Welcome to INFO216:
Knowledge Graphs
Spring 2022

Andreas L Opdahl
<Andreas.Opdahl@uib.no>

About me

- Background:
  - siv ing (1988), dr ing (1992) from NTH/NTNU
  - Univ of Bergen (early 1990-ies)
  - part-time programmer / consulting for industry
  - several Forskningsråd and EU projects and networks
- Central research interest:
  - modelling of information systems and enterprises
  - semantic modelling and modelling languages
  - semantic technologies
  - knowledge graphs in the media sector

Recent project: BDEM

- Leveraging *Big Data for Emergency Management*
  - how can semantic technologies play a part?
  - developed a new Master course: INFO319

Recent project: UBIMOB

Ubiquitous Data-Driven Urban Mobility

WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

NTNU
Norwegian University of Science and Technology

UNIVERSITAS BERGENSIS

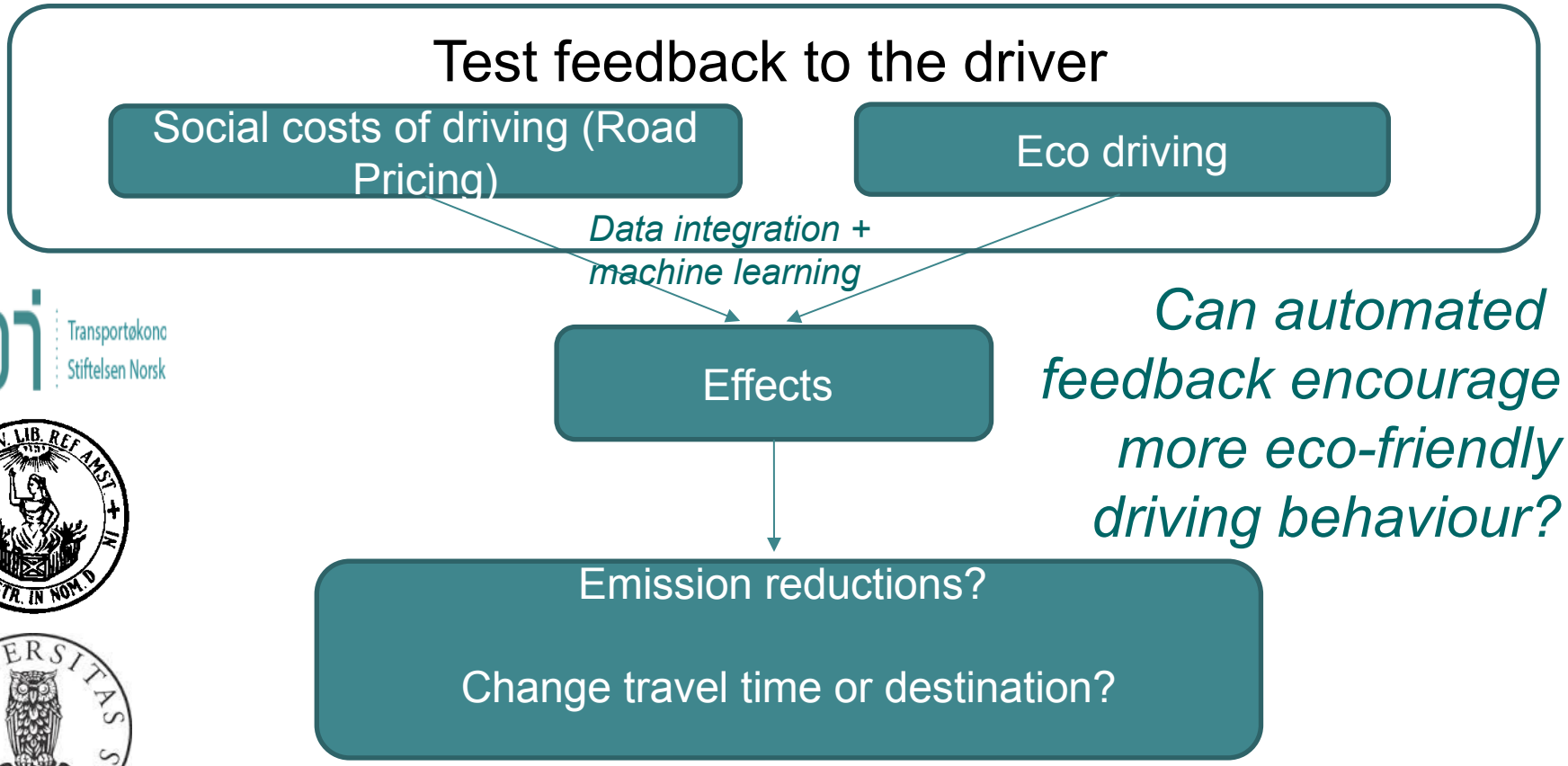
toi
Transportøkonomisk institutt
Stiftelsen Norsk senter for samferdselsforskning

Forschungszentrum · Research Center
L3S

The University Of Sheffield.

telenor

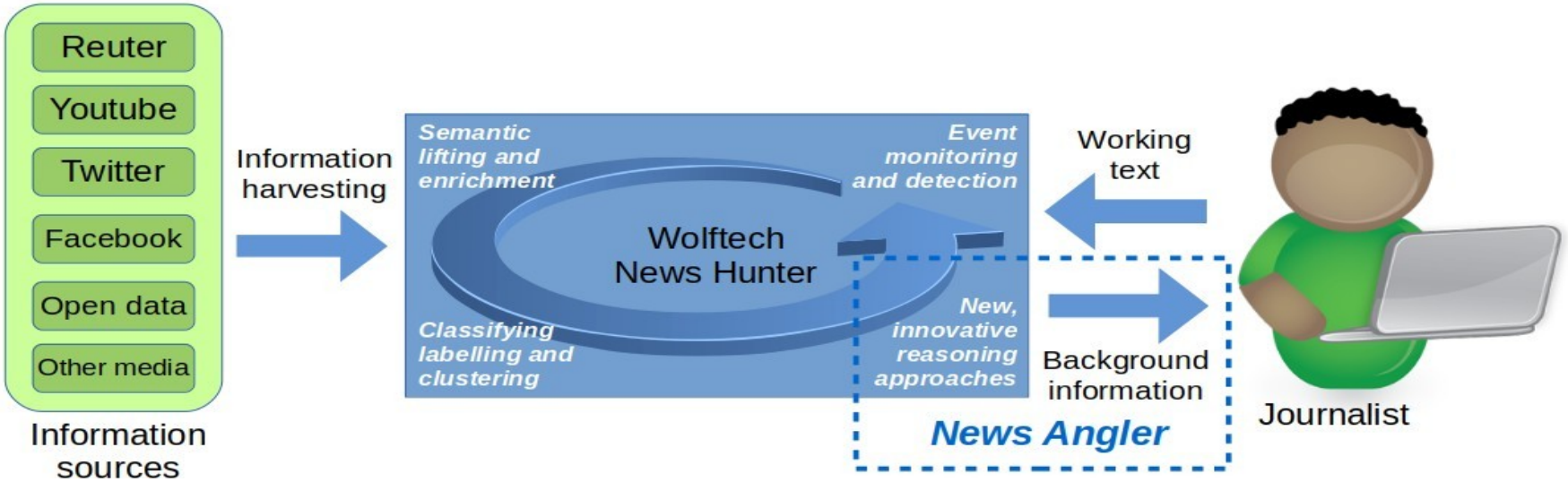
Recent project: Transfeed



Can automated feedback encourage more eco-friendly driving behaviour?

