

The DAE Platform

Document Mining, Interpretation and Semantics Through the Prism of Performance Evaluation

Bart Lamiroy

LORIA – Université de Lorraine – France

In collaboration with
Lehigh University



The DAE Platform



Loria



LEHIGH
UNIVERSITY

*Bart Lamiroy, November 2011
Slide 1*

Speaker Background

- **Assistant Professor @ INPL – France**

<http://www.inpl-nancy.fr>

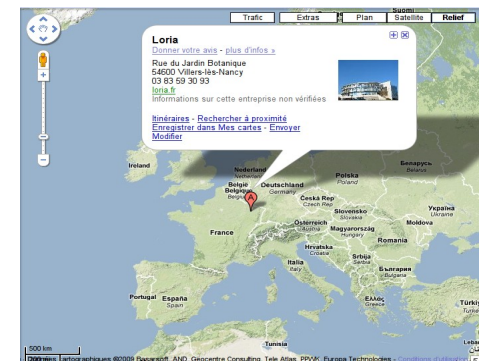
- **Research @ LORIA – INRIA**

<http://www.loria.fr>

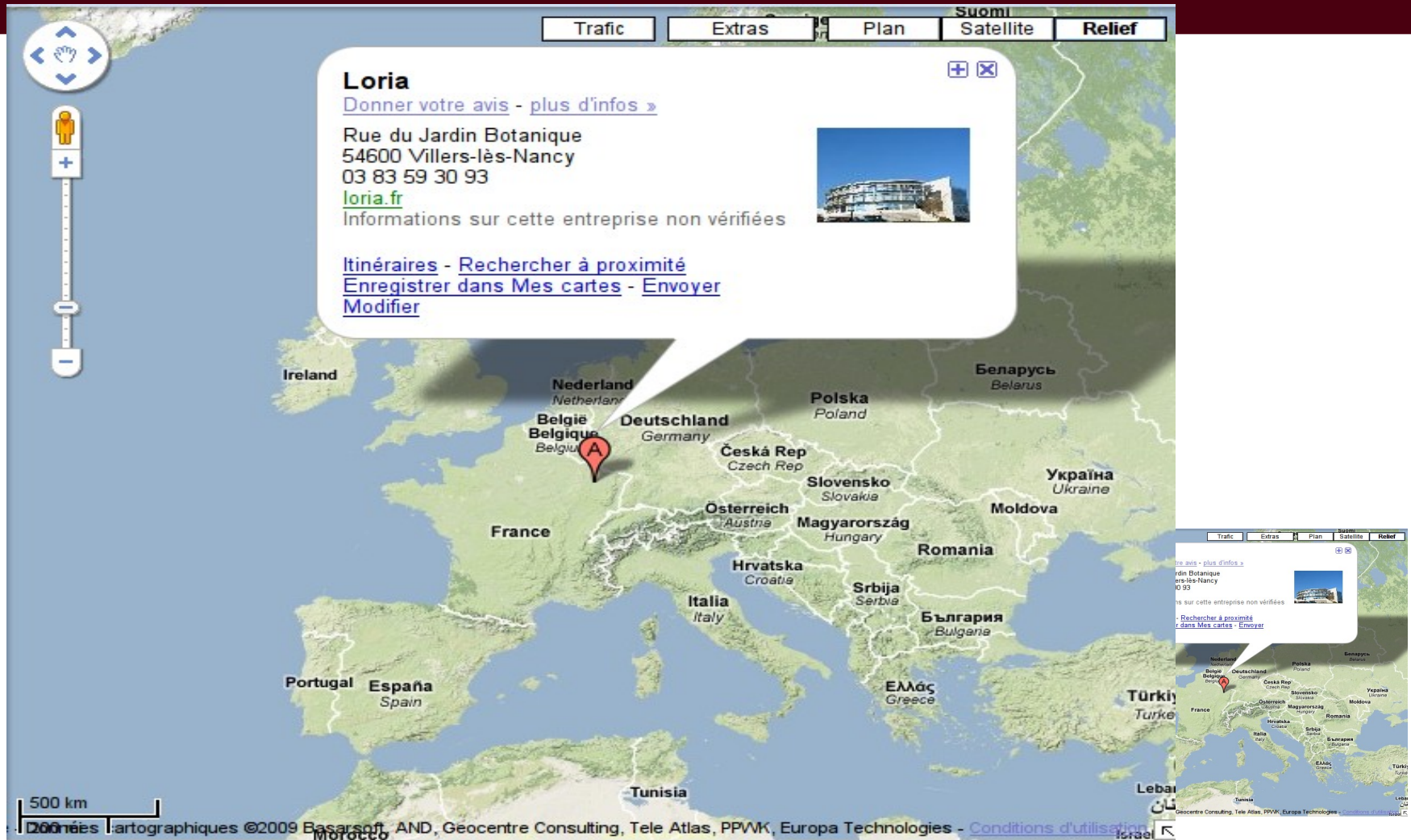
- **Visiting Scientist @ Lehigh (2010-2011)**

