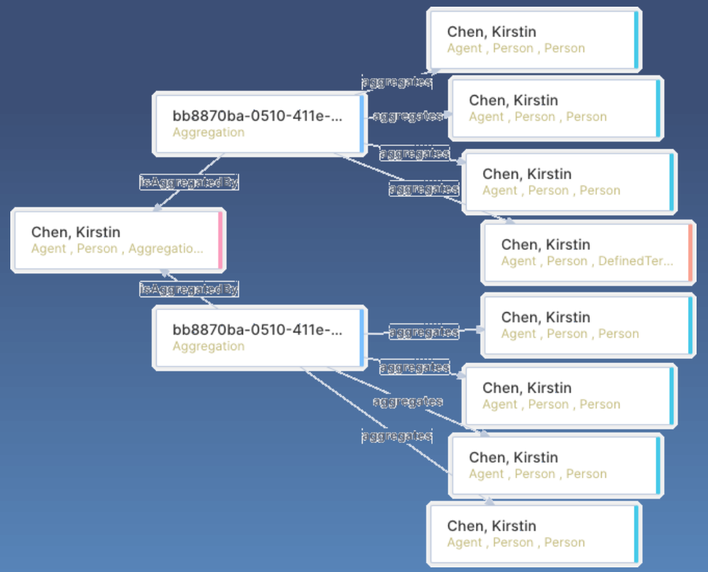
From Ambition to Go Live

The National Library Board of Singapore's Journey to an operational Linked Data Management and Discovery System



Richard Wallis
Evangelist and Founder
Data Liberate
richard.wallis@dataliberate.com
@dataliberate

15th Semantic Web in Libraries Conference
SWIB23 - 12th September 2023 - Berlin

CC BY

Data Liberate

Data Liberate

Independent Consultant, Evangelist & Founder

40+ Years – Computing
30+ Years – Cultural Heritage technology
20+ Years – Semantic Web & Linked Data

W3C Community Groups:

- Bibframe2Schema (Chair) – Standardised conversion path(s)
- Schema Bib Extend (Chair) - Bibliographic data
- Schema Architypes (Chair) - Archives
- Financial Industry Business Ontology – Financial schema.org
- Tourism Structured Web Data (Co-Chair)
- Schema Course Extension
- Schema IoT Community
- Educational & Occupational Credentials in Schema.org

Worked With:

- **Google** – Schema.org vocabulary, site, extensions. documentation and community
- **OCLC** – Global library cooperative
- **FIBO** – Financial Industry Business Ontology Group
- Various Clients – Implementing/understanding Linked Data, Schema.org:
  National Library Board Singapore
  British Library — Stanford University — Europeana

richard.wallis@dataliberate.com — @dataliberate

National Library Board Singapore

Public Libraries

Network of 28 Public Libraries, including 2 partner libraries*

Reading Programmes and Initiatives

Programmes and Exhibitions targeted at Singapore communities

*Partner libraries are libraries which are partner owned and
funded but managed by NLB/NLB's subsidiary Libraries and Archives Solutions Pte Ltd. Library@Chinatown and the Lifelong Learning Institute Library are Partner libraries.

National Library

Preserving Singapore's Print and Literary Heritage, and Intellectual memory
Reference Collections
Legal Deposit (including electronic)

National Archives

Transferred from NHB to NLB in Nov 2012

Custodian of Singapore's Collective Memory: Responsible for Collection, Preservation and Management of Singapore's Public and Private Archival Records