http://newsangler.uib.no

Active project: News Angler

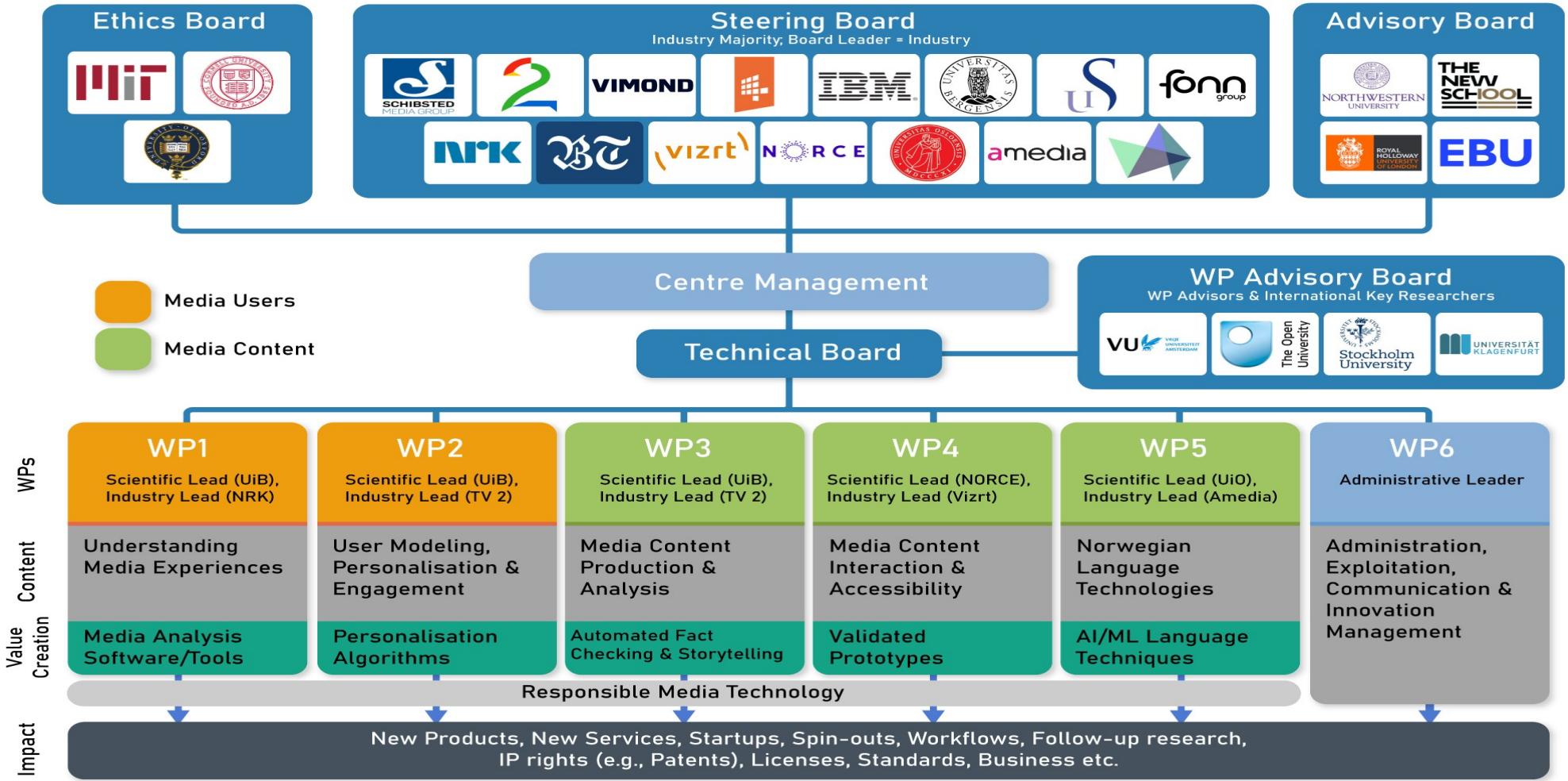


"Wolftech News supports and improves the workflows in a newsroom through mobile solutions for field work that are integrated with central systems for news monitoring, resource management, news editing, and multi-platform publishing"

1) Harvesting and analysing messages
2) Growing a semantic news graph
   - concepts, named entities, context…
3) Analysing working texts (stories)
4) Identifying background information
5) Prioritising and preparing
6) Journalistic and editorial preferences

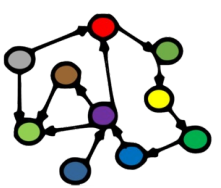
Research: graph, searches, preparation, preferences, language, scaling

Active centre: Media Futures



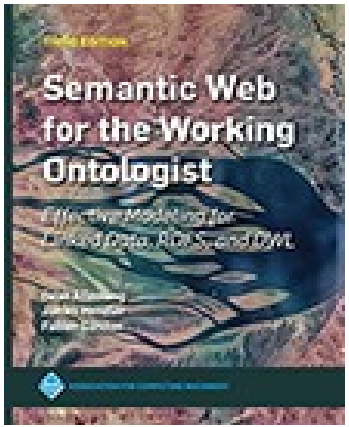
Session 1: Introduction to Knowledge Graphs

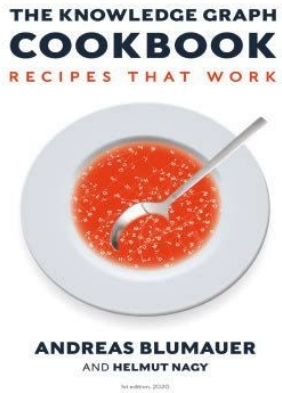
- Themes:
  - *what are knowledge graphs (KGs)?*
    - and *who uses them?*
    - examples of important *open KGs*
  - *background*
    - what are the *semantic web*, *semantic technologies*, and *linked data?*
  - *about INFO216*
    - organisation of the course
    - practical information



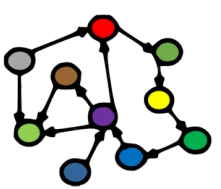
Readings

- Sources:
  - Allemang, Hendler & Gandon (2020): Semantic Web for the Working Ontologist, 3rd edition (chapters 1-2)
  - Blumauer & Nagy (2020): Knowledge Graph Cookbook – Recipes that Work (pages 27-55, 105-122, *supplementary*)
- Material at http://wiki.uib.no/info216:
  - Tim Berners-Lee talks about the semantic web
  - links to important open KGs