- **Contact Information:**

Bart.Lamiroy@loria.fr

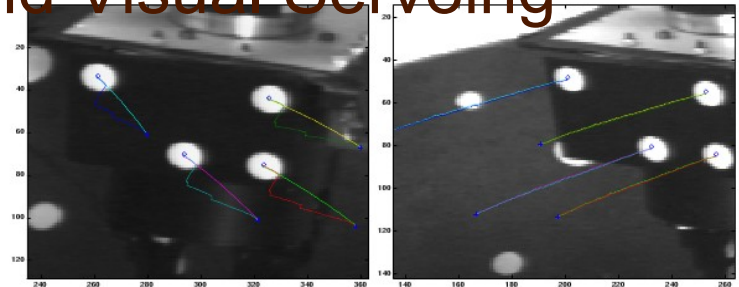


Speaker Background

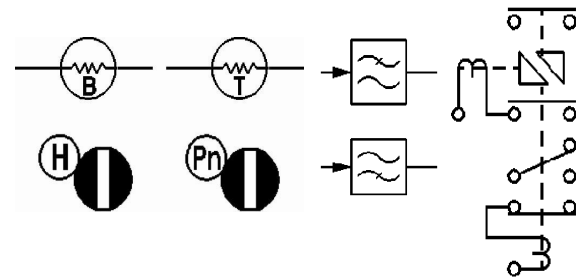


Scientific Background

- **1998:** Image Indexing (Ph.D.)
- **1998-2000:** Stereo Tracking and Visual Servoing



- **2000-2009:** Graphical Document Analysis & “*Image Semantics*”



- **2010-2011:** The DAE project (<http://dae.cse.lehigh.edu>)



The DAE Platform: Document Mining, Interpretation and Semantics Through the Prism of Performance Evaluation



