products



Freshness

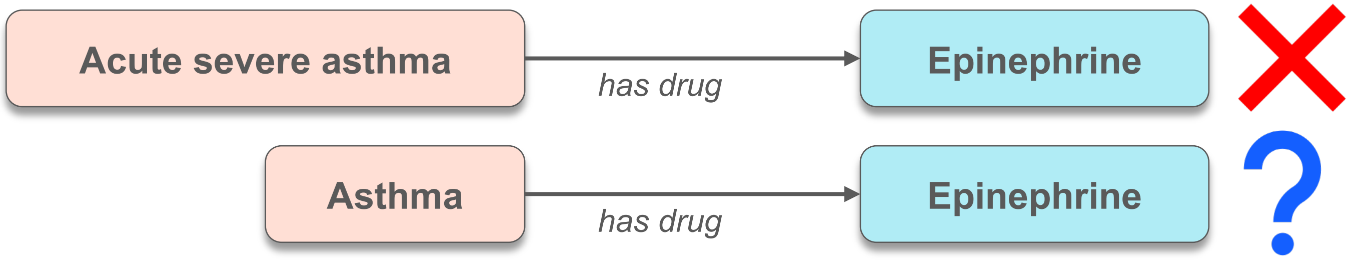
Knowledge is **continuously evolving and expanding** with new research and discoveries



1. http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/index.html
2. Noy, N., et al. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue (2019)*, 48-75.

Knowledge Extraction Challenges

Status asthmaticus, or acute severe asthma, is a refractory state that does not respond to standard therapy such as inhaled beta-agonists or subcutaneous epinephrine



- **Medical language is complex**: Speculation, negation, context (e.g., age, ethnicity), multiple entities etc.

Optum

Kamdar, Maulik R., et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* 2020 (2020): 288.

Knowledge Extraction Challenges

Status asthmaticus, or acute severe asthma, is a refractory state that does not respond to standard therapy such as inhaled beta-agonists or subcutaneous epinephrine

Acute severe asthma $\xrightarrow{\text{has drug}}$ Epinephrine ✗

Asthma $\xrightarrow{\text{has drug}}$ Epinephrine ?

Early retrospective studies suggested that reserpine was associated with breast cancer, but prospective studies and meta-analyses of case, control studies have shown only a weak association.

Breast Cancer $\xrightarrow{\text{has drug}}$ Reserpine ✗

- **Medical language is complex**: Speculation, negation, context (e.g., age, ethnicity), multiple entities etc.

Optum

Kamdar, Maulik R., et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* 2020 (2020): 288.

Knowledge Extraction Challenges

Status asthmaticus, or acute severe asthma, is a refractory state that does not respond to standard therapy such as inhaled beta-agonists or subcutaneous epinephrine

Acute severe asthma —*has drug*→ **Epinephrine**

Asthma —*has drug*→ **Epinephrine** ?

Early retrospective studies suggested that reserpine was associated with breast cancer, but prospective studies and meta-analyses of case, control studies have shown only a weak association.

Breast Cancer —*has drug*→ **Reserpine**

- **Medical language is complex**: Speculation, negation, context (e.g., age, ethnicity), multiple entities etc.
  - NLP is getting more advanced with the advent of **GPT-based models**

Kamdar, Maulik R., et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* 2020 (2020): 288.

Knowledge Extraction Challenges

Status asthmaticus, or acute severe asthma, is a refractory state that does not respond to standard therapy such as inhaled beta-agonists or subcutaneous epinephrine

Acute severe asthma → *has drug* → **Epinephrine**

Asthma → *has drug* → **Epinephrine**

?

Early retrospective studies suggested that reserpine was associated with breast cancer, but prospective studies and meta-analyses of case, control studies have shown only a weak association.