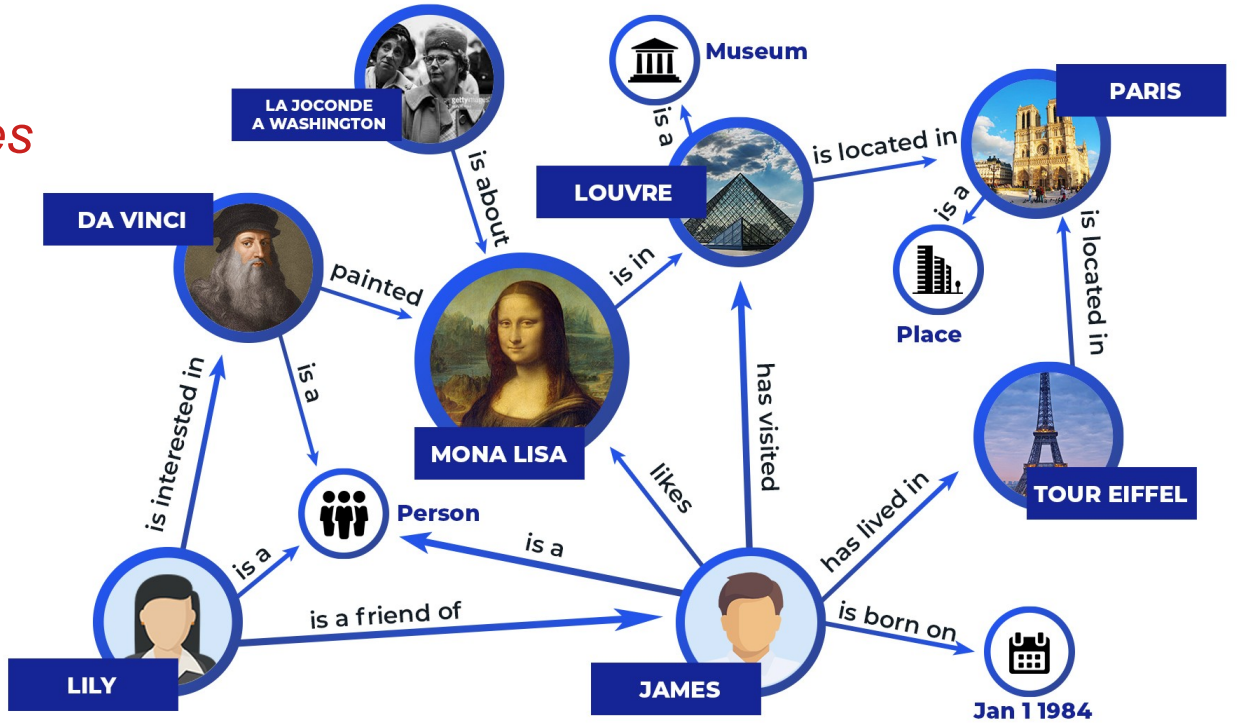


What are knowledge graphs?



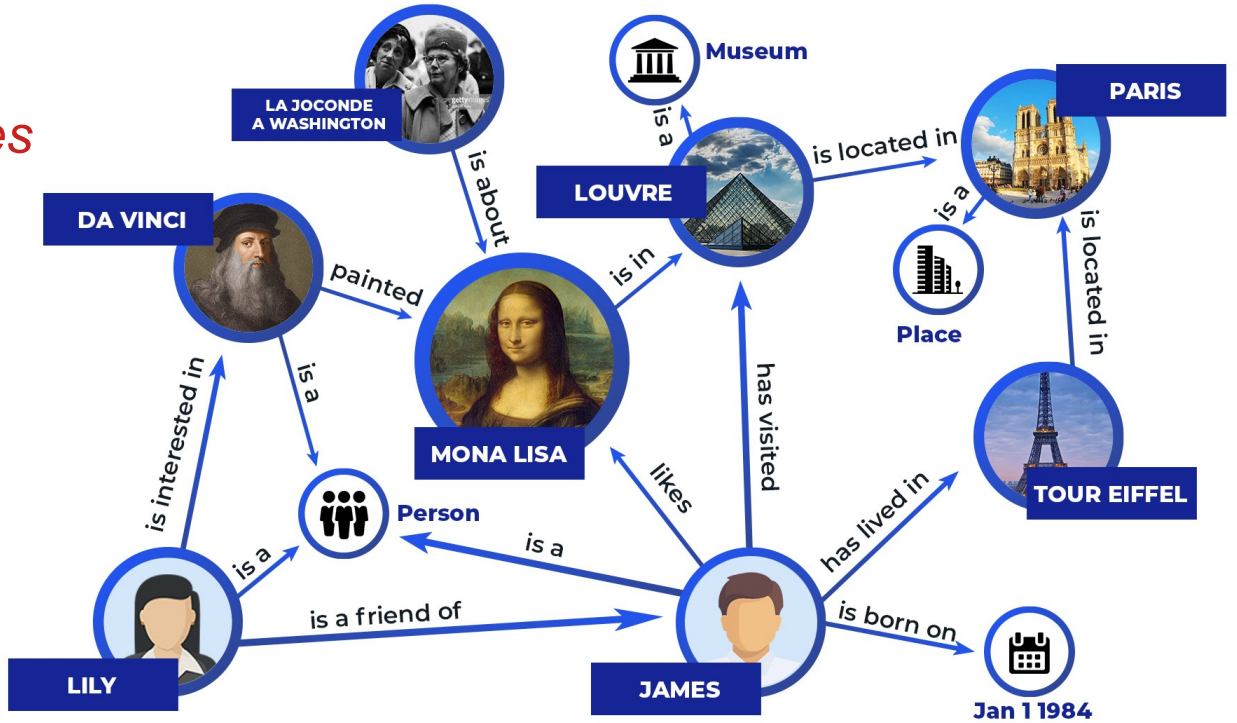
Knowledge graph

- A *graph* of *nodes*
  connected by directed *edges*



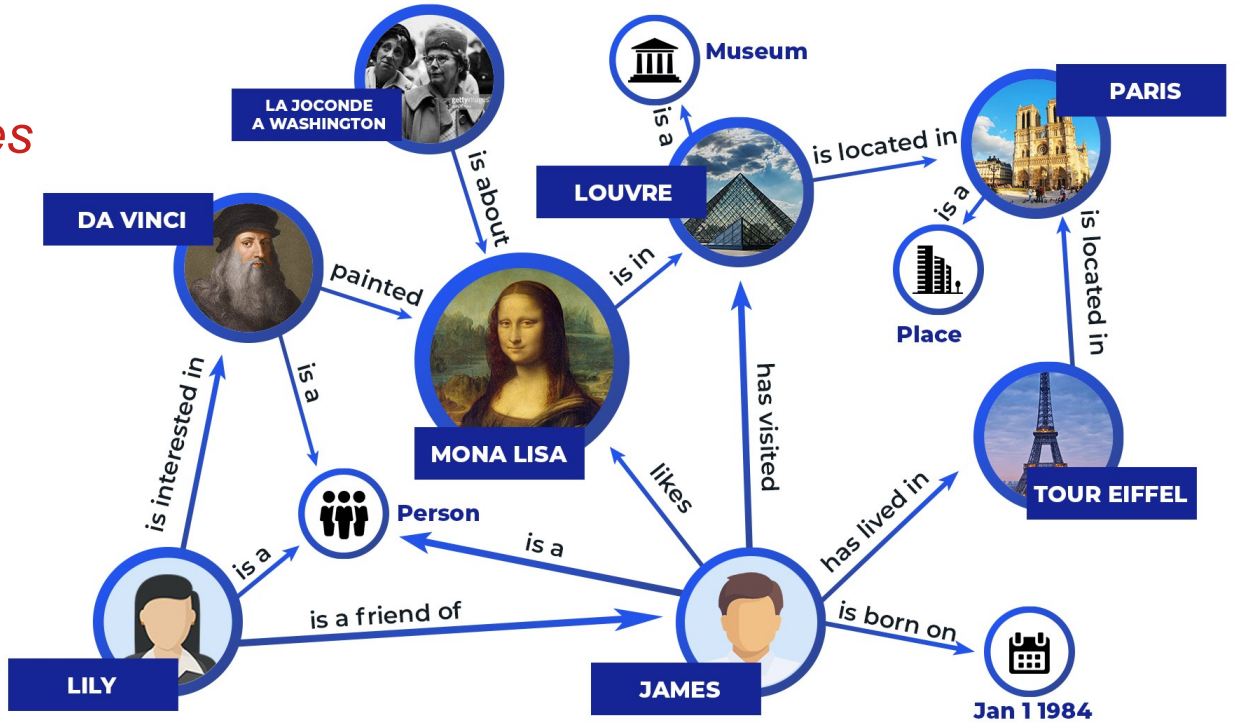
Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*