Promotes Public Interest in our Nation's History and Heritage

Data Liberate

National Library Board Singapore

Lending Collection

Over 5m
print collection

Over 1.7m
e-books and audio books

78
databases

Over 2.4m
music tracks

Over 7,400
e-newspapers and e-magazines titles

Over 8,000
e-learning courses

Reference Collection

Over 560,000
Singapore & SEA items

Over 147,000
Chinese, Malay & Tamil Languages items

Over 53,000
Arts items

Over 62,000
Social Sciences & Humanities items

Over 39,000
Science & Technology items

Over 19,000
Rare Materials items

Archival Materials

Over 1.14m
Photographs

Over 290,000
Government files & Parliament papers

Over 190,000
Audiovisual & sound recordings

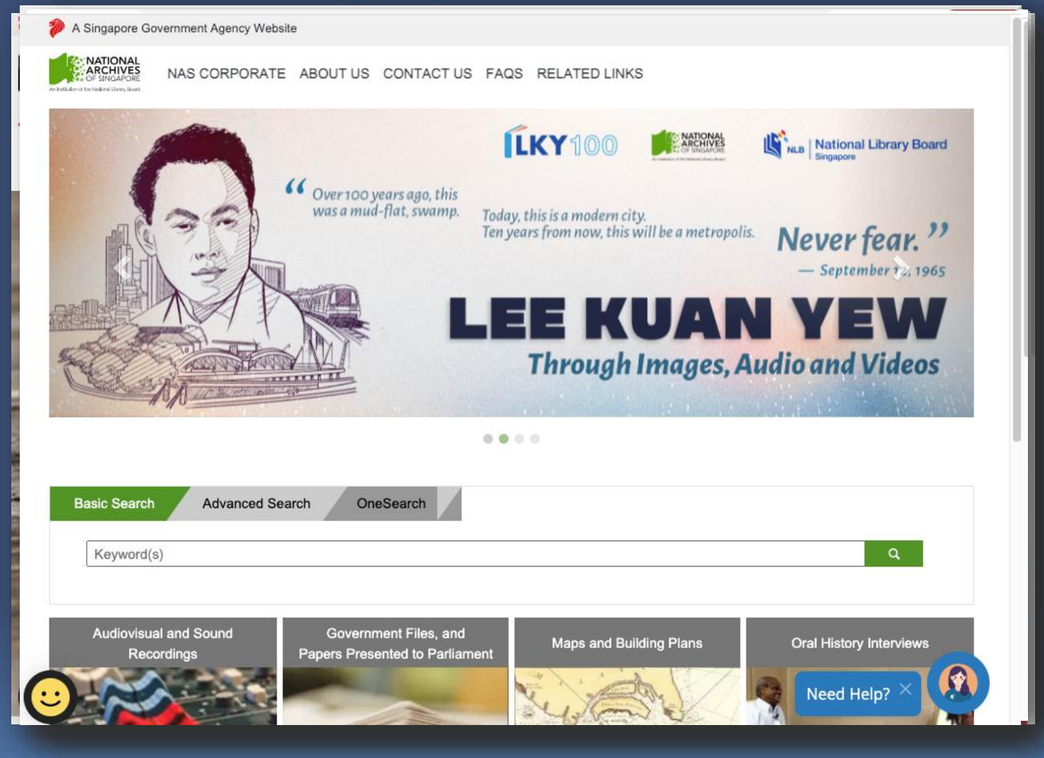
Over 70,000
Maps & building plans

Over 55,000
Speeches & press releases

Over 35,000
Oral history interviews

Over 7,000
Posters

National Library Board Online Services



Data Liberate

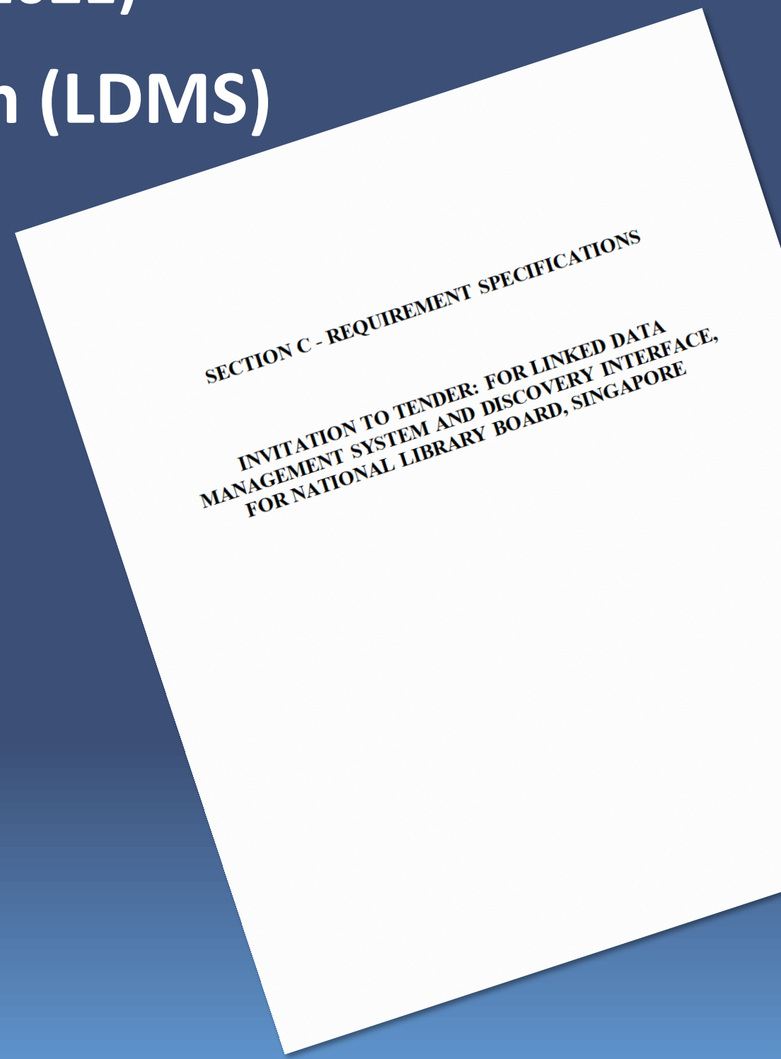
The Ambition

- To enable the discovery & display of entitles from different sources in a combined interface
- To bring together resources physical and digital
- To bring together diverse systems across the National Library, National Archives, and Public Libraries in a Linked Data Environment
- To provide a staff interface to view and manage all entities, their descriptions and relationships

The Journey Starts (March 2021)

A Linked Data Management System (LDMS)

- Cloud-based System
- Management Interface
- Source data curated at source
  - ILS, Authorities, CMS, NAS
- Linked Data
- Public Interface – Google friendly
  - Schema.org
- Shared standards
  - Bibframe, Schema.org


SECTION C - REQUIREMENT SPECIFICATIONS

INVITATION TO TENDER: FOR LINKED DATA MANAGEMENT SYSTEM AND DISCOVERY INTERFACE, FOR NATIONAL LIBRARY BOARD, SINGAPORE

Contract Awarded

metaphactory platform
Low-code knowledge graph platform
Semantic knowledge modeling
Semantic search & discovery

AWS Partner
Public sector partner
Singapore based

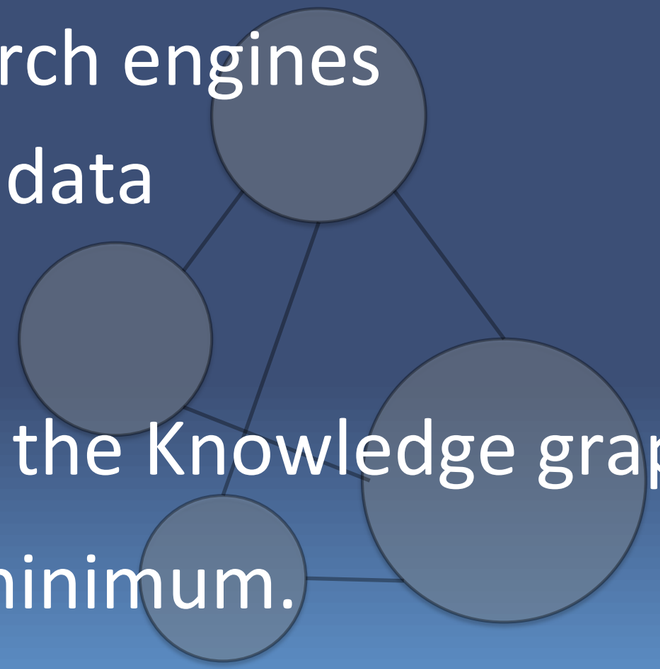
Linked Data, Structured data, Semantic Web, bibliographic meta data, Schema.org and management systems consultant

Initial Challenges

- Data – lots of it!
  - Record-based formats: MARC-XML, DC-XML, CSV
- Entities described in the data – lots and lots!
- Regular updates – daily, weekly, monthly
- Automatic ingestion
- Not the single source of truth
- Manual update capability
- Entity-based search & discovery
- Web friendly – Schema.org output

Basic Data Model

- Linked Data
  - BIBFRAME to capture detail of bibliographic records
  - Schema.org to deliver structured data for search engines
  - Schema.org representation of CMS, NAS, TTE data
  - Schema.org enrichment of BIBFRAME
- Schema.org as the ‘lingua franca’ vocabulary of the Knowledge graph
  - All entities described using Schema.org as a minimum.



Data Data Data!

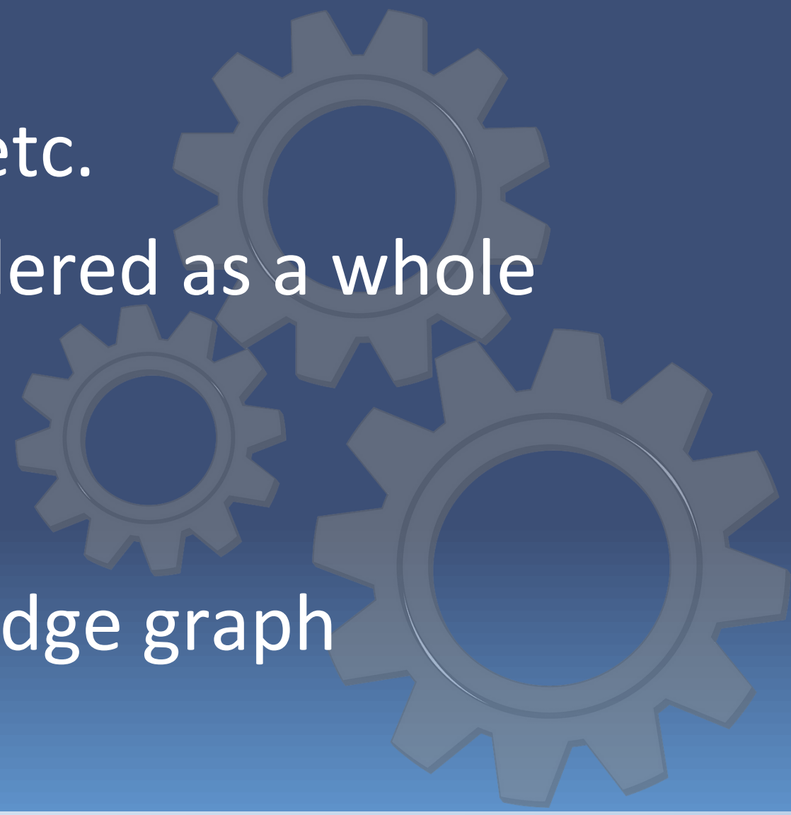
Data Source	Source Records	Entity Count	Update Frequency
ILS	1.4m	56.8m	Daily
CMS	82k	228k	Weekly
NAS	1.6m	6.7m	Monthly
TTE	3k	317k	Monthly
	3.1m	**70.4m**	

Data Ingest Pipelines – ILS

- Source data in MARC-XML files
- Step 1: marc2bibframe2 scripts
  - Open source – shared by Library of Congress
  - “standard” approach
  - Output BIBFRAME as individual RDF-XML files
- Step 2: bibframe2schema.org script
  - Open source – Bibframe2Schema.org Community Group
  - SPARQL-based script
  - Output Schema.org enriched individual BIBFRAME RDF files for loading into Knowledge graph triplestore

Data Ingestion Pipelines – TTE (Authorities)

- Source data in bespoke CSV format
- Exported in collection-based files
  – People, Places, organisations, time periods, etc.
- Interrelated references so needed to be considered as a whole
- Bespoke python script
- Creating Schema.org entity descriptions
- Output RDF format file for loading into Knowledge graph triplestore