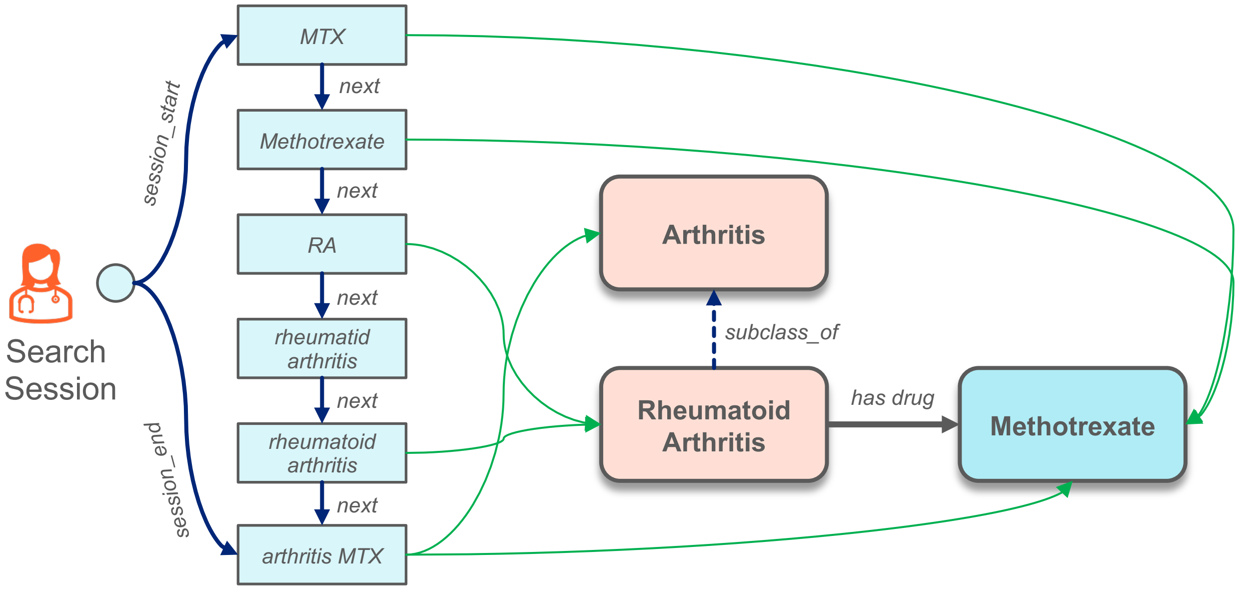
Breast Cancer → *has drug* → **Reserpine**

- **Medical language is complex**: Speculation, negation, context (e.g., age, ethnicity), multiple entities etc.
  - NLP is getting more advanced with the advent of **GPT-based models**
- **Provenance and confidence metrics** can be stored within the knowledge graph and used for querying (e.g., thresholding or combining on metrics)
  - **Human-in-the-loop** can decide depending on the use case(s) and the risk(s).

Optum

Kamdar, Maulik R., et al. "Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings." *AMIA Summits on Translational Science Proceedings* 2020 (2020): 288.

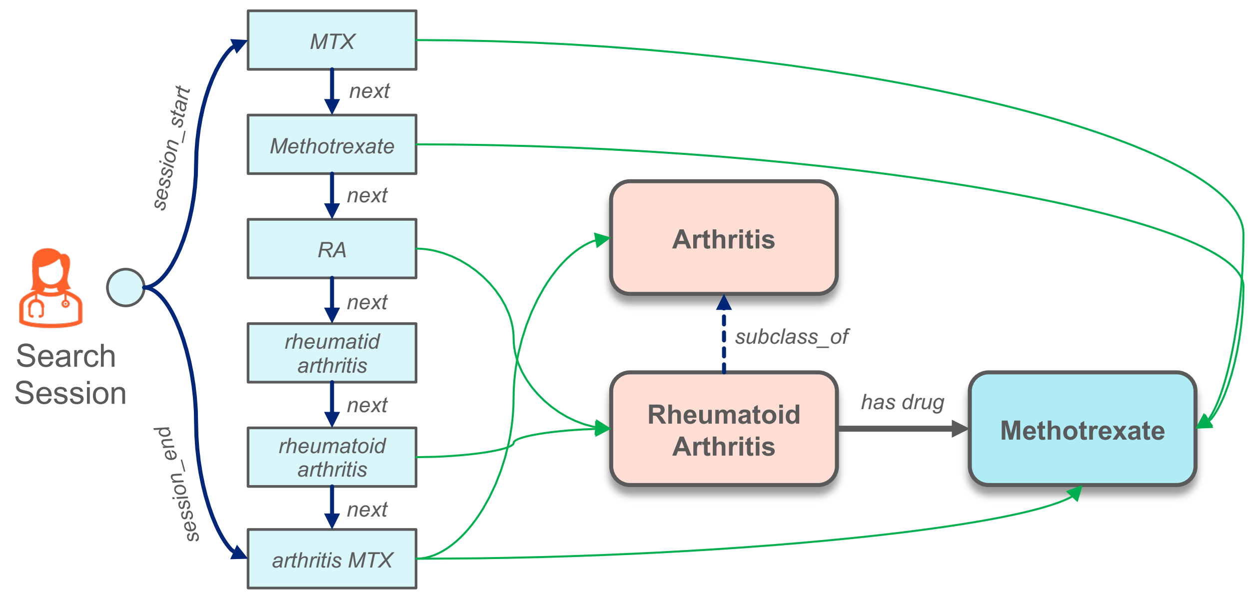
Representation, Storage, and Querying Challenges