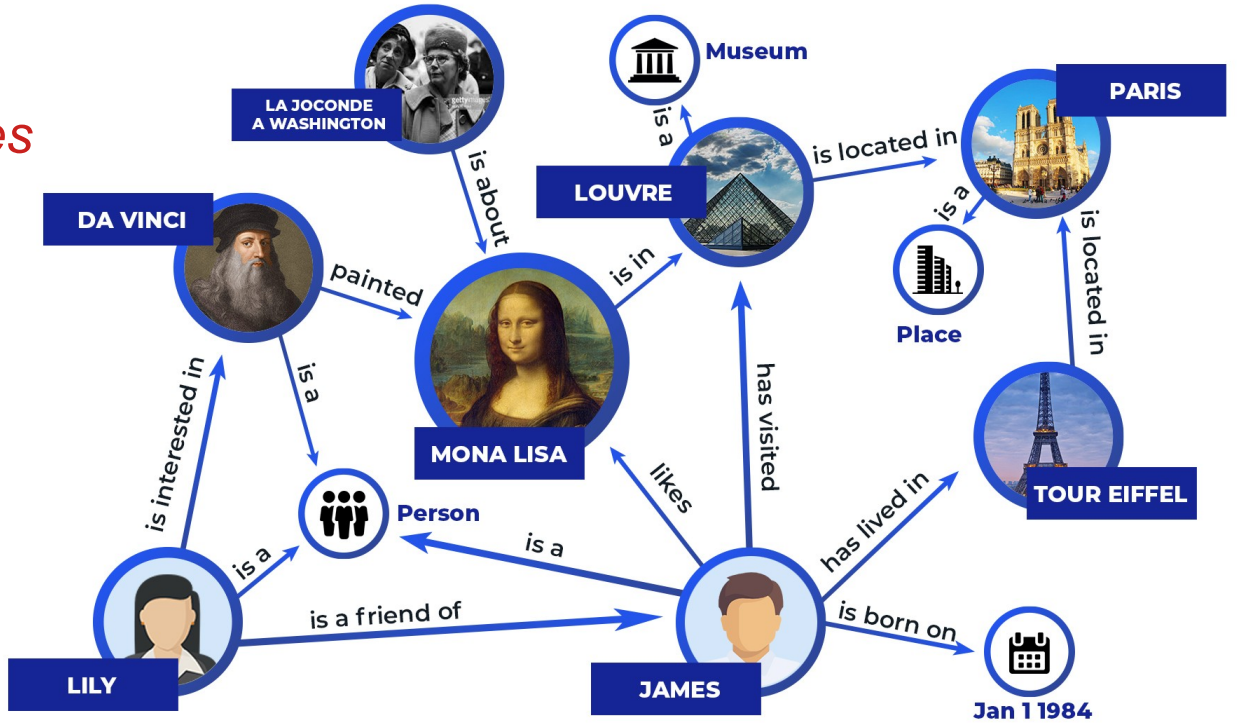
Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*
- Edges represent *relations*



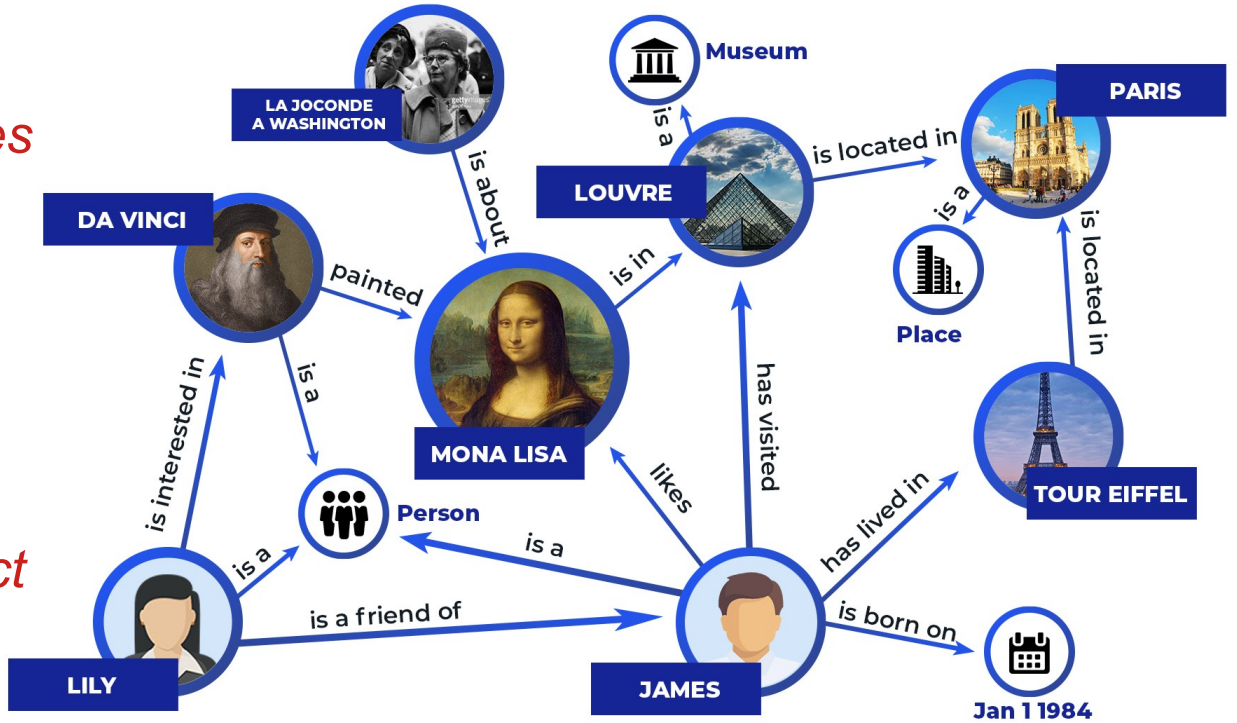
Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*
- Edges represent *relations*
- Each node–edge–node *triple* represents a *fact*



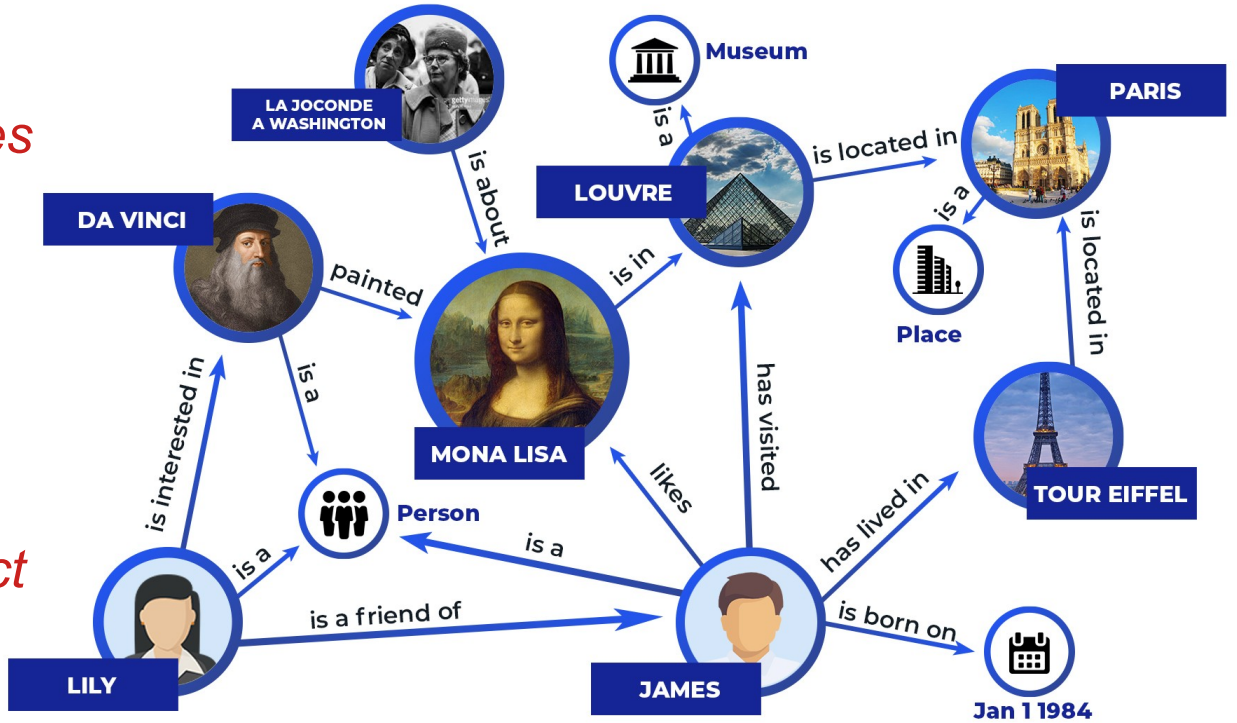
Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*
- Edges represent *relations*
- Each node–edge–node *triple* represents a *fact*
  - *subject–predicate–object*