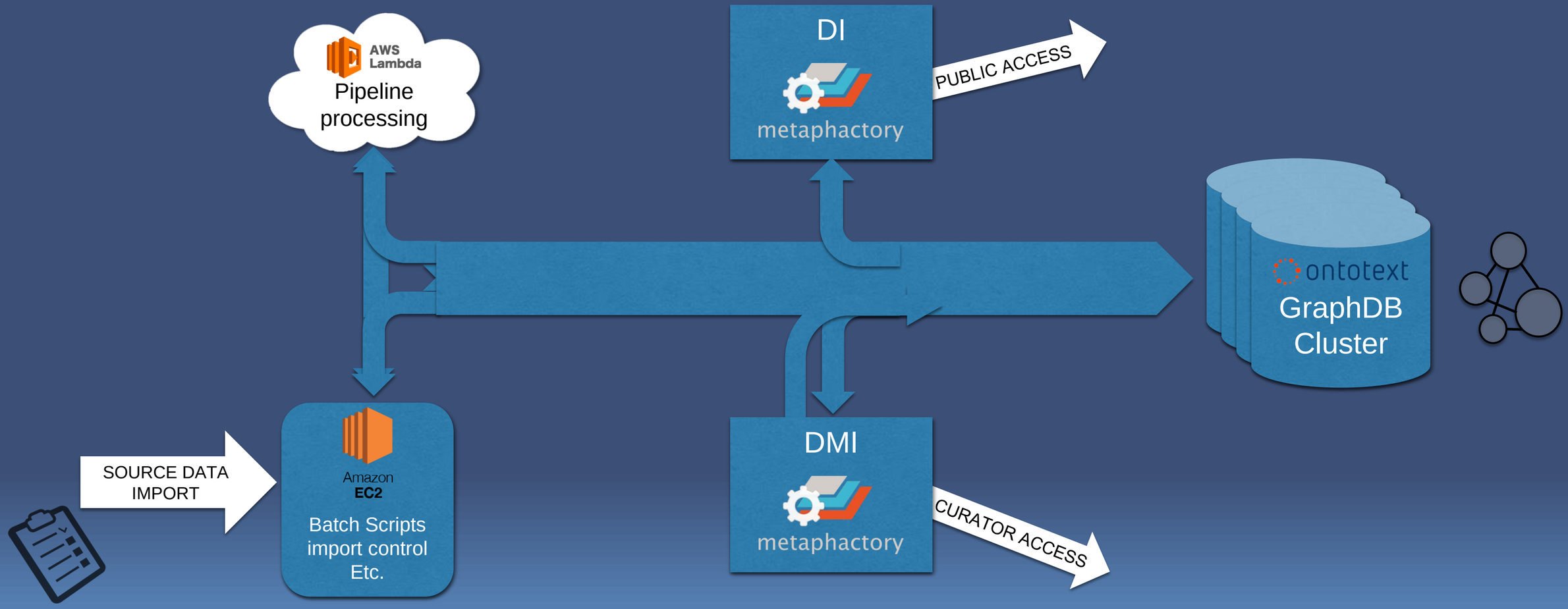


Data Ingestion Pipelines – NAS & CMS

- Source data in DC-XML files
- Step 1: DC-XML to DC-Terms RDF
  - Bespoke XSLT script to translate from DC record to DCT-RDF
- Step 2: TTE lookup
  - Lookup identified values against TTE entities in Knowledge graph
    - use URI if matched
- Step 3: Schema.org entity creation
  - SPARQL script create schema representation of DCT entities
- Output individual RDF format files for loading into Knowledge graph

Technical Architecture (simplified)

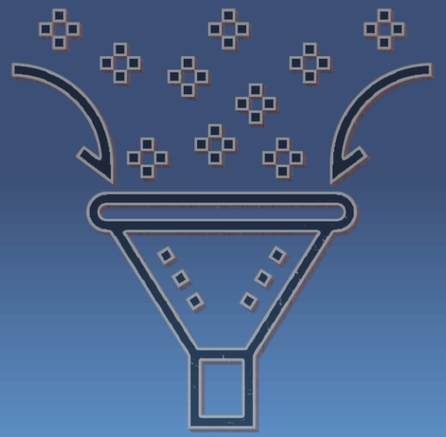
Hosted on Amazon Web Services



Data Liberate

A need for entity reconciliation …..

- Lots (and lots and lots) of source entities – 70.4 million entities
- Lots of duplication
  - Lee, Kuan Yew – 1st Prime Minister of Singapore
    - 160 individual entities in ILS source data
  - Singapore Art Museum
    - Entities from source data
    - 21 CMS, 1 NAS, 66 ILS, 1 TTE
- Users only want 1 of each!

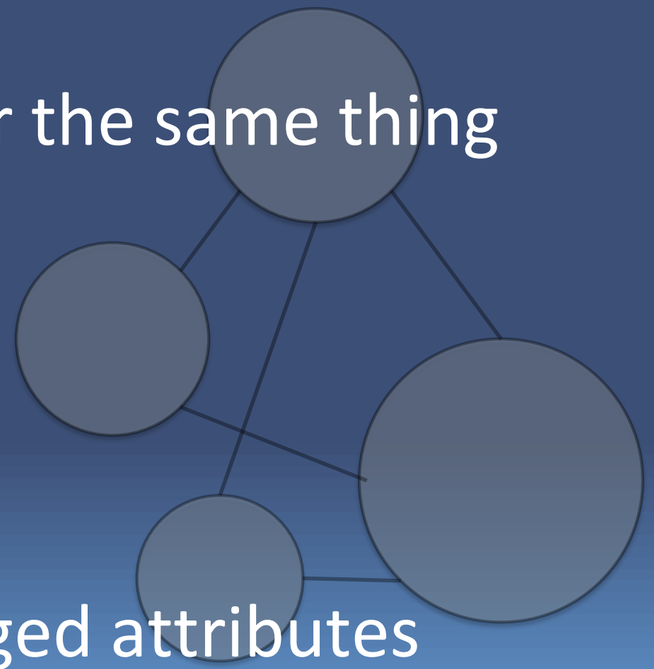


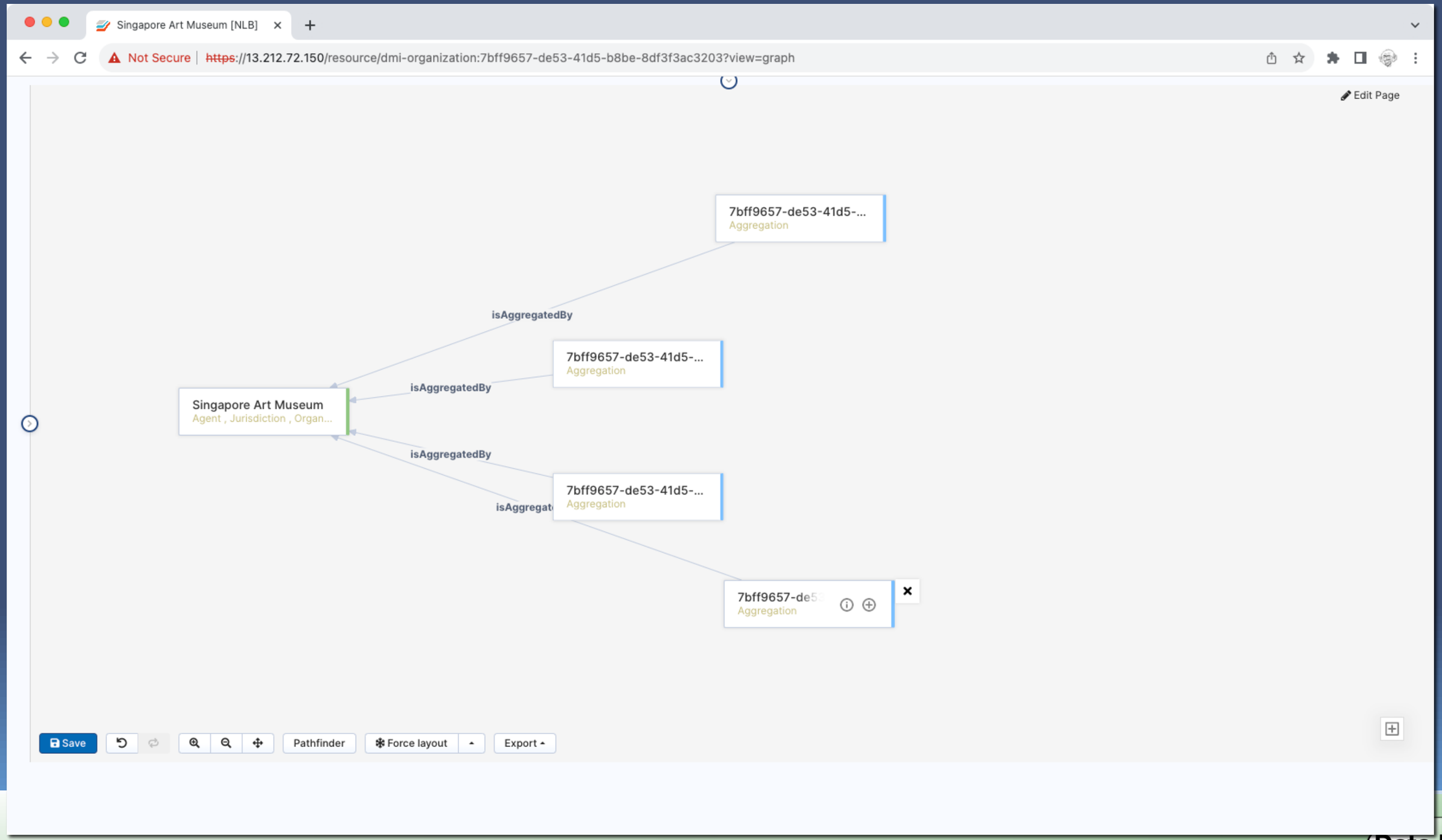
70.4 (million) into 6.1 (million) entities does go!