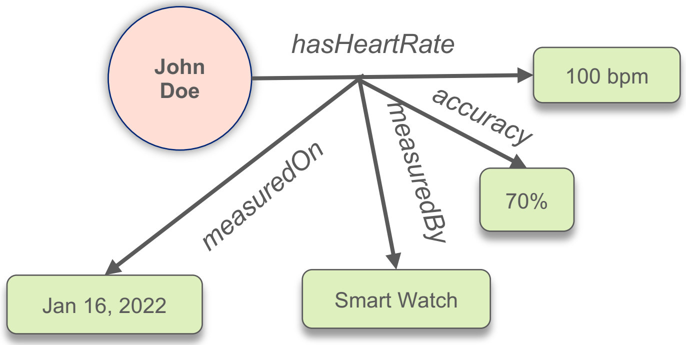


- Modeling, representation, and querying of **temporal sequences, probabilistic data, grouped set of concepts**

1. https://w3c.github.io/rdf-star/cg-spec/editors_draft.html
2. Whang, J. J., et al (2020). MEGA: Multi-view semi-supervised clustering of hypergraphs. *Proceedings of the VLDB Endowment*, *13*(5), 698-711.
3. Fatemi, B., et al. Knowledge hypergraphs: Prediction beyond binary relations. *arXiv preprint (2019).*