Acknowledgements

Bryan Hodgson

Jeanne Steinberg

Xingjian Zhang

Yingjie Li

Yang Yu

Daniel Lopresti

Henry Baird

Weidong Chen

Dezhao Song

Qihan Long

Tao Sun

Sameer Antani

Mike Kot

Jeff Heflin

Hank Korth

Brian Davison

Sai Lu Mon Aung

Mike Caffrey

Austin Borden



Outline of Talk

- ◉ **General Context**

Challenges of reproducible and traceable experimental research

- ◉ **The DAE Platform**

A way of addressing the challenges, technical architecture and design choices

- ◉ **First Experiments**

Reports of actual uses of the platform

- ◉ **Looking Beyond and New Challenges**



Outline of Talk

- ◉ **General Context**

Challenges of reproducible and traceable experimental research

- ◉ **The DAE Platform**

A way of addressing the challenges, technical architecture and design choices

- ◉ **First Experiments**

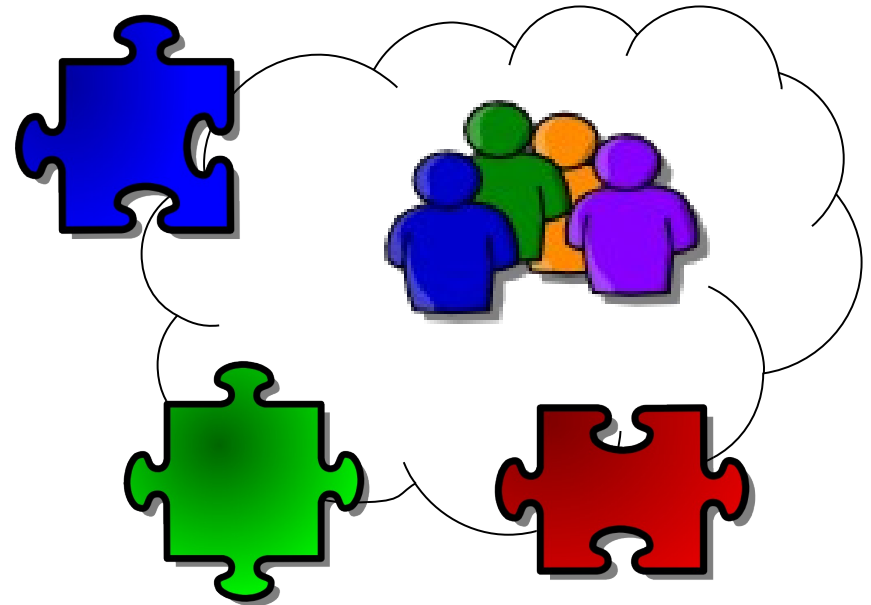
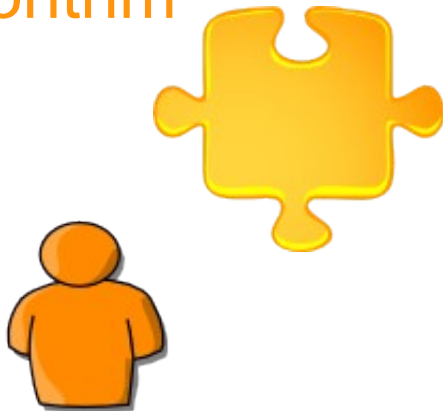
Reports of actual uses of the platform

