

தமிழ்ச் சூழலில் திறந்த இணைப்புத்  
தரவுக்கான மெய்ப்பொருளிய  
உருவாக்கம் நோக்கி

Towards Building Ontologies for Linked Open  
Data in the Tamil Context

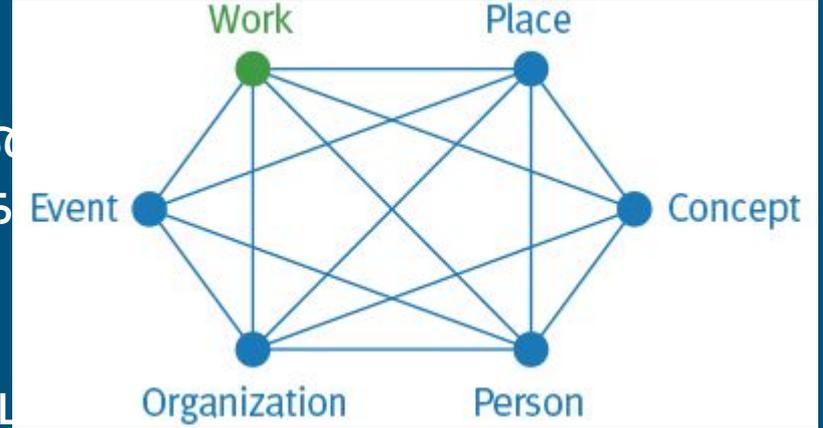
இ. நற்கீரன்

நூலக நிறுவனம் &  
ரொறன்ரொ பல்கலைக்கழகம் - ஸ்கார்புரோ

25-08-2017 - உத்தமம் த.இ.மா 2017

# பொருளடக்கம்

- கலைச்சொற்கள்
- இணைப்புத் தரவு என்றால் என்ன?
- இணைப்புத் தரவின் பயன்பாடுகள்
- மெய்ப்பொருளியம்:
  - என்றால் என்ன?
  - தேவைகளும் நோக்கங்களும்
  - வடிவமைப்புக் கொள்கைகள்
  - உருவாக்க முறையியல்
  - விபரிப்பு
  - நுட்பக் கட்டமைப்பு
- அடுத்த கட்ட பணிகள்



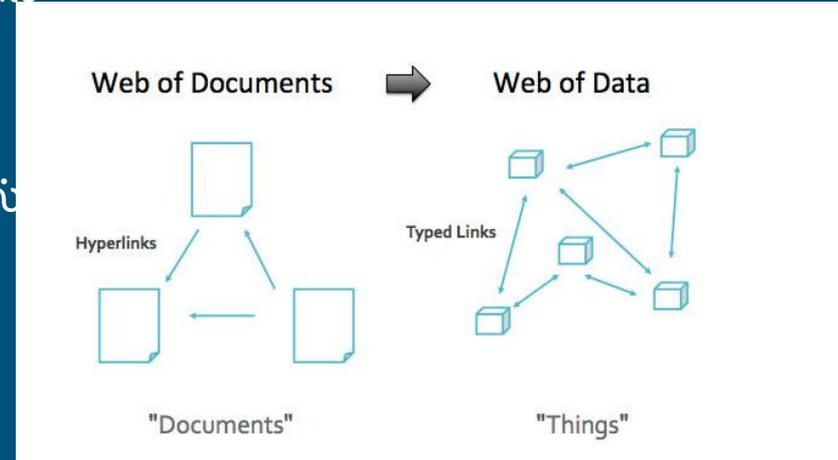
# கலைச்சொற்கள்

---

- உலகளாவிய வலை - World Wide Web
- உரலி - URL/IRI - Internationalized Reference Identifier
- இணைப்புத் தரவு - Linked Data
- பொருளுணர் வலை - Semantic Web
- மெய்ப்பொருளியம் - Ontology
- வள விபரிப்புச் சட்டகம் - Resource Description Framework
- எசுபார்க்கிள் - SPARQL Protocol and RDF Query Language
- மும்மை - Triple
- தரவுத்தளம் - Database
- மெய்ப்பொருளிய விபரிப்பு மொழிகள் - RDFS, OWL, SKOS

# இணைப்புத் தரவு என்றால் என்ன?

- வலை ஆவணங்கள் கட்டமைப்பில் இருந்து வலைத் தரவுகள் கட்டமைப்பு (web of documents to web of data)
- அமைப்பற்ற தரவில் இருந்து அமைப்புள்ள தரவுக்கு (unstructured data to structured data)
- தொடர்புசால் (Relational) அல்லது எக்ச.எம்.எல் (XML based) தரவு ஒழுங்கமைப்பில் இருந்து கோட்டுரு (Graph) தரவு அமைப்புக்கு
- ஒவ்வொரு தரவுக்கும் ஓர் உரலி
- எழுவாய் - பயனிலை - செயற்படுபொருள் (subject-predicate-object) என்ற இயற்கை மொழி வசனத்தின் அமைப்பைக் கொண்ட கூற்றுக்களால் தரவுகளை விபரித்தல் -> RDF - வளை விபரிப்பும் சட்டம்