Representation, Storage, and Querying Challenges

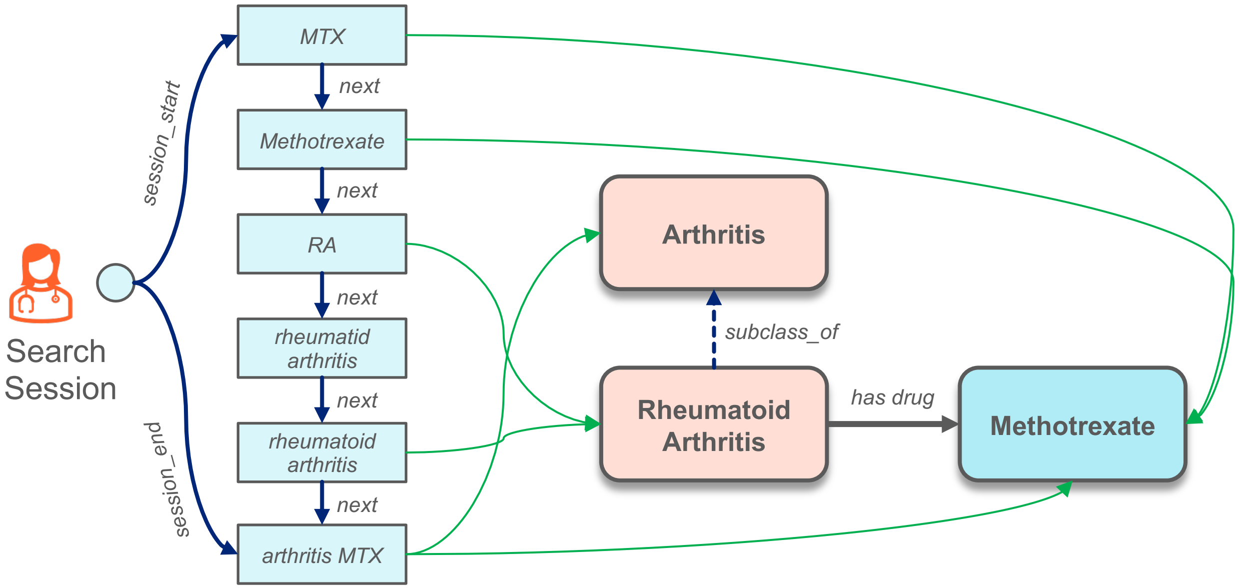


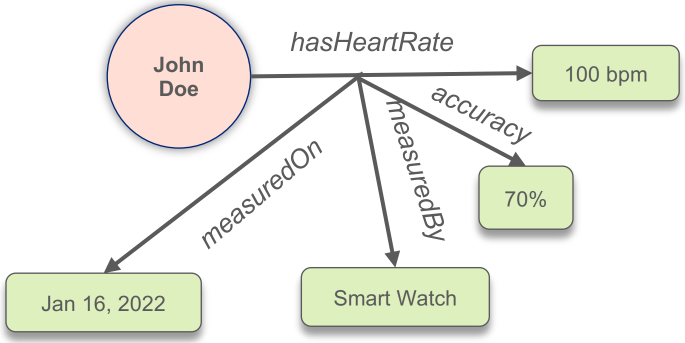


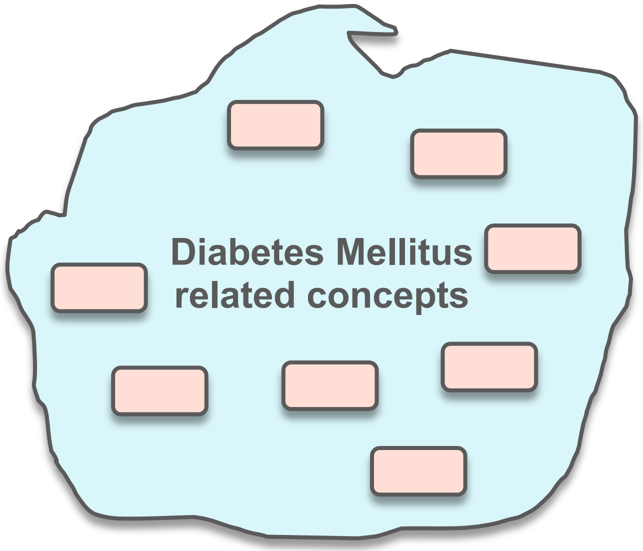
- Modeling, representation, and querying of **temporal sequences, probabilistic data, grouped set of concepts**

1. https://w3c.github.io/rdf-star/cg-spec/editors_draft.html
2. Whang, J. J., et al (2020). MEGA: Multi-view semi-supervised clustering of hypergraphs. *Proceedings of the VLDB Endowment*, *13*(5), 698-711.
3. Fatemi, B., et al. Knowledge hypergraphs: Prediction beyond binary relations. *arXiv preprint (2019).*