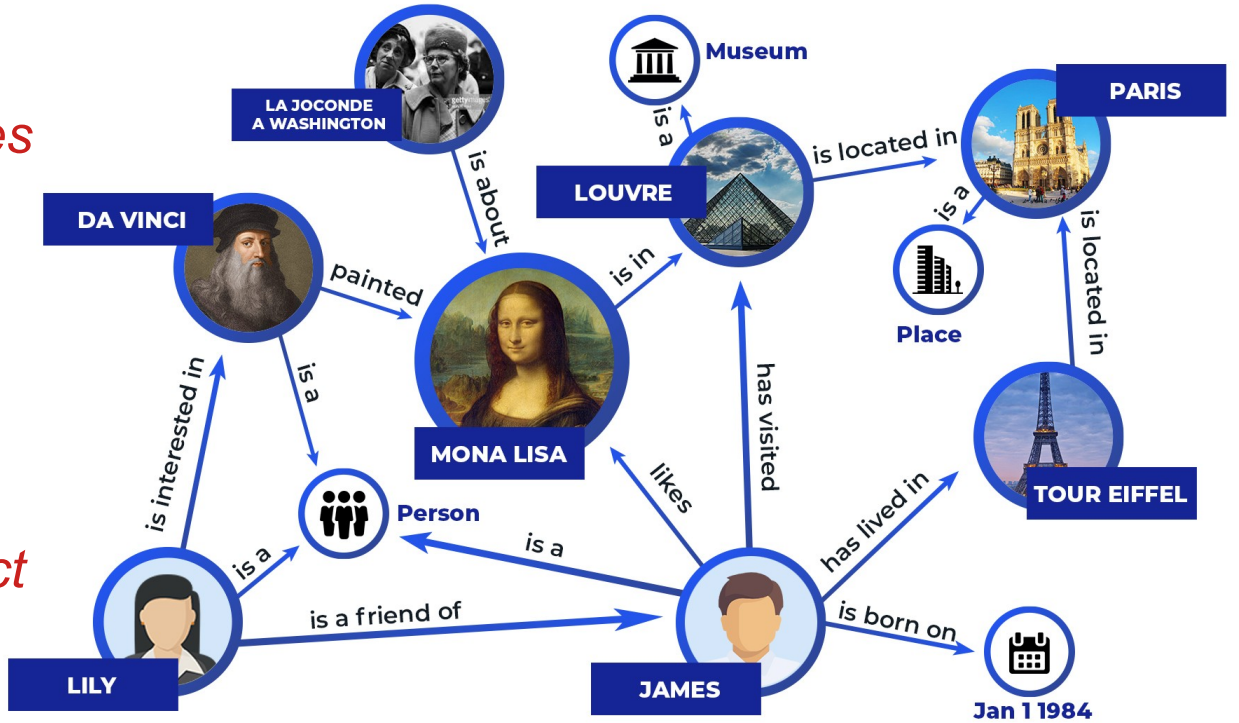
Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*
- Edges represent *relations*
- Each node–edge–node *triple* represents a *fact*
  - *subject–predicate–object*
  - *head–relation–tail*

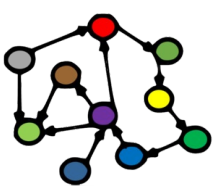


Knowledge graph

- A *graph* of *nodes* connected by directed *edges*
- Nodes can represent *resources* or *values*
- Edges represent *relations*
- Each node–edge–node *triple* represents a *fact*
  - *subject–predicate–object*
  - *head–relation–tail*
- A *knowledge graph* represents *knowledge* as connected *facts*

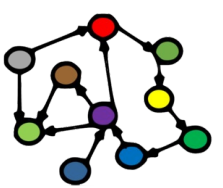


And there is more...



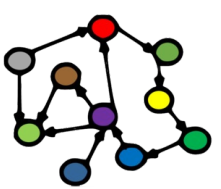
And there is more...

- Technical:
  - standard formats for storing and exchanging graphs
    - including types of values (strings, numbers, times, dates, etc.)
  - specialised databases and standard query languages
  - APIs for all major languages



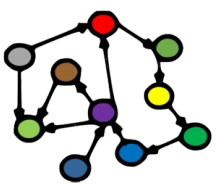
And there is more...

- Technical:
  - standard formats for storing and exchanging graphs
    - including types of values (strings, numbers, times, dates, etc.)
  - specialised databases and standard query languages
  - APIs for all major languages
- Semantic:
  - large repositories of unique identifiers for individual resources
  - vocabularies with unique identifiers for resource types and relations
  - graph embeddings and graph neural networks for machine learning



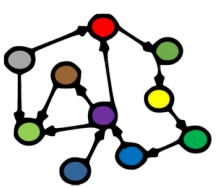
And there is more...

- Technical:
  - standard formats for storing and exchanging graphs
    - including types of values (strings, numbers, times, dates, etc.)
  - specialised databases and standard query languages
  - APIs for all major languages
- Semantic:
  - large repositories of unique identifiers for individual resources
  - vocabularies with unique identifiers for resource types and relations
  - graph embeddings and graph neural networks for machine learning
- Formal:
  - rule languages and inference engines
  - formal logic systems and reasoning engines



Why knowledge graphs?

- Ease of exchanging, reusing information
  - inherent semantics become clearer
  - less dependency on context
- Ease of interlinking, enriching information
  - semantic data can be combined in new ways
  - open reference datasets
  - general and specialised knowledge bases
- Schema independence
  - no pre-defined schemas (“schema-on-read”)
  - easy to add new types of entities and new relations
- *Well-matched with the needs of big data and machine learning!*

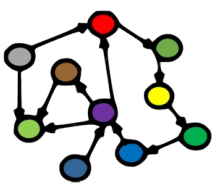


Who is using this?