- ◉ **Looking Beyond and New Challenges**



General Context

New idea – Claim –
Algorithm

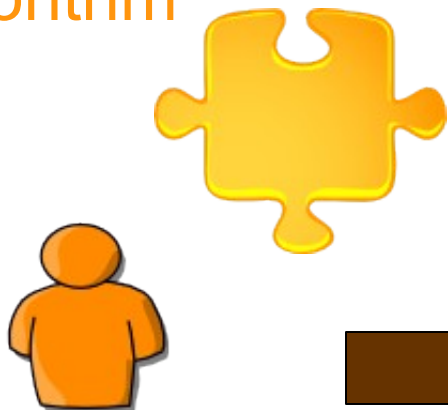


Existing algorithms – State-
of-the-Art – Other claims

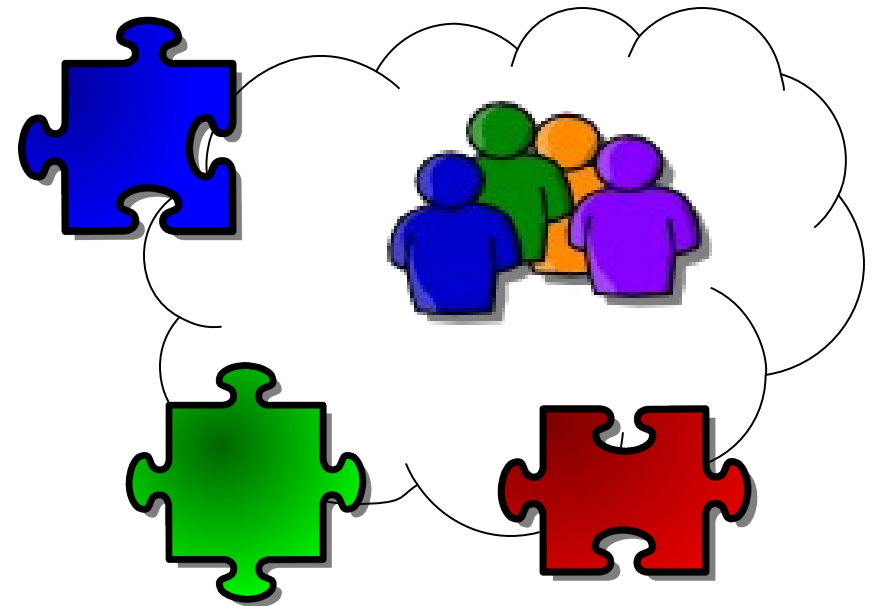


Reproducible and Traceable Experimental Research

New idea – Claim –
Algorithm



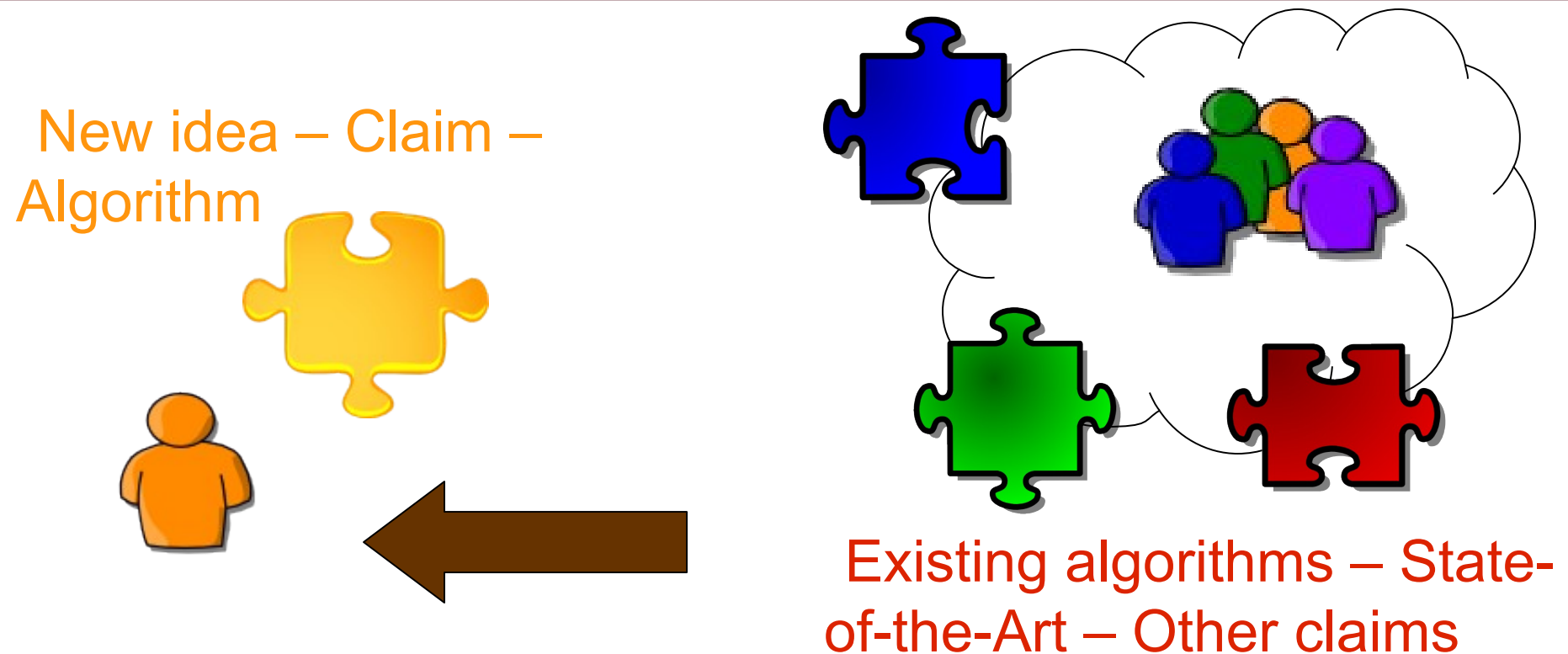
How do I perform?
How do I compare to the
State-of-the-Art?



Existing algorithms – State-
of-the-Art – Other claims



Reproducible and Traceable Experimental Research



What is the legitimacy of this claim?
Can it be verified/reproduced?



A Generic Scenario

Jane, a young starting researcher

Goal: solve a specific knowledge extraction problem (examples follow)

Requirements:

- Find appropriate experimental data
- Prove genericity of approach (or establish boundaries)
- Compare with state of the art
- Report results



Example 1

The screenshot displays the Cervigram CBIR System interface. The main window shows a query image of a cervix with a blue outline highlighting a lesion. The interface includes a menu bar (File, Help), a toolbar with various icons, and a search bar. Below the query image, there is a 'Region' tab with a list of features: AW, OS, All, Blood, Polyps, Cyst, SCJ, Mucus, SM, and a color selection area. A 'Start Retrieving' button is located at the bottom left. The right side of the interface shows 'Retrieved Results (2148 Regions)' with six small thumbnail images of cervixes, each with a blue outline highlighting a lesion. The thumbnails are labeled with their respective IDs: 1: 39803.89, 2: 43897.2046, 3: 42374.706, 4: 41645.430, 5: 44109.842, and 6: 43330.1739. A 'Page 1 of 358' indicator is visible at the bottom right of the results section.