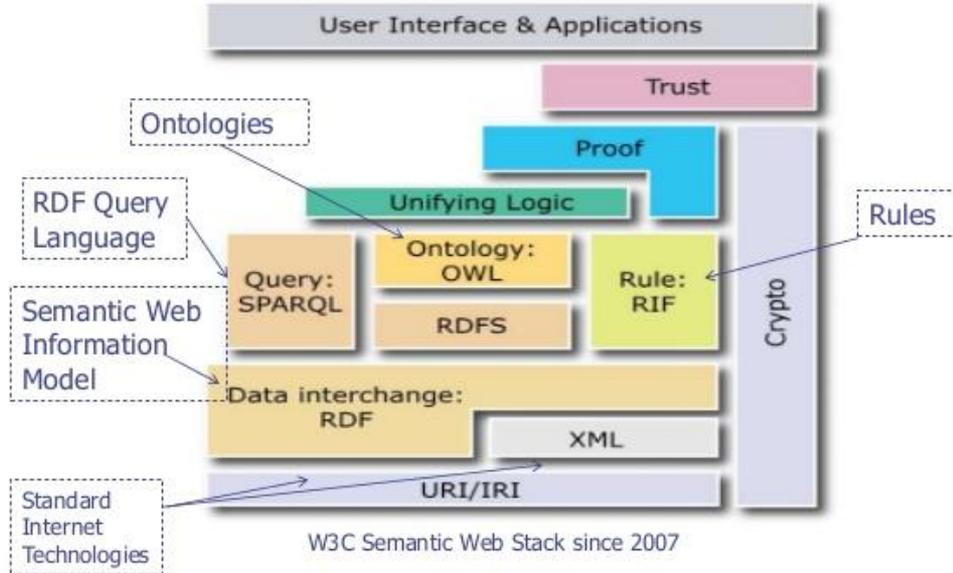


# இணைப்புத் தரவின் பயன்பாடுகள்

- தரவு மாதிரியாக்கம் & உருவகப்படுத்தல் - Data Modeling & Representation
- பகுத்தறில் & வினவுதல் - Reasoning & Querying
- தரவுப் பரிமாற்றம் - Data Exchange
- தரவு ஒருங்கிணைப்பு - Data Interchange/Integration
- தரவு/தகவல் வெளியீடு - Storing and Publishing Data
- தரவை இணைத்தல், அணுக்கப்படுத்தல் - Linking/relating the Data, Access to Data
- தரவைத் தேடல், கண்டுபிடித்தல் - Search and Discovery
- மென்பொருள் கட்டமைப்புக்கு - Software Development Platform
- Enterprise Knowledge Graphs (Google KG, FB, IBM, Siemens, SAP etc)

# நுட்ப அடுக்கு - Technology Stack

## The (current) W3C Semantic Web Stack



# மெய்ப்பொருளியம்

---

மெய்ப்பொருளியம் ஒரு குறிப்பிட்ட ஆய்வுத் துறையின் கருத்துருக்களையும் (concepts) அவற்றுக்கு இடையேயான உறவுகளையும் (relationships) பண்புகளையும் (properties) உருவகப்படுத்த, விபரிக்க, ஒழுங்குபடுத்த உதவுகின்றது.

எ.கா: FOAF, GeoNames, Dublin Core, Schema.org, Open Biomedical Ontologies

தமிழ்ச் சூழலில் நினைவு நிறுவனங்கள் தமது தரவுகளை ஒரு பொது மெய்ப்பொருளியத்தைப் பயன்படுத்தி இணைப்புத் தரவாக வெளியிடும் தேவை உள்ளது.

# மையபொருளியத்தின் தேவைகளும நோக்கங்களும் (Requirements and Objectives)

- நூலகம், ஆவணகம், அருங்காட்சியம் உட்பட்ட நினைவு நிறுவனங்கள்
- நூல்கள், ஆவணங்கள், அரும்பொருட்கள்
- உருவகப்படுத்தலில் சிக்கல்: திருக்குறளுக்கு எப்படி உரலி வழங்குவது?
- சிக்கலான தகவல் வளங்கள் (எ.கா தொடர்கள்), பலபடித்தான தகவல் வலங்களை கையாழுதல் (heterogenous information)
- வகைப்படுத்தல், ஒழுங்குபடுத்தல் (எ.கா டிவி தசம முறை)
- நிறுவனங்களுக்கு இடையே தரவுப் பரிமாற்றம்
- மையப்படுத்தப்பட்ட கட்டுப்பாடுகள் அற்றது (decentralized),  
உயர்நிலையில் இலகுவாக நினைவுகளைக் கையாள்

# தகுதிக் கேள்விகள் - Competency Questions

- இடம் - இந்த ஊரோடு தொடர்புடைய நிகழ்வுகள் எவை?
- இடம் - இந்த ஊரில் போரில் இறந்த மக்கள் யார்?
- 
- கருத்துரு - இந்தக் கருத்துப் பற்றி தொடர்புடைய ஆக்கங்கள் எவை?
- கருத்துரு - இந்தக் கருத்துரு/துறைசார்ந்த எழுத்தாளர்கள் யார்?
- நபர் - இவர் எக்காலத்தவர்? இவரின் சம காலத்தவர்கள் யார்? அவர்கள் எந்தத் துறைகளைச் சேர்ந்தவர்கள்?
- நபர் - இவர் எழுதிய நூல்கள் எந்தத் துறைகள் பற்றி பெரிதும் பேசுகின்றன?
- நபர் - இவர் எழுதிய நூல்கள் எந்தத் துறைகளைப் பற்றிப் பேசுகின்றன?