- *All the big players!*
- Google's Knowledge Graph
- Microsoft's Satori
- Amazon's Product Graph
- ...and many others

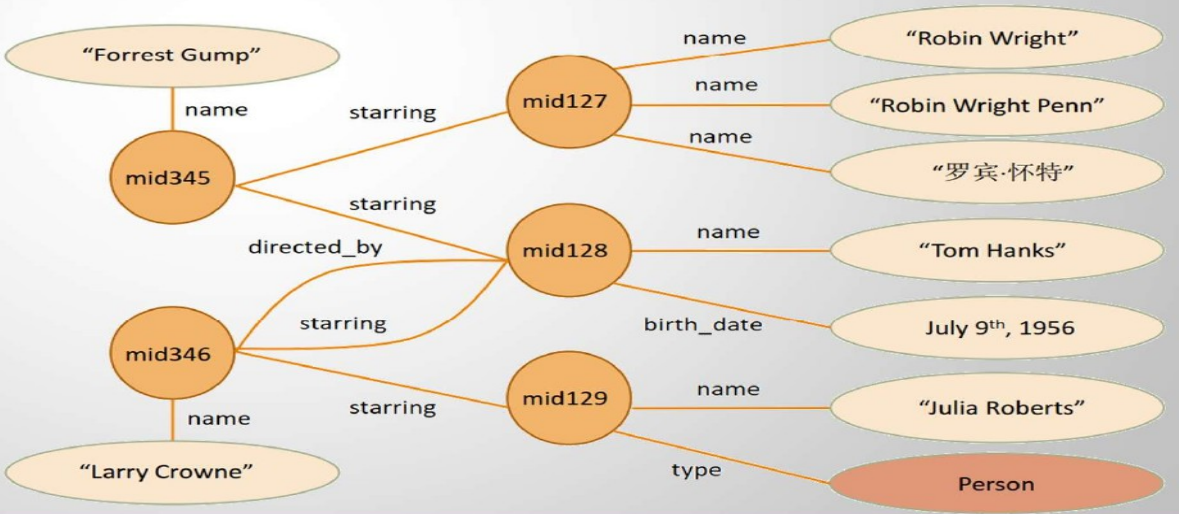
Knowledge graphs at Amazon

- Let shoppers find the best products that fit their needs
  - allow greater variation in search terms
  - allow complex queries
- Ambition: *to structure all of the world's information as it relates to everything available on Amazon*
- Describe every product on Amazon
  - concrete and abstract concepts
  - products and non-products
  - link different entities
- Enriched customer experience
  - visit Amazon to see what's new or interesting
  - discover ways to simplify and enrich their lives