- We know that now - but how did we get there?
- Requirements hurdles to be cleared:
  - Not a single source of truth
  - Regular automatic updates – add / update / delete
  - Manual management of combined entities
    - Suppression of incorrect or ’private’ attributes from display
    - Addition of attributes not in source data
    - Creating / breaking relationships between entities
- Near real-time updates

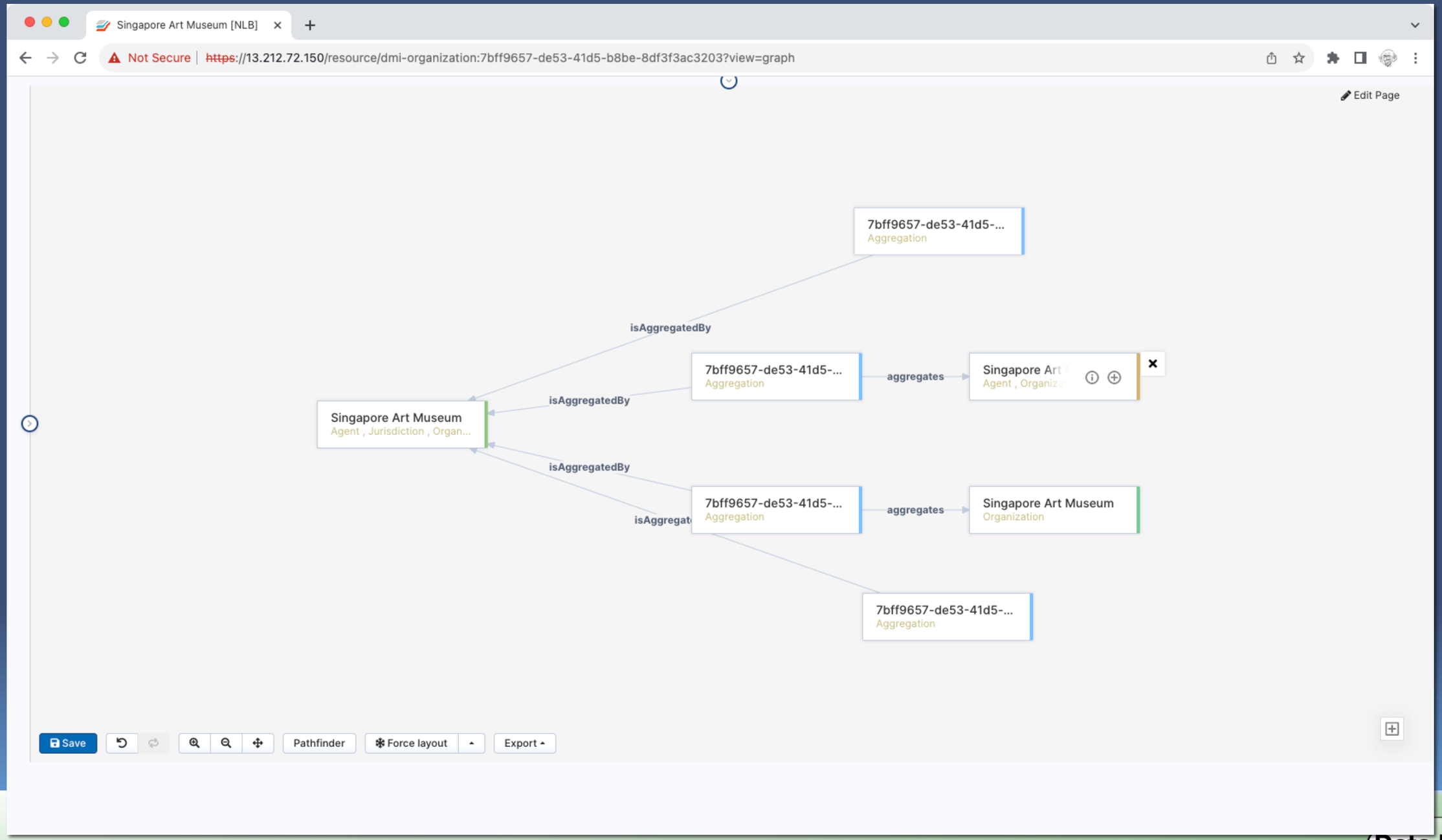
Adaptive Data Model Concepts

- **Source entitles**
  - Individual representation of source data
- **Aggregation entities**
  - Tracking relationships between source entities for the same thing
  - No copying of attributes
- **Primary Entities**
  - Searchable by users
  - Displayable to users
  - Consolidation of aggregated source data & managed attributes

