Data for History Vision

George Bruseker (FORTH-ICS)

Second Data for History Meeting
LARHRA Lyon
25/5/2018

Data for History Vision

An international group of (groups of) historians, producing and wishing to share quality analytic data on which to base historical analyses. Data for History adopts semantic data representation principles and toolsets in order to make open, sourced and reusable data for use in historical research. As a standard for conceptual modelling, the group adopts the CIDOC CRM ontology, while intending to extend this model with concepts and relations adapted to the scope of historical research and information representation. Wherever possible, the group aims to adopt, enrich and/or create standardized reference resources in the sense of thesauri for terminology alignment and authority records for instance alignment.

The development of a longterm, sustainable semantic data strategy for history falls within the remit and interest of the Data for History group. It aims to support best practice in adopting semantic technology and creating reusable analytic resources. As such it is interested in conceptualizing and supporting the complete semantic data life cycle without prejudice to particular tools or software brands, but focussing instead on understanding the complete workflow necessary to design, implement and sustain semantic data projects and identifying the necessary kinds of components to achieve those ends. In line with these overall goals, the group adopts and attempts to follow its own findings and recommendations in conducting its work.

Data for History aims to be a catalyzer for historical research based on and generating semantic data, enabling researchers and research teams to join forces and benefit from intellectual collaboration, putting semantic models to work towards the common end of the better understanding and appreciation of history.
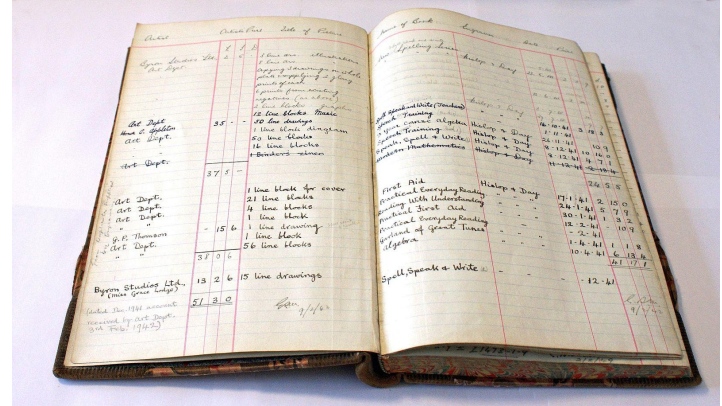
SEMANTIC DATA WORKFLOW MANAGEMENT

REGISTER

Edinburgh City of Print CC BY 2.0
https://commons.wikimedia.org/wiki/File:Ledger_detailing_external_work_commissioned_at_Holmes_McDougall_(4268190563).jpg

WHAT:

A database listing the resources available to the researcher/institution/consortium and their present state, which is updated as semantic project results are generated.

WHY:

We need to know what the present state of affairs is, what's on offer, what data is there, what state it is in, what partners are available, what resources can be accessed and we need to track what has been done, based on what.

Tools Required	● Register
Tool Examples	1. Parthenos Entities and Architecture 2. Data for History Drupal

PROJECT DESIGN

WHAT:

Working on the basis of the known present state of affairs, project design in terms of existing dataset selection (if any), software toolkit selection, partner selection, research question design

WHY:

Proper Research Methodology informed by knowledge of latest state of affairs in resource base.

Tools Required	- Pen - Paper - Education - Project Management Software
Tool Examples	1. Bic 2. Moleskin 3. Not taking sides 4. RedMine

Modelling (Selection/Extension)

INPIVIC CC BY 2.0 https://www.flickr.com/photos/inpivic/5205915907

WHAT:

Selection of an adequate conceptual model to the needs of the particular project. The model may be fully adequate already or require extensions to an existing model in order to be realized. Can also document best practices for model.

WHY:

Ensure adequate semantic representation for the field of research in order to answer questions, establish conditions of interoperability

| Tools Required	● ontology development management environment ● application profile designer
Tool Examples	1. OntoMe 2. Arches Designer Tool
Output Types	● Ontologies (RDF/OWL) ● Application Profiles

Modelling Implementation/Mapping

WHAT:

For data production, a data curation environment should be chosen which can implement the model.

For pre-existing data, a data mapping environment should be chosen which can transform data to the model.

WHY:

Ensure new project data conformant to model while reusing and valorizing existing datasets.

https://pixabay.com/en/network-head-face-grid-human-516333/
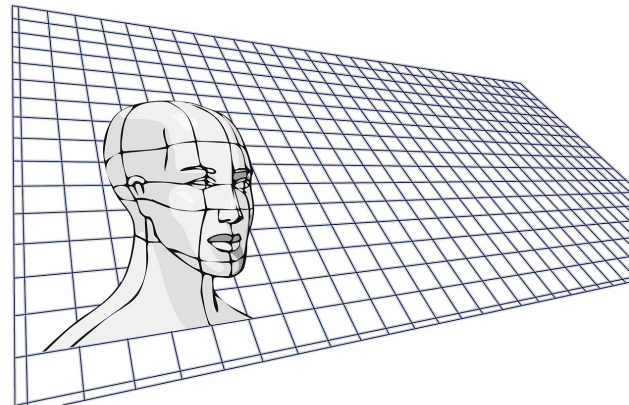
Tools Required	● Data Curation Environment ● Data Mapping Environment
Tool Examples	1. ResearchSpace, Wisski, Arches, Qoqnus 2. 3M, Karma
Output Types	● Model Compatible RDF

Data Aggregation Processes

Geoff Gallice CC BY 2.0
https://commons.wikimedia.org/wiki/File:Flickr_-_ggallice_-_Caterpillar_aggregation.jpg

WHAT:

Once all desired sources are expressed in a conformant format (at schema and data level), semantic data should be ported to a common knowledge base for common use and exploration.

WHY:

Ensure reproducible workflow of aggregation to common environment and provenance of semantic data.

Tools Required	● Aggregation Management Environment
Tool Examples	1. DNet

Data Manipulation & Visualization

Eric Delcroix CC BY-NC-SA 2.0
https://www.flickr.com/photos/eric-delcroix/24519077825

WHAT:

Use and interpret aggregate data to generate new knowledge

WHY:

Creating new knowledge is the final cause of all these processes.

Tools Required	● Knowledge graph search and visualization platform
Tool Examples	1. ResearchSpace

More than a day's work