OVERCOMING RESISTANCE TO TECHNOLOGY CHANGE

A LINKED DATA PERSPECTIVE

PHUSE EU CONNECT 2018
Frankfurt, Germany
November, 2018

Views are my own, not my employer's, not PhUSE...
ALTERNATE TITLE
BRIDGING THE GAP BETWEEN THE LINKED DATA NERDS
AND EVERYONE ELSE
INSPIRED BY RECENT EVENTS

"Linked Data is really cool, but..."

1. "...I can already do that in a relational database."
   - US Connect18

2. "...I don't see how to apply it to what we do."
   - US CSS 2018

3. "...it will never happen in pharma."
   - EU Connect17
"Linked Data is really cool, but..."

4. "...I need a solution today."
   ○ *Almost everyone in Pharma?*
WHAT IS LINKED DATA?
RESOURCE DESCRIPTION FRAMEWORK (RDF)

Subject

eg:Drug1

ncit:study

eg:Study1

Predicate

Identifier

Drug 1

Key=Value Pair

study=Study1

“Drug1 has study Study1”

Unique identification of "things"
  • Define once, use many
WHAT IS A **KNOWLEDGE GRAPH**?

**LINKED DATA BACKED BY AN ONTOLOGY**

- Model classes of things in a domain
- Relationships, Rules, Restrictions
- Ontology is RDF
- Infer new information using a Reasoner
- Semantic Web
  - Original vision for the internet
"Linked Data is really cool, but..."

1. "...I can already do that in a relational database."
   - US Connect18

**Technology**

**Argument:**
- No need to change.
- RDBMS satisfies current (and future) requirements.
RDBMS STRENGTHS

- Transactional
- High-velocity
- Skill set
- Vendors
- Validated
- *Et cetera*

Comfort Zone
GRAPH DATABASE STRENGTHS

- One value One Time
- Web Ontology Language : semantics, inference
- Multidimensional
- Flexible and extendable schema
- ...and more!
Technology

CONCLUSION

- RDBMS & Graph DB Coexist
- Bridge Data Silos with Graph Databases
- Develop Ontology Expertise
- Build Enterprise Knowledge Graphs
"Linked Data is really cool, but..."

2. "...I don’t see how to apply it to what we do."
   - US CSS 2018

Practical Application

Responsible:
- Linked Data Nerds
- Standards Organizations (industry + internal)
- Everyone
FUNDAMENTALLY CHANGE HOW WE THINK ABOUT DATA

*When I say* **data**, you think:

<table>
<thead>
<tr>
<th>studyid</th>
<th>domain</th>
<th>usubjID</th>
<th>vsseq</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDISCPILOT01</td>
<td>VS</td>
<td>01-701-1015</td>
<td>1</td>
</tr>
<tr>
<td>CDISCPILOT01</td>
<td>VS</td>
<td>01-701-1015</td>
<td>2</td>
</tr>
<tr>
<td>CDISCPILOT01</td>
<td>VS</td>
<td>01-701-1015</td>
<td>3</td>
</tr>
<tr>
<td>CDISCPILOT01</td>
<td>VS</td>
<td>01-701-1015</td>
<td>4</td>
</tr>
<tr>
<td>CDISCPILOT01</td>
<td>VS</td>
<td>01-701-1015</td>
<td>5</td>
</tr>
<tr>
<td>CDI</td>
<td>dm</td>
<td>01-701-1015</td>
<td>1015</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1023</td>
<td>1023</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1028</td>
<td>1028</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1033</td>
<td>1033</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1034</td>
<td>1034</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1047</td>
<td>1047</td>
</tr>
<tr>
<td>CDI</td>
<td>DM</td>
<td>01-701-1057</td>
<td>1057</td>
</tr>
</tbody>
</table>
FUNDAMENTALLY CHANGE HOW WE THINK ABOUT DATA

When I say **data**, you think:
LINKED DATA NERDS: IMPROVE

ELITISM

"It almost seemed as if the authors intentionally wanted to encourage the perception of high priesthood, only to be fathomed by the chosen few."
- David McComb
"People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated."
- Dan Brickley and Libby Miller
IS THIS SIMPLE?
IS THIS *BETTER*??

STUDY CDISCPILOT01, SUBJECT 1015
DM, VS, EX : SUBSET OF 250 TRIPLES
FOCUS ON PRACTICAL SOLUTIONS

Problems? We've got a few...
SUBMISSIONS TO FDA
Conformance Criteria

68% No Issues
32% At least 1 issue

D+
SUBMISSIONS TO FDA

Janus Database Uploads

80% Succeed

20% Fail

B-
MORE PROBLEMS

- Quality
- Integration
- Traceability and Provenance
- Standards
- Metadata
- Human Error
- ...


SOLUTION: KNOWLEDGE GRAPHS

- Clinical trial data as RDF backed by Taxonomies, Ontologies (+Inference)
- Entities represented across the clinical trial lifecycle.

WARNING: Scope issue.
Linked Data Business Case

Image Attribution: https://bit.ly/2x0Hjmd
"You can't just buy a vendor product to catch up. [It] Takes a long term investment in skills development."
- Dave Newman #EDW18

• But the benefits are immense
• And there is a practical path!
THE ROOFSHOT / MOONSHOT MANIFESTO

**Roofshot**
*Incremental impacts*
SDTM domains as RDF

**Moonshot**
*Invent & apply state-of-the-art*
**Knowledge Graph**
Clinical trial lifecycle

Concept & Image Attribution: https://rework.withgoogle.com/blog/the-roofshot-manifesto/
Practical Application

CONCLUSION

- **Roofshot**: Prototypes & Projects
  - PhUSE
  - You, at your company
- **Moonshot**
  - Ontologies, Taxonomies
  - Standards
  - End-to-end
"Linked Data is really cool, but..."

3. "...it will never happen in pharma."
   - EU Connect17

4. "...I need a solution today."
   - Almost everyone in Pharma?

Adoption

- Challenges
- Approaches
"NEED A SOLUTION TODAY"

- Fails to plan for the future
- Perpetuates antiquated SAS XPT

PLAN FOR THE FUTURE TODAY

- Graph may not mature for Pharma for 2-3 years
- Start adoption NOW.
EVERYONE: IMPROVE

PESSIMISM

Excuses:
• We’ve always done it this way.
• It is not validated.
• The regulatory agencies...
• The standards agencies...
• My IT department....
• ...and on...and on...and on...
TIMELINE OF DISILLUSIONMENT

1980: Created (1983)
1990: Candy of the Year (1985)
2000: Not the best candy
2010: Not the best format
2020: Linked Data?

XPT: Development (late 1980’s)
HOW TO BREAK THROUGH?

... a clue from another nerd.
"...the old medium is always the content of the new medium."  
PHUSE PROJECT: CLINICAL TRIAL DATA AS RDF
WHO SHOULD LEAD ADOPTION?

- Industry?
- Regulatory?
- Standards: CDISC? TransCelerate?
- PhUSE?

YOU!
Adoption
CONCLUSION

• Several large pharma investing in Knowledge Graphs
BE INSPIRED!

"Linked Data is really cool, and..."

1. "...it will complement RDBMS."
2. "...there is a practical adoption path."
3. "...it IS HAPPENING."

- EU Connect18
Thank you!