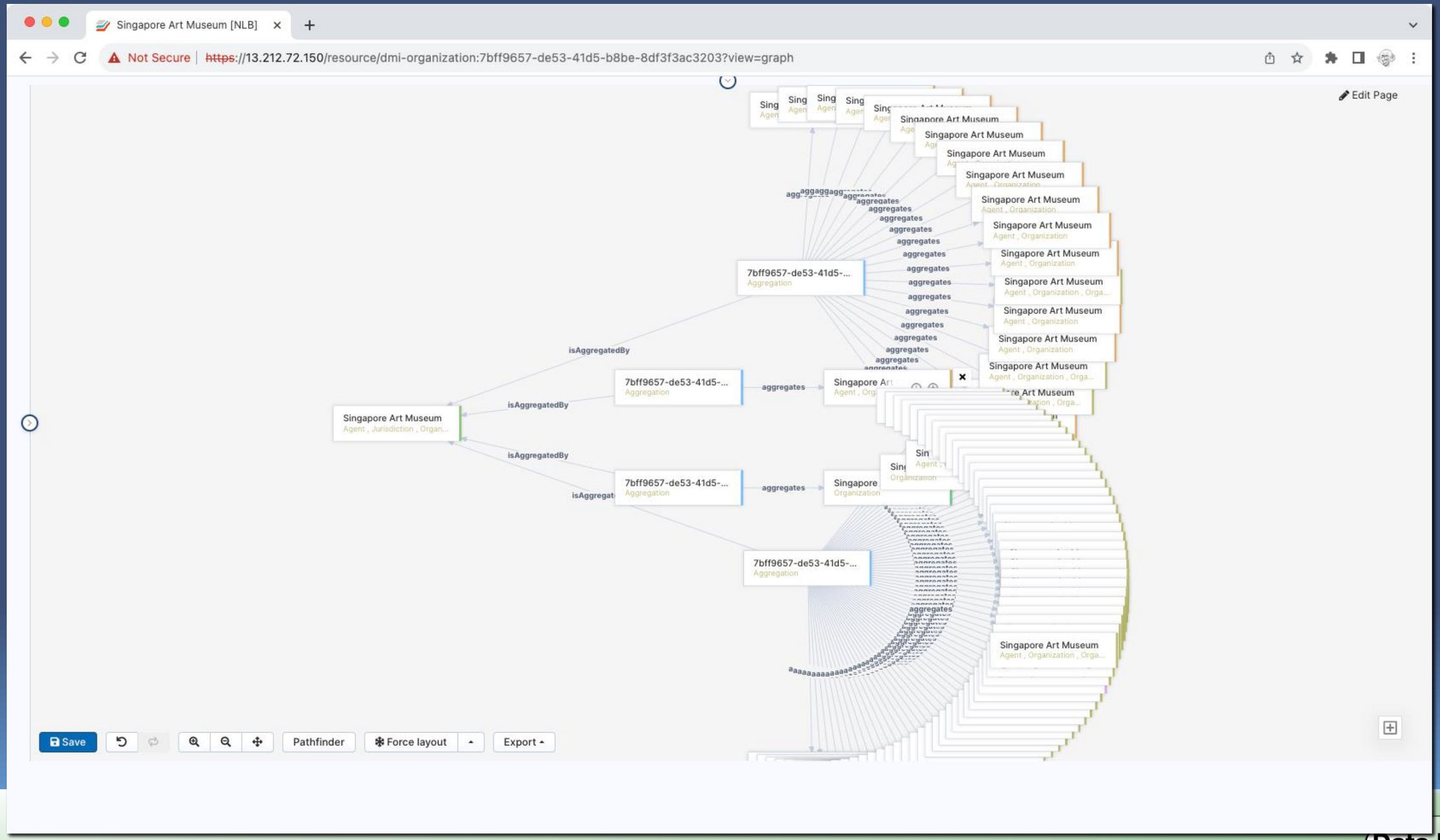




Data Liberate



Data Liberate

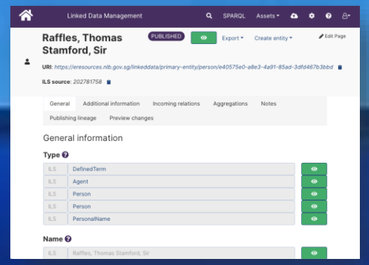


Data Liberate

Entity Matching

- Candidate matches based on schema:names
  - Lucene indexes matching
  - Levenshtein similarity refined
  - Same entity type
  - Entity type specific rules eg:
    - Work: name + creator / contributor / author / sameAs
    - Person: name + birthDate / deathDate / sameAs