- Lesion vs. healthy tissue
- Pathology classification
- Localization



Example 2

- Author – Title – Abstract localization
- Language Identification
- Topic Identification
- Reference extraction

Graefes Arch Clin Exp Ophthalmol
(2002) 240:101–105
DOI 10.1007/s00417-001-0389-z

LABORATORY INVESTIGATION

Toshiya Sakurai
Eiichi Yukawa
Yoshiaki Hara
Norio Miyata
Manabu Mochizuki

Indocyanine green angiography in patients with human T cell-lymphotropic virus type 1 uveitis

Received: 26 March 2001
Revised: 12 September 2001
Accepted: 12 September 2001
Published online: 29 January 2002
© Springer-Verlag 2002

This study was presented in part at the 4th International Symposium on ICG Angiography, Baden-Baden, Germany, September 1999

This study was approved by the ethics committee of Tane Memorial Eye Hospital

T. Sakurai (✉)
Tane Memorial Eye Hospital,
1-1-39 Sakaigawa, Nishi-ku, Osaka,
550-0024, Japan
e-mail: sakurai@wonder.ocn.ne.jp
Tel.: +81-6-65815800
Fax: +81-6-65815063

E. Yukawa · Y. Hara
Department of Ophthalmology,
Nara Medical University, Nara, Japan
N. Miyata
Miyata Eye Hospital, Miyakonojo,
Miyazaki, Japan
M. Mochizuki
Tokyo Medical and Dental University,
Tokyo, Japan

Abstract *Purpose:* To determine the indocyanine green (ICG) angiographic features and to evaluate the choroidal involvement of human T-cell lymphotropic virus type 1 (HTLV-1)-associated uveitis. *Methods:* We performed ICG angiography using scanning laser ophthalmoscopy in 54 eyes of 27 patients (8 men and 19 women) diag-

nosed with HTLV-1 uveitis. The patient's mean age was 51.5 years with a range of 24–65 years. *Results:* The early phase of ICG angiography revealed ICG leakage from the choroidal vessels in the posterior pole, hyperfluorescent spots that which were not detected with fluorescein angiography, and small hypofluorescent lesions in the macula which most likely corresponded to microcirculatory disturbances in the choriocapillaris. *Conclusions:* We suggest that the ICG angiographic findings reflect choroidal lesions such as infiltration with leukocytes and edema. ICG angiography may provide useful information on chorioidopathy in HTLV-1 uveitis.

Introduction

Human T cell-lymphotropic virus type 1 (HTLV-1) is a human retrovirus that is highly endemic in some regions of the world. HTLV-1 carriers have been detected in the Caribbean islands, parts of central Africa, and the south of Kyushu island, Japan. HTLV-1 is not easily transmissible under normal conditions, because cell-to-cell contact is generally required. The virus is transmitted by three major routes: vertical transmission from mother to child via breast-feeding (HTLV-1 antigen-positive lymphocytes in milk), and horizontal transmission from person to person via sexual contact (semen) or blood transfusion [16]. The virus is known to cause adult T-cell leukemia (ATL) and HTLV-1-associated myelopathy (HAM). Recent studies have suggested that HTLV-1 is also related to many other diseases [15].

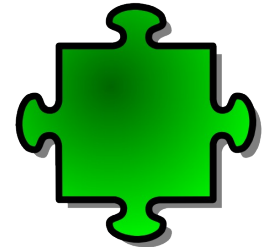
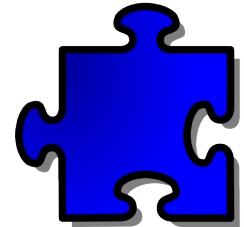
In 1989 [10], Oba et al. reported a relationship between HTLV-1 and uveitis. Mochizuki et al. performed epidemiologic, clinical, and virologic studies to investigate the correlations between HTLV-1 and uveitis and defined HTLV-1-associated uveitis. The diagnosis of HTLV-1 uveitis is based on seropositivity to HTLV-1 and the exclusion of other uveitic entities with defined etiology [1, 5, 6, 7, 8, 9, 11, 14, 18, 20].