Representation, Storage, and Querying Challenges



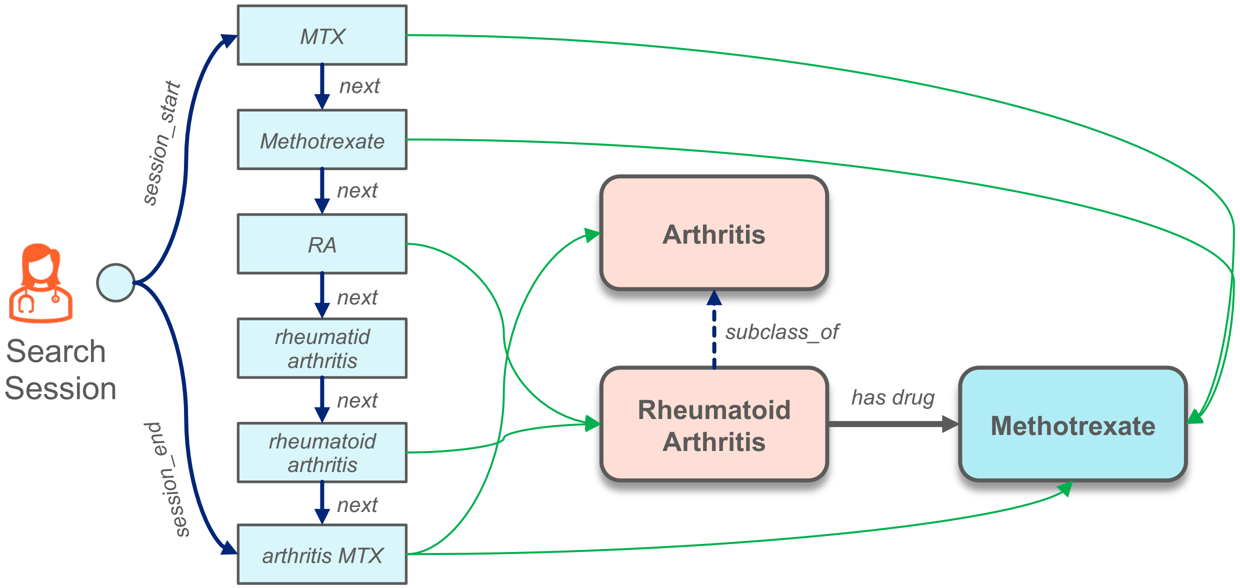


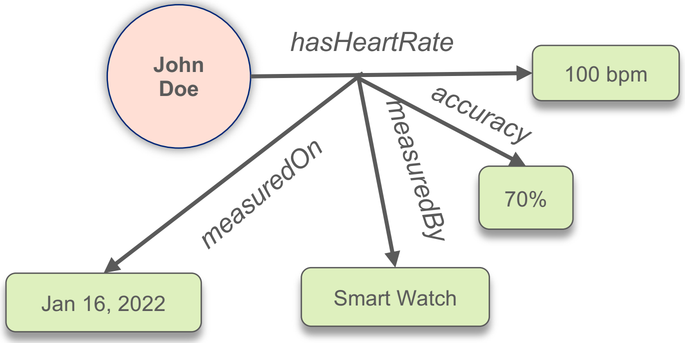


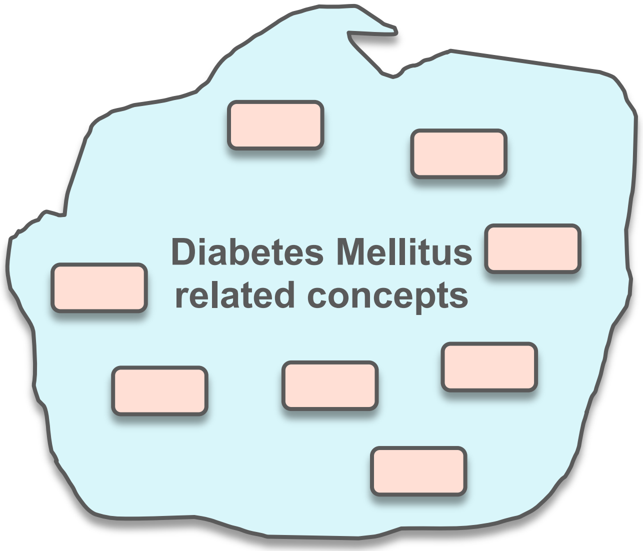
- Modeling, representation, and querying of **temporal sequences, probabilistic data, grouped set of concepts**

1. https://w3c.github.io/rdf-star/cg-spec/editors_draft.html
2. Whang, J. J., et al (2020). MEGA: Multi-view semi-supervised clustering of hypergraphs. *Proceedings of the VLDB Endowment*, *13*(5), 698-711.
3. Fatemi, B., et al. Knowledge hypergraphs: Prediction beyond binary relations. *arXiv preprint (2019).*