The entity iceberg



Management

Primary

Discovery

Aggregation

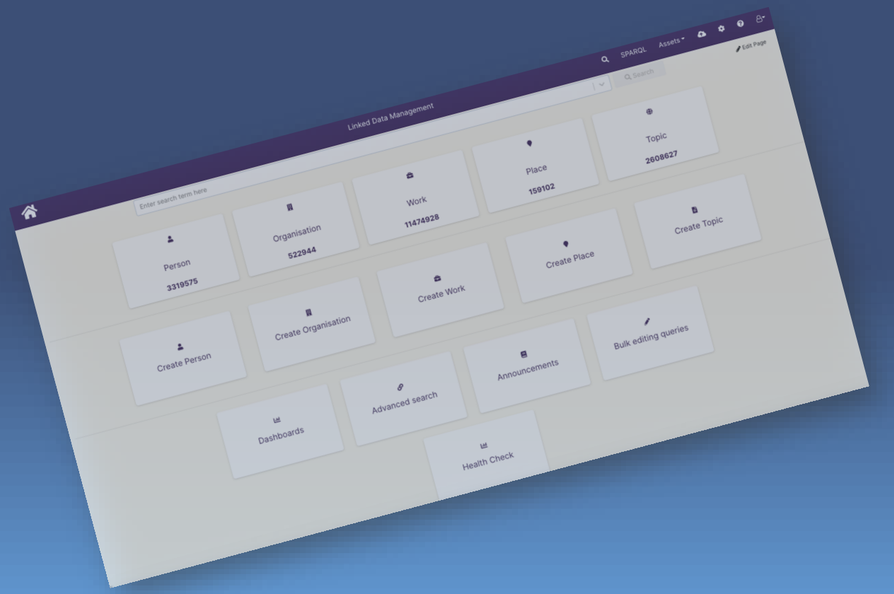
Source

Ingestion Pipelines

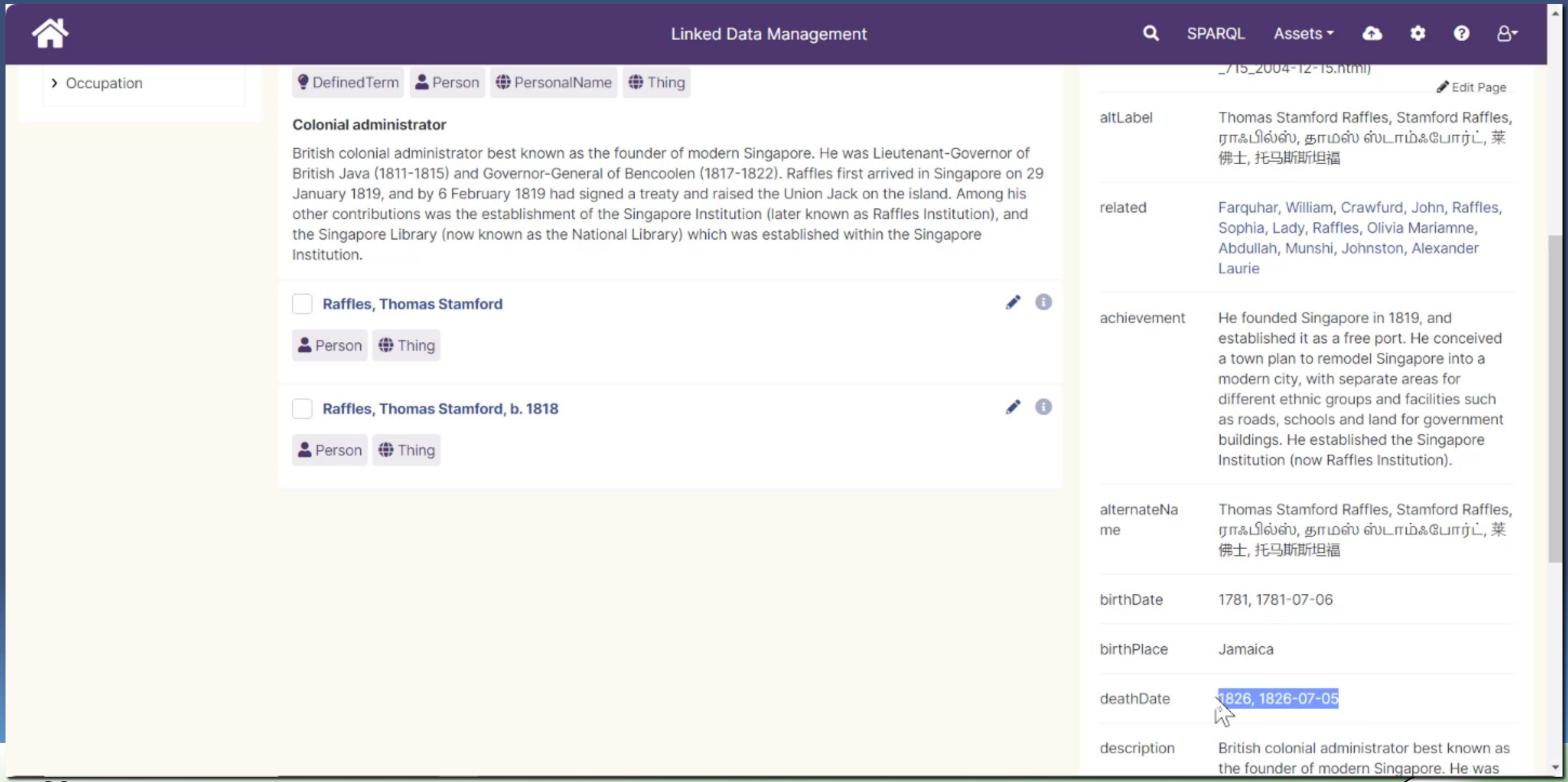
Data Liberate

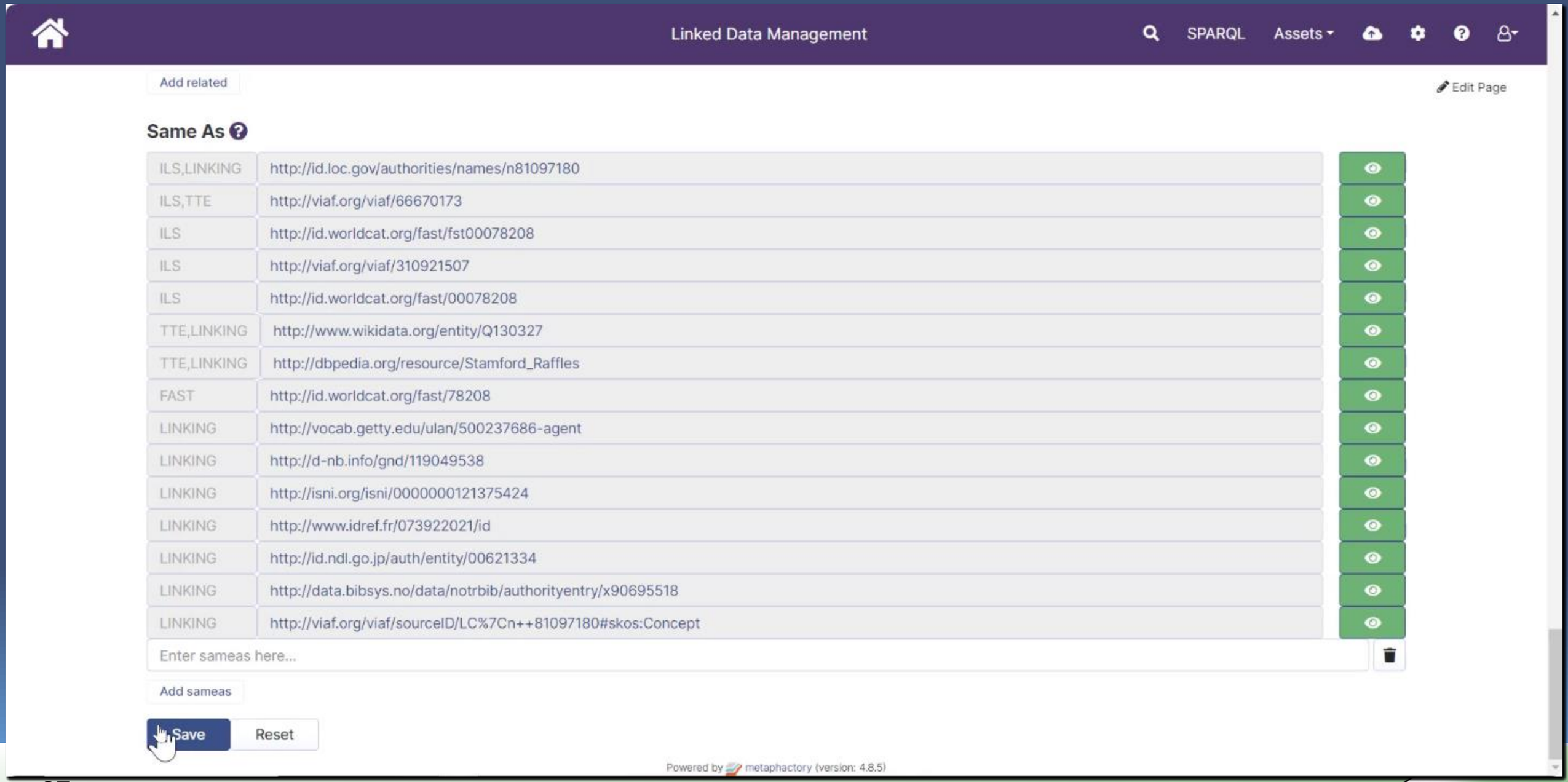
Data Management Interface

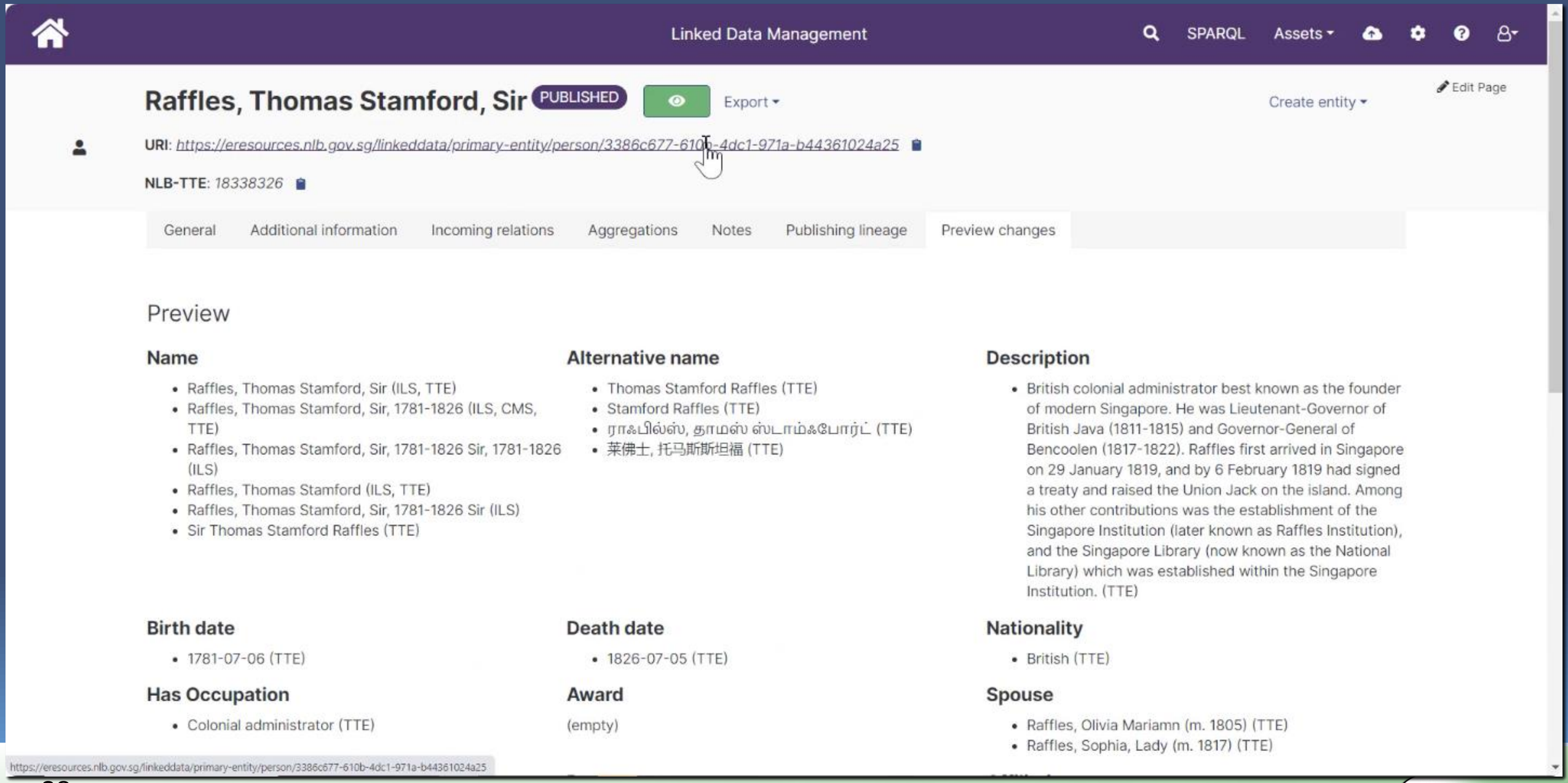
- Rapidly developed – easily modified
- Using mataphactory's semantic visual interface
- A collaboration environment for data curators
- Pre-publish entity management
- Search & discovery – emulates DI
- Dashboard view
- Entity management / Creation



Data Liberate


Linked Data Management
SPARQL
Assets
Occupation
DefinedTerm
Person
PersonalName
Thing
Colonial administrator
British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution.
Raffles, Thomas Stamford
Person
Thing
Raffles, Thomas Stamford, b. 1818
Person
Thing
_715_2004-12-15.html)
Edit Page
altLabel
Thomas Stamford Raffles, Stamford Raffles, ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட், 莱佛士, 托马斯斯坦福
related
Farquhar, William, Crawfurd, John, Raffles, Sophia, Lady, Raffles, Olivia Mariamne, Abdullah, Munshi, Johnston, Alexander Laurie
achievement
He founded Singapore in 1819, and established it as a free port. He conceived a town plan to remodel Singapore into a modern city, with separate areas for different ethnic groups and facilities such as roads, schools and land for government buildings. He established the Singapore Institution (now Raffles Institution).
alternateName
Thomas Stamford Raffles, Stamford Raffles, ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட், 莱佛士, 托马斯斯坦福
birthDate
1781, 1781-07-06
birthPlace
Jamaica
deathDate
1826, 1826-07-05
description
British colonial administrator best known as the founder of modern Singapore. He was