# மெய்ப்பொருளிய வடிவமைப்புக்கான கொள்கைகள்

---

ரிம் பேர்னேர்ஸ்-லீ இன் நான்கு இணைப்புத் தரவுக் கொள்கைகள்: (1) தகவல்கள் வளங்களையும், தகவல் அல்லா வளங்களையும் உரலி (URI/IRI) கொண்டு இனங்காட்டுதல் (2) HTTP ஐப் பயன்படுத்தி dereferencing செய்தல் (3) உரலியில் பயன்படக்கூடிய தகவல்களை வழங்கல் (4) பிற வளங்கோடு இணைத்து தருதல்

எளிதாக - ஏற்கனவே பயன்பாட்டி உள்ள மெய்ப்பொருளியங்களை இயன்றவரைப் பயன்படுத்த.

நேர்ப்படுத்தல் (Mapping and Alignment) சிக்கலைத் தவிர்க்க மேநிலை மெய்ப்பொருளியங்கள் (Upper Ontologies) அல்லது குறிப்பு மெய்ப்பொருளியங்கள் (Reference Ontologies) பயன்படுத்தல்

# மெய்ப்பொருளியத்தை உருவாக்குவதற்கான முறையியல்

---

- செயற்பரப்பு -> கருத்துருக்கள் -> வகுப்புகள் -> உறவுகள்/பண்புகள் -> (அடிகோல்கள்/விதிகள்/செயற்கூறுகள்)
- அனைத்துலக அருங்காட்சியகங்கள் மன்றத்தின் கருத்துரு குறிப்பு மாதிரி (CIDOC Conceptual Reference Model - சி.ஆர்.எம்)
- துறைசார் மெய்ப்பொருளியங்கள் (Domain Ontologies) பிப்ஃபேரம் (BIBFRAME), போர்ட்லன்ட் பொது தரவு மாதிரி (Portland Common Data Model), திறந்த ஆவணக தரவு மாதிரி (Open Archive Data Model), டப்பிளின் கருவகம் (Dublin Core), எசுசீமா (Schema.org), FOAF

# வடிவமைப்பு - அடிப்படை வகுப்புகள்

---

- E73 Information Object - தகவல் பொருள்
  - E78 Collection - சேகரிப்பு
  - F1 Work - படைப்பு
- E28 Conceptual Object - கருத்துருப் பொருள்
- E18 Physical Thing - பௌதீகப் பொருள்
- E21 Person - நபர்
- E74 Group - குழு
- E53 Place - இடம்
- E5 Event - நிகழ்வு
- E52 Time-Span - நேர-காலம்

# வடிவமைப்பு - வகுப்பின் உறவுகளை, பண்புகளை வரையறை செய்தல்

NF Ontology - Classes and Properties ☆

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

fx

	A	B	C	D	E	F	G
1	Term	Description	Type	Domain	Range	Example Values	Usage Comments
2	rdf:type		Relation			schema:Person	Required
3	skos:prefLabel		Property				Required
4	skos:note		Property				Required
5	schema:familyName		Property				Required
6	schema:givenName		Property				
7	schema:honorificPrefix		Relation ?	Person	Set: Mr. Mrs. Phd		
8	schema:birthDate		Relation	Person	xsd:date or Event?		
9	schema:deathDate		Relation		xsd:date or Event?		
10	dbp:knownFor		Property	Person			
11	schema:memberOf		Relation				
12	schema:relatedTo		Relation				
13	dbo:birthPlace		Relation	Person	Place		
14	dbo:residence		Relation	Person	Place		
15	dbo:citizenship		Relation ?	Person	Place		
16	dbo:deathPlace		Relation	Person	Place		
17							

# ஐலண்டோரா குளோவில் (Islandora CLAW) நிறுவுதல்

---

Next generation digital repository platform. Fully open source. Horizontally scalable.

## Technical Stack

- Drupal 8 - CMS
- Fedora - Linked Data Platform
- BlazeGraph - Triplestore
- Alpaca (Apache Camel + ActiveMQ) - the plumbing that connects them together
- Micro Services (ex Derivative Creation, FITS, OCR etc)

# அடுத்த கட்டப் பணிகள்

- ஈடுபாடு உள்ளவர்களை இணைத்து, மேலதிக பயனர் கதைகளை (user stories) கருத்தில் எடுத்தல்
- வகுப்புகளை, உறவுகளை, பண்புகளை இறுதிப்படுத்தல்
- OWL போன்ற மெய்ப்பொருளிய வடிவமைப்பு மொழியில் வெளியிடுதல்
- முன்மாதிரிகளை உருவாக்கல்
- இணைந்து செயற்படுவோம்