Representation, Storage, and Querying Challenges





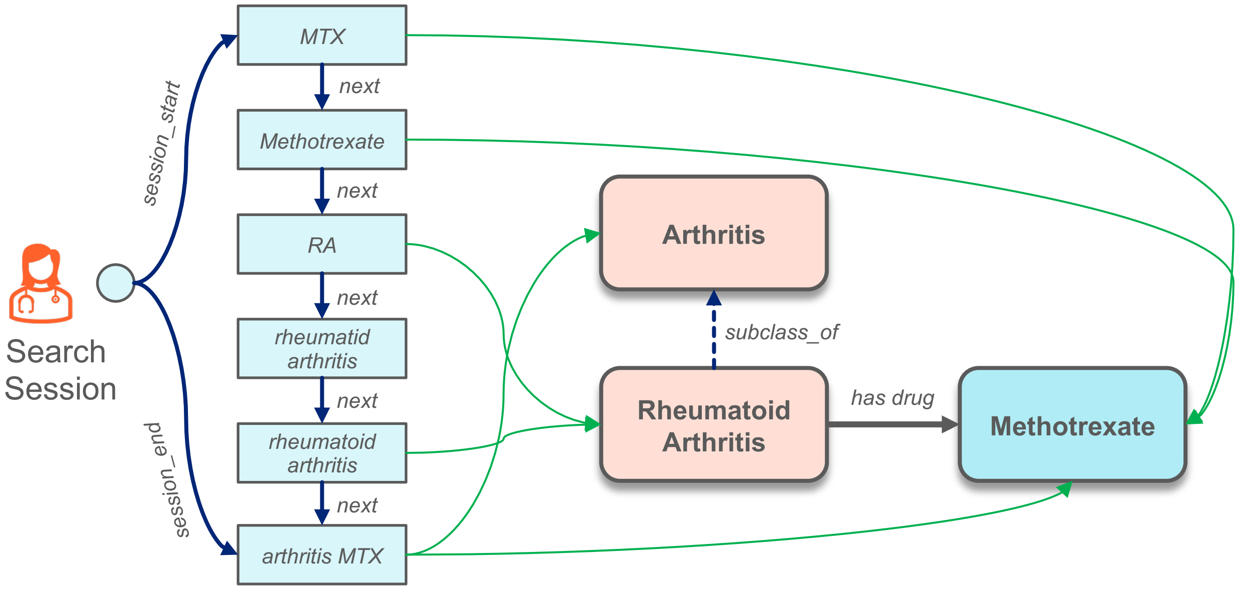


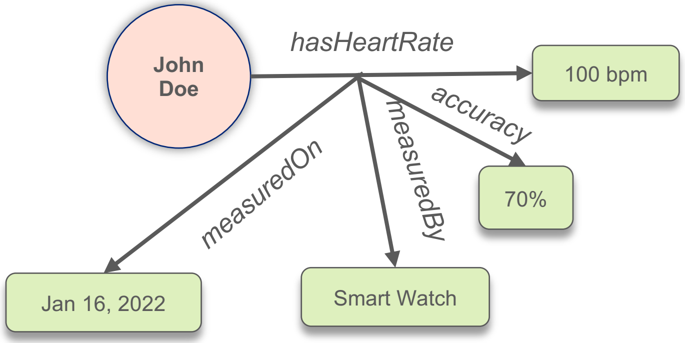
- Modeling, representation, and querying of **temporal sequences, probabilistic data, grouped set of concepts**
  - Reification, RDF*/SPARQL*, labelled property graphs, hypergraph formalisms

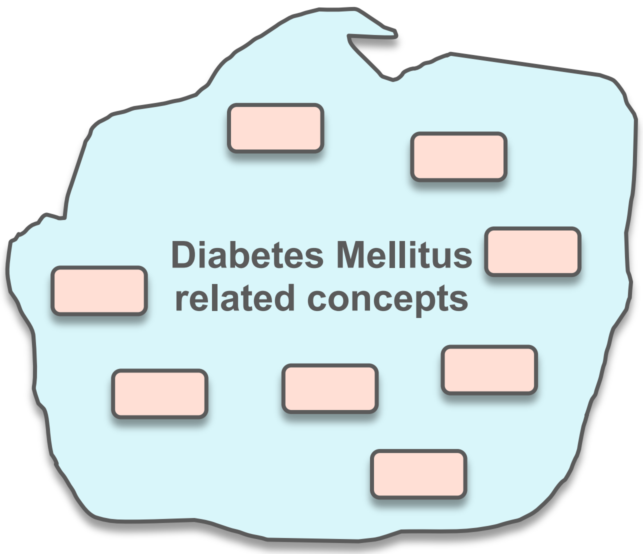
1. https://w3c.github.io/rdf-star/cg-spec/editors_draft.html
2. Whang, J. J., et al (2020). MEGA: Multi-view semi-supervised clustering of hypergraphs. *Proceedings of the VLDB Endowment*, *13*(5), 698-711.
3. Fatemi, B., et al. Knowledge hypergraphs: Prediction beyond binary relations. *arXiv preprint (2019).*

Optum

Representation, Storage, and Querying Challenges







- Modeling, representation, and querying of **temporal sequences, probabilistic data, grouped set of concepts**
  - Reification, RDF*/SPARQL*, labelled property graphs, hypergraph formalisms
- **Data copies** and **redundant duplication** across enterprise with knowledge graphs
  - Privacy and security with patient and user context data (e.g., search histories)

1. https://w3c.github.io/rdf-star/cg-spec/editors_draft.html
2. Whang, J. J., et al (2020). MEGA: Multi-view semi-supervised clustering of hypergraphs. *Proceedings of the VLDB Endowment*, *13*(5), 698-711.
3. Fatemi, B., et al. Knowledge hypergraphs: Prediction beyond binary relations. *arXiv preprint (2019).*

Optum

The Social Factor in Enterprise Knowledge Graphs (KGs)

Stakeholder buy-in needed for development, quality assurance, maintenance, and adoption of the KGs in business products

KGs can be made more **tangible** – user-friendly visualization interfaces, developer-friendly API services, business use case-driven KG development, etc.

Thank you!

Email: maulik_kamdar@optum.com

Twitter: @maulikkamdar