Linked Data Management
SPARQL
Assets
Add related
Edit Page
Same As
ILS,LINKING http://id.loc.gov/authorities/names/n81097180
ILS,TTE http://viaf.org/viaf/66670173
ILS http://id.worldcat.org/fast/fst00078208
ILS http://viaf.org/viaf/310921507
ILS http://id.worldcat.org/fast/00078208
TTE,LINKING http://www.wikidata.org/entity/Q130327
TTE,LINKING http://dbpedia.org/resource/Stamford_Raffles
FAST http://id.worldcat.org/fast/78208
LINKING http://vocab.getty.edu/ulan/500237686-agent
LINKING http://d-nb.info/gnd/119049538
LINKING http://isni.org/isni/0000000121375424
LINKING http://www.idref.fr/073922021/id
LINKING http://id.ndl.go.jp/auth/entity/00621334
LINKING http://data.bibsys.no/data/notrbib/authorityentry/x90695518
LINKING http://viaf.org/viaf/sourceID/LC%7Cn++81097180#skos:Concept
Enter sameas here...
Add sameas
Save
Reset
Powered by metaphactory (version: 4.8.5)


Linked Data Management
SPARQL
Assets
Raffles, Thomas Stamford, Sir
PUBLISHED
Export
Create entity
Edit Page
URI: https://eresources.nlb.gov.sg/linkeddata/primary-entity/person/3386c677-610b-4dc1-971a-b44361024a25
NLB-TTE: 18338326
General
Additional information
Incoming relations
Aggregations
Notes
Publishing lineage
Preview changes
Preview
Name
Raffles, Thomas Stamford, Sir (ILS, TTE)
Raffles, Thomas Stamford, Sir, 1781-1826 (ILS, CMS, TTE)
Raffles, Thomas Stamford, Sir, 1781-1826 Sir, 1781-1826 (ILS)
Raffles, Thomas Stamford (ILS, TTE)
Raffles, Thomas Stamford, Sir, 1781-1826 Sir (ILS)
Sir Thomas Stamford Raffles (TTE)
Alternative name
Thomas Stamford Raffles (TTE)
Stamford Raffles (TTE)
ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட் (TTE)
莱佛士, 托马斯斯坦福 (TTE)
Description
British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution. (TTE)
Birth date
1781-07-06 (TTE)
Death date
1826-07-05 (TTE)
Nationality
British (TTE)
Has Occupation
Colonial administrator (TTE)
Award
(empty)
Spouse
Raffles, Olivia Mariamn (m. 1805) (TTE)
Raffles, Sophia, Lady (m. 1817) (TTE)
https://eresources.nlb.gov.sg/linkeddata/primary-entity/person/3386c677-610b-4dc1-971a-b44361024a25

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

Raffles, Thomas Stamford, Sir

Person **Colonial administrator**

British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution.

Also known as: Thomas Stamford Raffles
Stamford Raffles
ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட்
莱佛士, 托马斯斯坦福

Birth place: Onboard the ship "Ann" off of the coast of Jamaica

Birth date: 1781-07-06

View Full Detail

External Links 4 results

Library of Congress - Linked Data Service

Works by Raffles, Thomas Stamford, Sir 56 results

Contributions by Raffles, Thomas Stamford, Sir 88 results

Works About Raffles, Thomas Stamford, Sir 268 results

https://13.212.72.150

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

Raffles, Thomas Stamford, Sir

Person **Colonial administrator**

British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution.

External Links 4 results

- Library of Congress - Linked Data Service
- VIAF - Virtual International Authority File
- VIAF - Virtual International Authority File
- Wikidata

Also known as: Thomas Stamford Raffles
Stamford Raffles
ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட்
莱佛士, 托马斯斯坦福

Birth place: Onboard the ship "Ann" off of the coast of Jamaica

Birth date: 1781-07-06

Death date: 1826-07-05

Spouse: Raffles, Olivia Mariamn (m. 1805)
Raffles, Sophia, Lady (m. 1817)

Parent: Raffles, Benjamin
Lyde, Ann

Related: Farquhar, William
Crawfurd, John
Raffles, Sophia, Lady
Raffles, Olivia Mariamne
Abdullah, Munshi
Johnston, Alexander Laurie

Data Liberate

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

Raffles, Thomas Stamford, Sir

Person **Colonial administrator**

British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution.

Also known as: Thomas Stamford Raffles
Stamford Raffles
ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட்
莱佛士, 托马斯斯坦福

Birth place: Onboard the ship "Ann" off of the coast of Jamaica

Birth date: 1781-07-06

View Full Detail

External Links 4 results

Library of Congress - Linked Data Service

Works by Raffles, Thomas Stamford, Sir 56 results

Contributions by Raffles, Thomas Stamford, Sir 88 results

Works About Raffles, Thomas Stamford, Sir 268 results

https://13.212.72.150

Data Liberate

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

Raffles, Thomas Stamford, Sir

Person **Colonial administrator**

British colonial administrator best known as the founder of modern Singapore. He was Lieutenant-Governor of British Java (1811-1815) and Governor-General of Bencoolen (1817-1822). Raffles first arrived in Singapore on 29 January 1819, and by 6 February 1819 had signed a treaty and raised the Union Jack on the island. Among his other contributions was the establishment of the Singapore Institution (later known as Raffles Institution), and the Singapore Library (now known as the National Library) which was established within the Singapore Institution.

Also known as: Thomas Stamford Raffles
Stamford Raffles
ராஃபில்ஸ், தாமஸ் ஸ்டாம்ஃபோர்ட்
莱佛士, 托马斯斯坦福

Birth place: Onboard the ship "Ann" off of the coast of Jamaica

Birth date: 1781-07-06

View Full Detail

External Links 4 results

Library of Congress - Linked Data Service

Works by Raffles, Thomas Stamford, Sir 56 results

Quick search

A history of Java

Raffles, Thomas Stamford, Sir

Work Book English

Data Liberate

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

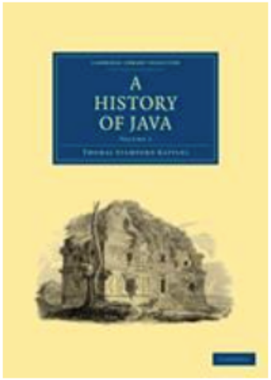
A history of Java

Work Book English

Author: Raffles, Thomas Stamford, Sir

Language: English

About: Malayan languages
a41b0bbf-40b8-49d7-93ae-fb8329189c0a
57aff1ec-3c49-4b78-983d-bf73a8a00496
Indonesia--Java.
3e98b3f2-b757-4132-bb39-7b8bfc03a2a8
Java (Indonesia)--Description and travel.



Version 1 result

Data Liberate

A Singapore Government Agency Website

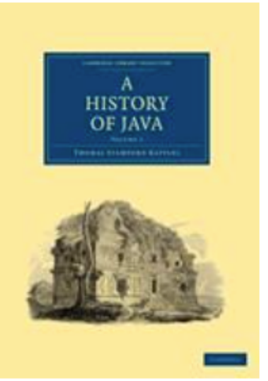
Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

A history of Java

Work Book English



Author: Raffles, Thomas Stamford, Sir

Language: English

About: Malayan languages
a41b0bbf-40b8-49d7-93ae-fb8329189c0a
57aff1ec-3c49-4b78-983d-bf73a8a00496
Indonesia--Java.
3e98b3f2-b757-4132-bb39-7b8bfc03a2a8
Java (Indonesia)--Description and travel.

Version 1 result

Quick search

A history of Java

Raffles, Thomas Stamford, Sir

Work Book English

ill., maps, music Reprint. Originally published: London : Printed for Black, Parbury, and Allen, 1817.

View In Library

Data Liberate

A Singapore Government Agency Website

Edit Page

NLB KNOWLEDGE DISCOVERY

Search for books, audios, videos, and more...

Indonesia--Java.