Amazon

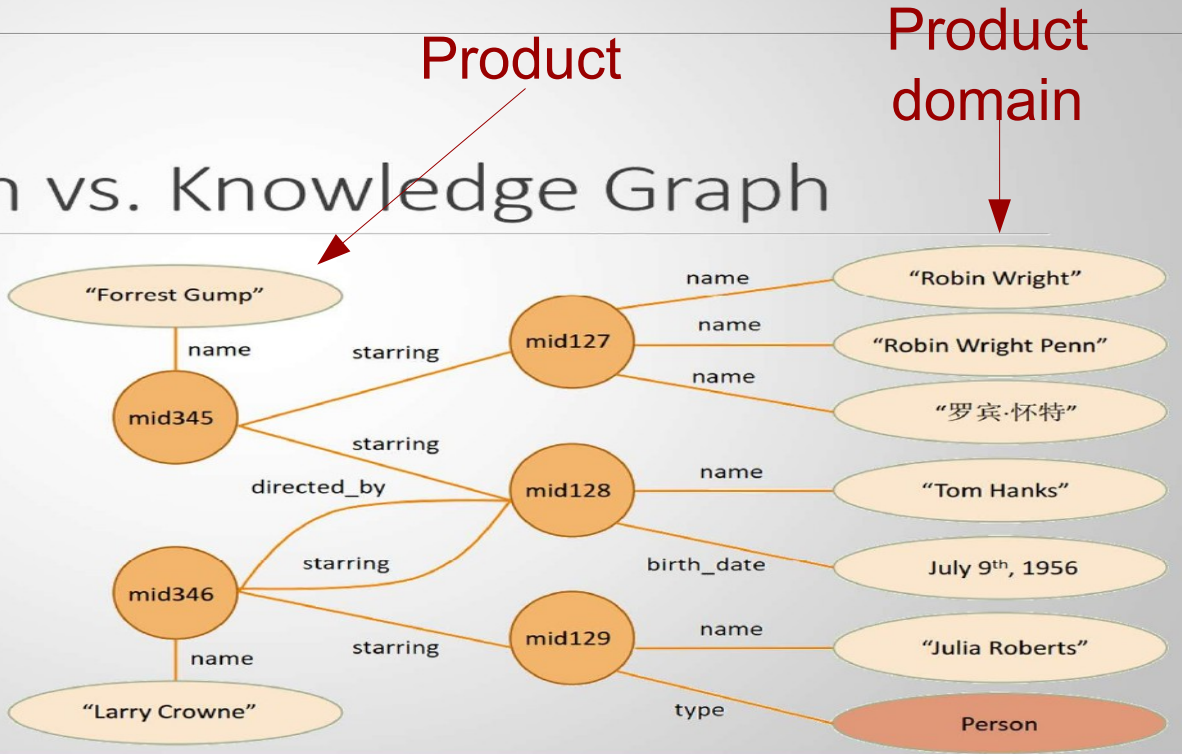
Product Graph vs. Knowledge Graph



Frank van Harmelen (2018): Keynote at CAiSE'18

Amazon

Product Graph vs. Knowledge Graph



Frank van Harmelen (2018): Keynote at CAiSE'18

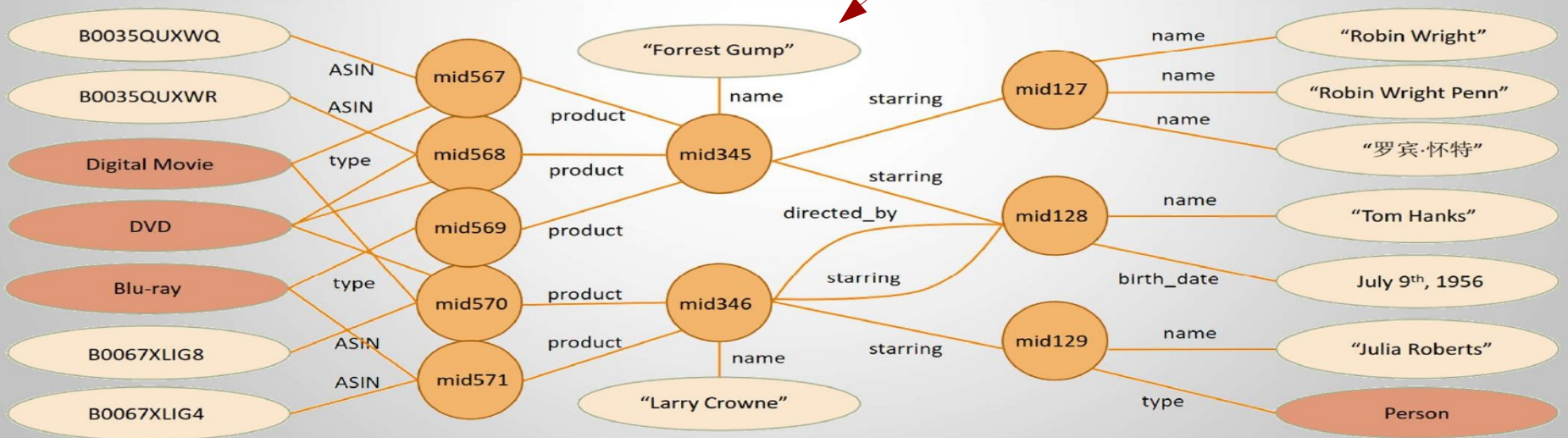
Amazon

Product details

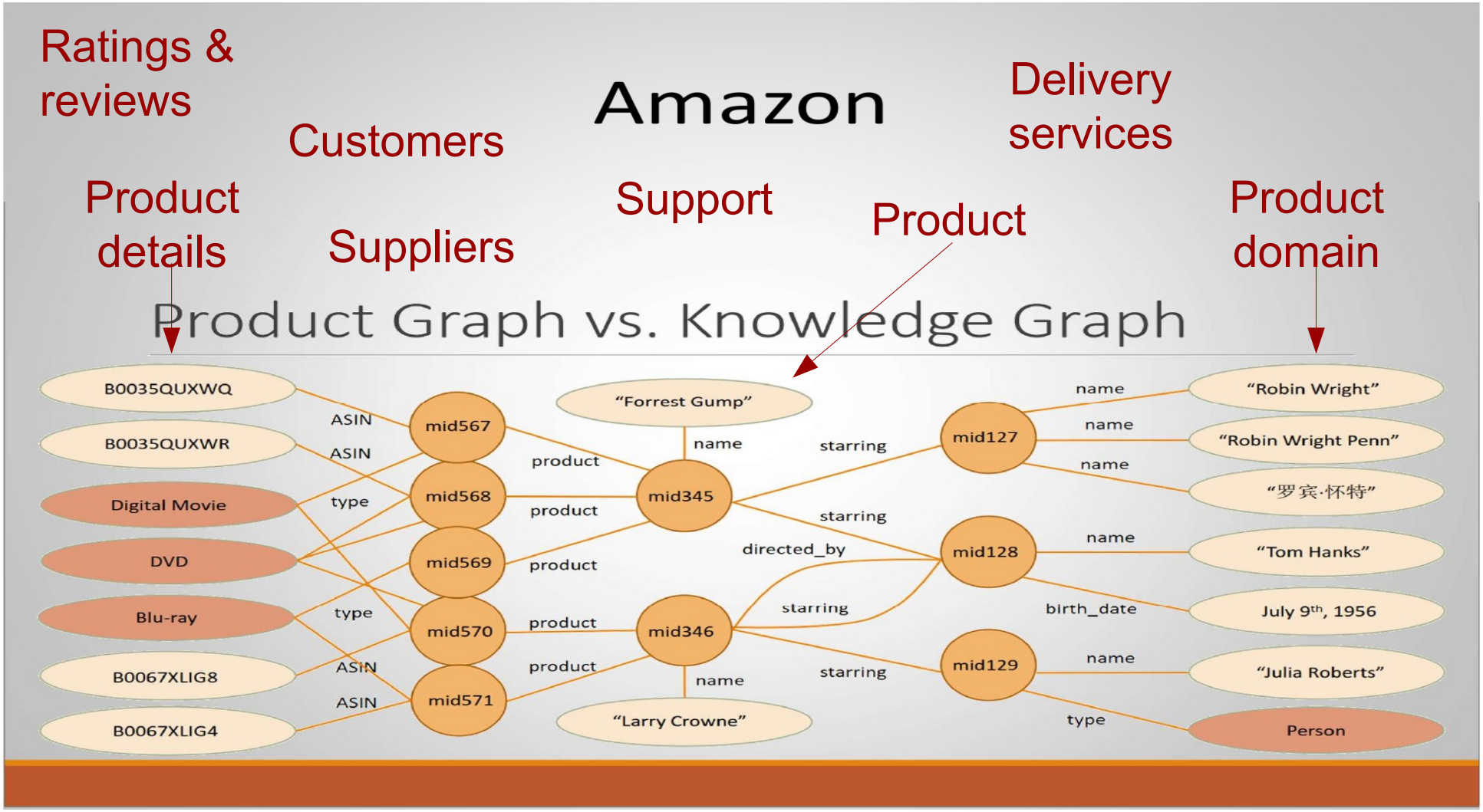
Product

Product domain

Product Graph vs. Knowledge Graph



Frank van Harmelen (2018): Keynote at CAiSE'18

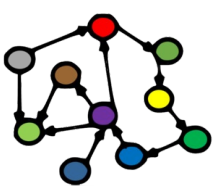


Frank van Harmelen (2018): Keynote at CAiSE'18

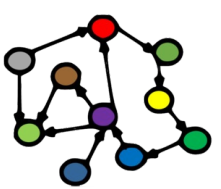
And many others...

- BBC's content management, ontologies, BBC Things
- Google, Bing, Yahoo… (schema.org) (2011)
- Google's Knowledge Graph (2012), Microsoft's Satori
- Facebook's Open Graph and Graph Search (2013)
- Thomson Reuters, Bloomberg...
- Amazon's Product Graph (2017), Neptune
- Uber Eats' food graph

Frank van Harmelen's keynote at CAiSE 2018.

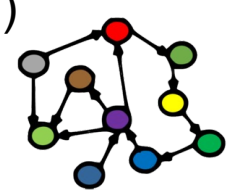


Can I have a look at one?

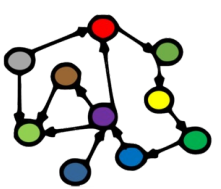


Some knowledge graphs we will look at in INFO216

- You have already seen Google's KG many times:
  - the "knowledge panels" in search results
- Wikidata (https://www.wikidata.org/)
  - part of the Wikimedia family, feeds factual information to Wikipedia
- DBpedia (https://www.dbpedia.org, https://dbpedia.org/page/Bergen)
  - extracts factual information from Wikipedia
- GeoNames (https://www.geonames.org/)
  - global database of place names (toponyms), relations and types
- BabelNet (https://babelnet.org/)
  - a multi-lingual dictionary and thesaurus
- Linked Open Vocabularies (LOV, https://lov.linkeddata.es/dataset/lov/)
  - a collection of knowledge graphs that describe vocabularies (also called ontologies) for other knowledge graphs



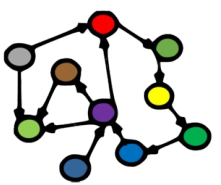
A little background



Tim Berners-Lee's call for a transition


Tim Berners-Lee

- From the early 1990-ies: creation of a *Web of Documents*
  - the "plain old web" (PoW)
  - document-centric
  - document-to-document links
  - for humans
- From the late 1990-ies: transition to a *Web of Data*
  - also called the *Semantic Web*, *Web 3.0*, *Web of Knowledge, Linked Open Data (LOD) Cloud*, *Giant Global Graph (GGG)*
  - document- *and data-centric*
  - doc-to-doc *and data-to-data links*
  - for humans *and machines*




Tim Berners-Lee's call for a transition

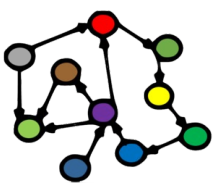

Tim Berners-Lee

- There's an enormous amount of data on the web
  - ...but the data are mostly not linked
    (think of a world wide web without document links!)
  - availability, accessibility does not go all the way
  - *what if we had standard ways of representing data so that linkable data could always be automatically linked?*
  - *enormous potential to solve, simplify, speed up... many critical information handling problems*
- This is the purpose of *semantic technologies*
- This is the vision that led to today's *semantic knowledge graphs*

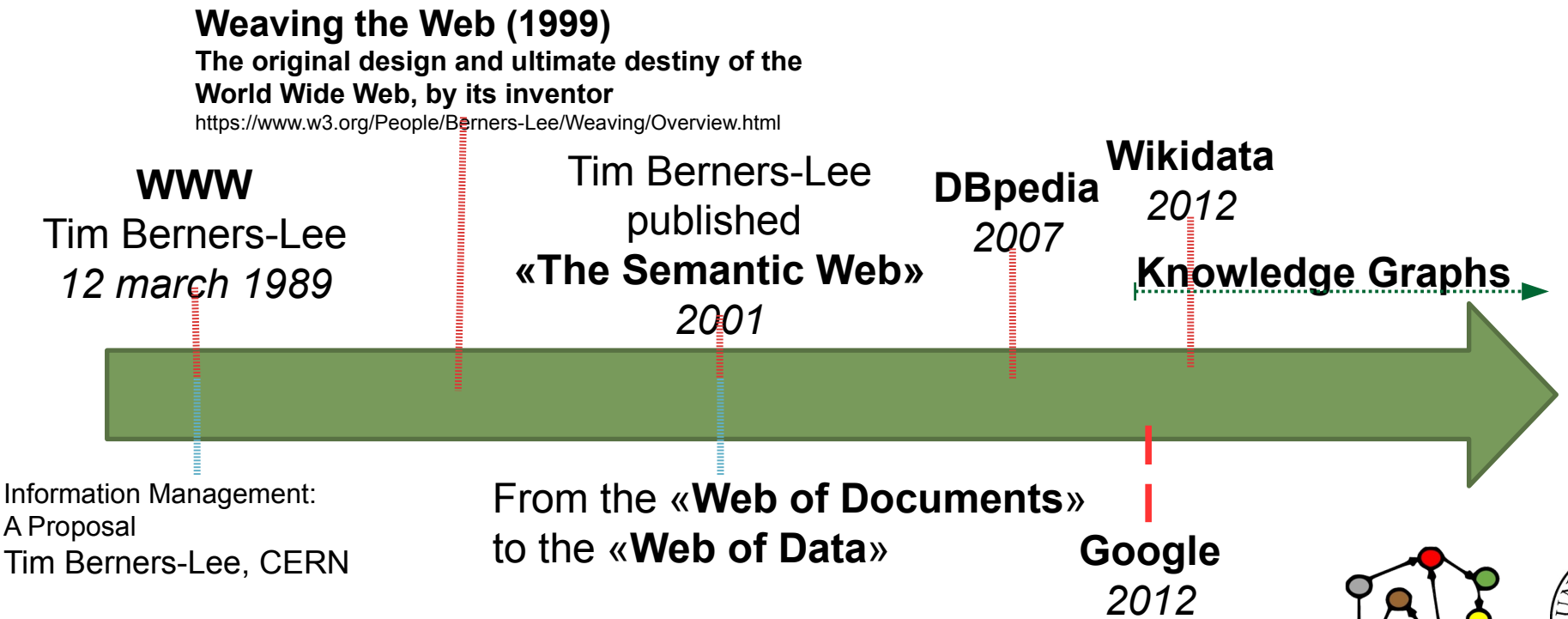
Tim Berners-Lee: <http://www.youtube.com/watch?v=HeUrEh-nqtU>

Many independent, but related developments

- The Linked Open Data (LOD) cloud:
  - interlinking semantic datasets, making them openly available: DBpedia (2007-), Wikidata (2012-), …
  - the Giant Global Graph (GGG)
- *Knowledge graphs:*
  - general term for semantic graph representations of (primarily) factual information (from 2012)
- Enterprise knowledge graphs:
  - company-internal semantic data
  - linked open data and semantic-web technologies used inside an enterprise or cluster



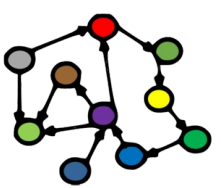
Semantic web and WWW history



Tim Berners-Lee: http://www.youtube.com/watch?v=HeUrEh-nqtU
Information Management: A Proposal: https://cds.cern.ch/record/369245/files/dd-89-001.pdf

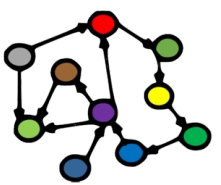
Common themes

- Graph representations of knowledge
  - RDF, RDFS, OWL, SPARQL
  - a recent alternative: labelled-property graph databases
- Semantically tagged data
  - well-defined tags (terms)
    - defined in standard vocabularies
    - formal ontologies, description logic
- From the start open, community-based
  - (re-)using many of the same standards, technologies, resources, etc.
  - openness and global interlinking

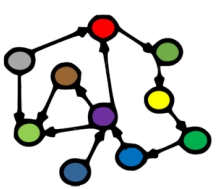


The LOD cloud

- *http://lod-cloud.net/*
  - which datasets mention resources in other datasets?
  - 1301 datasets with 16283 links between them
    - started in 2007
    - exponential-like growth for a few years
    - still growing, but more slowly now
- *How big is the LOD cloud?*
  - hard to measure exactly
  - at least 150G (150 000M) triples from ca 3000 data sets (ca 2020)
  - Wikidata is the largest general one:
    - 96M resources, 1,2G (1200M) triples

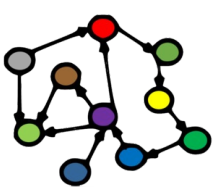


About INFO216



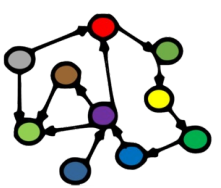
Purpose

- To learn theories, techniques, tools, and best practices for managing knowledge graphs.
- To acquire understanding and skills for programming applications that use and produce such data and metadata.
- To learn about existing sources of and standards for big, open, and semantic data.
- To gain practical experience in developing knowledge graph-based applications using technologies such as RDF, RDFS, OWL, SPARQL, and JSON-LD.



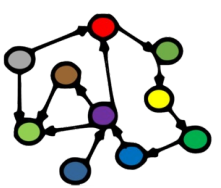
Requirements

- Required Previous Knowledge
  - INFO132 *Programming* or similar
- Recommended Previous Knowledge
  - basic data skills in *data management* and artificial intelligence
  - medium level skills in *programming*
  - for example:
    - INFO125 *Data Management*
    - INFO135 *Advanced Programming*
    - INFO180 *Methods in Artificial Intelligence*



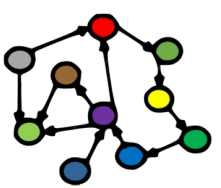
Curriculum

- Course book (*the whole book is mandatory*):
  - Allemang, Hendler & Gandon (2020).
    Semantic Web for the Working Ontologist,
    Effective Modeling for Linked Data, RDFS and OWL (Third Edition)
- Supplementary course book (*suggested, not mandatory*):
  - Blumauer & Nagy (2020).
    The Knowledge Graph Cookbook - Recipes that Work
- Additional readings (both *mandatory* and *suggested*) will be made available in the course wiki: https://wiki.uib.no/info216
- The lectures and lectures notes are also *mandatory* parts of the curriculum.



Practical

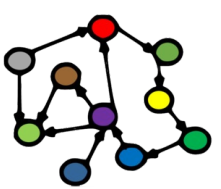
- 14 lectures:
  - Mondays 1015-1200, on Zoom
- 14 lab weeks:
  - starting next week (week 4), no labs week 15 (Easter) and 16
  - 2 hours of lab in groups + 2 hours consultation
  - seminar/lab leaders to be announced
- Evaluation:
  - individual, written 4-hour exam
  - could be shorter *if home exam*
- Requirements:
  - participation in 75% of labs: *not mandatory this spring*



Lecture plan (tentative)

1. Introduction to KGs
2. Representing KGs (RDF)
3. Querying and updating KGs (SPARQL)
4. Storing and sharing KGs
5. Open KGs
6. Enterprise KGs
7. Rules (RDFS)
8. Ontologies (OWL)
9. Vocabularies
10. Reasoning about KGs (DL)
11. Formal ontologies (OWL-DL)
12. KG embeddings
13. Knowledge engineering
14. Wrapping up

You learn KGs best through practice: do the lab exercises thoroughly!



Next week:
Representing KGs (RDF)