Subject

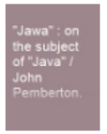
Works About Indonesia--Java. 1137 results

Quick search

"Jawa" : on the subject of "Java"

Pemberton, John

Work Book Indonesian



'T koninkryk Sunda met dat van lava by d'Inwoonderen voor een Eiland Gehouden, en door den Hr. I.B. Lavanha Als twee Eylanden Beschreven

Aa, Pieter van der

Work Map Dutch

This map of Sundra and Java appears in Pieter van der Aa's Cartes des itineraires et voyage moderne. Whilst incorrect in its view of the island, the map features a cartouche with Javanese natives encountering the Portuguese explorer Lopo vaz de Sampoya. The map is based on the work of Portuguese cartographer Joao de Barros, whose map of Java appeared in Quatra decada d'Asia in 1615. Barros own sources for this map are unknown, but it follows the Portuguese belief that Java comprised two islands, the smaller one to the west being known as Sundra. -- Dealer's description.

13 raja paling berpengaruh sepanjang sejarah kerajaan di Tanah Jawa

Achmad, Sri Wintala

Work Book Indonesian

Collective biography of kings and rulers of Java.

Data Liberate

A Singapore Government Agency Website

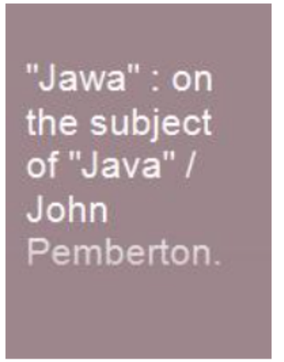
Edit Page

NLB Knowledge Discovery

Search for books, audios, videos, and more...

"Jawa" : on the subject of "Java"

Work Book Indonesian



Author: Pemberton, John

Language: Indonesian

About: Indonesia--Java.

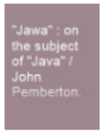
Version 1 result

Quick search

"Jawa" : on the subject of "Java"

Pemberton, John

Work Book Indonesian



ill. ports. Diterjemahkan dari: On the subject of "Java".

View In Library

1

Show 10 rows

Data Liberate

LDMS System Live – December 2022

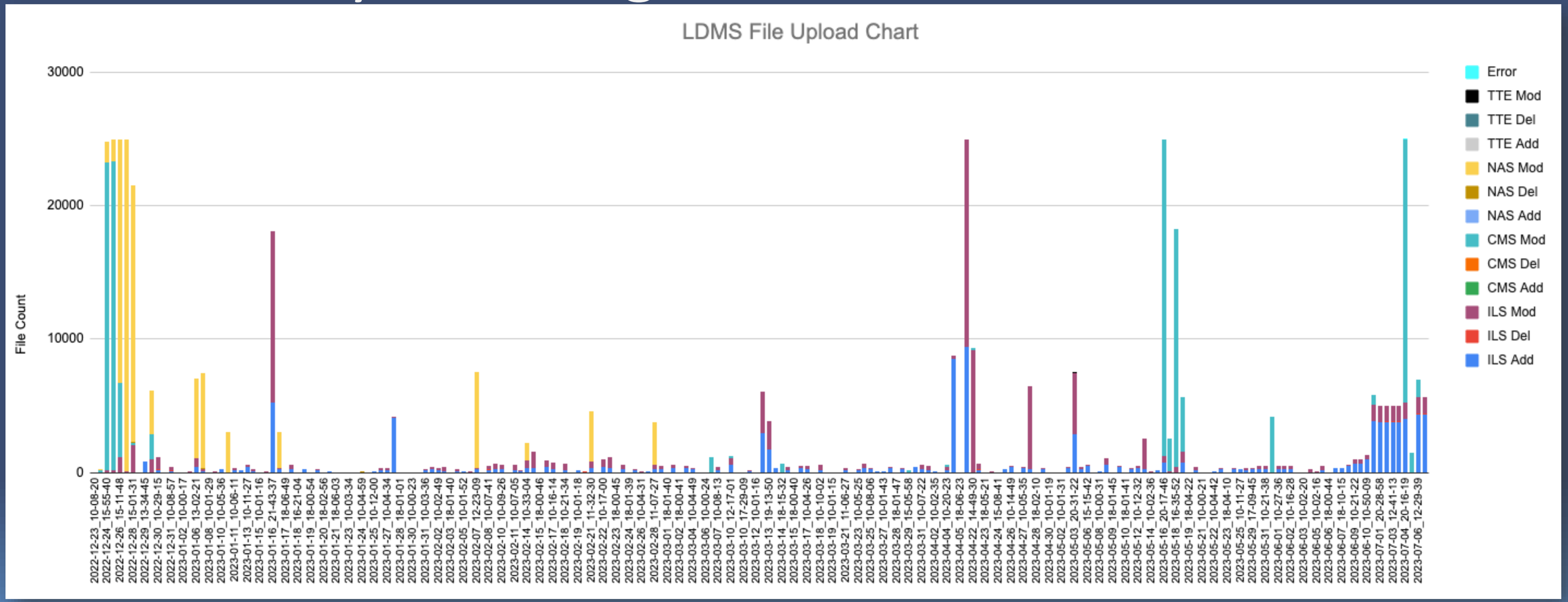
- Live – data curation team using DMI
  - Entity curation
  - Manually merging & splitting aggregations
- Live – data ingest pipelines
  - Daily / Weekly / Monthly
  - Delta update exports from source systems
- Live – Discovery Interface ready for deployment
- Live – Support
  - Issue resolution & updates

Discovery Interface

- Live and ready for deployment
  - Performance, load and user tested
- Designed & developed by NLB Data Team to demonstrate characteristics and benefits of a Linked Data service
- NLB services rolled-out by National Library & Public Library divisions
  - Recently identified a service for Linked Data integration
    - Currently undergoing fine tuning to meet service's UI requirements

Data Liberate

Daily Delta Ingestion & Reconciliation



Data Liberate

Some Challenges Addressed on the Way

- Inaccurate source data
  - Not apparent in source systems
  - Very obvious when aggregated in LDMS
    - Eg. Subject and Person URIs duplicated
  - Identified and analysed in DMI
  - Reported to source curators and fixed
  - Updated in next delta update
  - Data issues identified in LDMS – source data quality enhanced

Data Liberate

Some Challenges Addressed on the Way

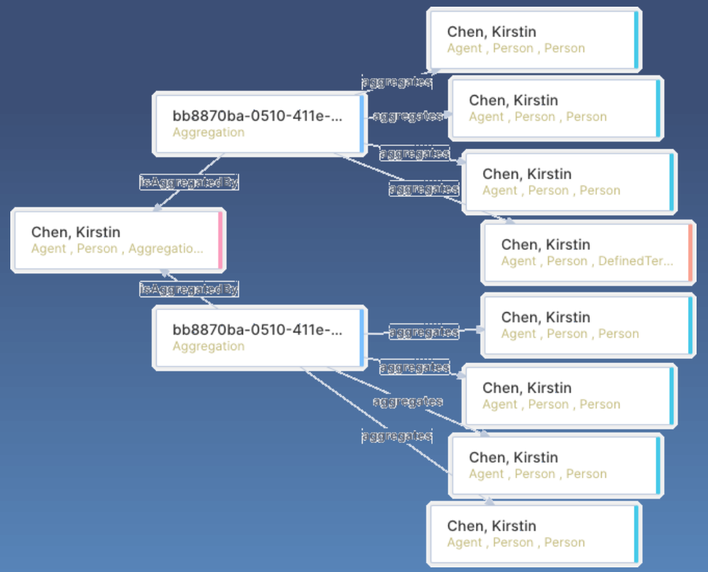
- Asian name matching
  - Often consisting of several short 3-4 letter names
  - Lucene matches not of much help
  - Introduction of Levenshtein similarity matching helped
  - Tuning is challenging – trading off European & Asian names

From Ambition to Go Live – A Summary

- Ambition based on previous years of trials, experimentation, and understanding of Linked Data potential
- Ambition to deliver a production LDMS to provide Linked Data curation and management capable of delivering a public discovery service
- With experienced commercial partners & tools as part of a globally distributed team
  - Benefiting from Open Source community developments where appropriate
- Unique and challenging requirements
  - Distributed sources of truth in disparate data formats
  - Auto updating, consolidated entity view – plus individual entity management
  - Web friendly – providing structured data for search engines
- Live and operational actively delivering benefits from December 2022

From Ambition to Go Live

The National Library Board of Singapore's Journey to an operational Linked Data Management and Discovery System



Richard Wallis
Evangelist and Founder
Data Liberate
richard.wallis@dataliberate.com
@dataliberate

CC BY

15th Semantic Web in Libraries Conference
SWIB23 - 12th September 2023 - Berlin

Data Liberate