The diagnostic criteria are idiopathic uveitis, HTLV-1 seropositivity, clinical features such as membranous vitreous opacities, and the presence of HTLV-1 infected cells in the aqueous humor. Fluorescein angiography provides useful information on the retinal vessels, but the choroidal vessels and choroidal circulation cannot be well visualized by this means. Recently, indocyanine green (ICG) angiography has been used to observe choroidal lesions [2, 3, 4, 19]. ICG dye limits leakage from



Current Consensus

- **Peer Reviewed Papers**
 - Re-implementation issues
 - Claim verification
- **Code Repositories**
 - Technology dependency
 - IP and Copyright
- **Reference Data Collections**
 - Maintenance cost
 - Non-evolution over time – static
 - Context restriction
- **Contests**
 - Snapshots



Problems with Consensus

- **Goal**

- Develop robust algorithms that approach human levels of performance for specific tasks of interest.

- **Observations**

- We want general algorithms, but often they are tested on small, overused datasets removed from real world.
- Published experimental results may be biased by algorithm developer's intimate knowledge of the data.
- Current practices fail to provide evidence of generality.
- What does “human levels of performance” mean? Even experts can disagree on all but the most trivial of cases.



Problems with Consensus

- **Goal**

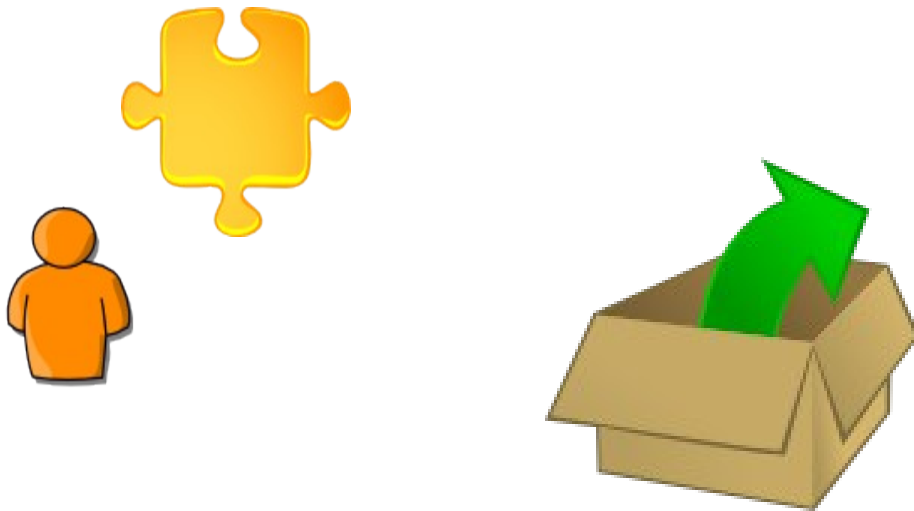
- Invent new methods that improve on known techniques.

- **Observations**

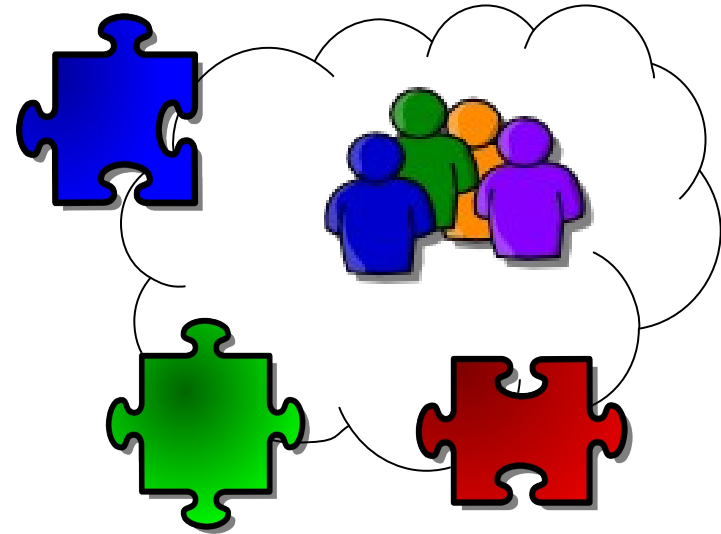
- How do we know when we have succeeded?
 - Need to compare against previously published results creates reliance on standard datasets (a bad idea).
 - Difficulties in re-implementing a published algorithm (incomplete descriptions, inherent conflict of interest).
 - Competitions can be useful, but are snapshots.
 - Many papers do not bother to make reproducible comparisons.



Needed Solution



- ▶ Find appropriate testing data
- ▶ Find reference validation
- ▶ Certify experimental results
- ▶ Reproduce published results



- ▶ Provide reference data
- ▶ Provide access to algorithms



Requirements for Solution

Formalizing Experiments and their Environment

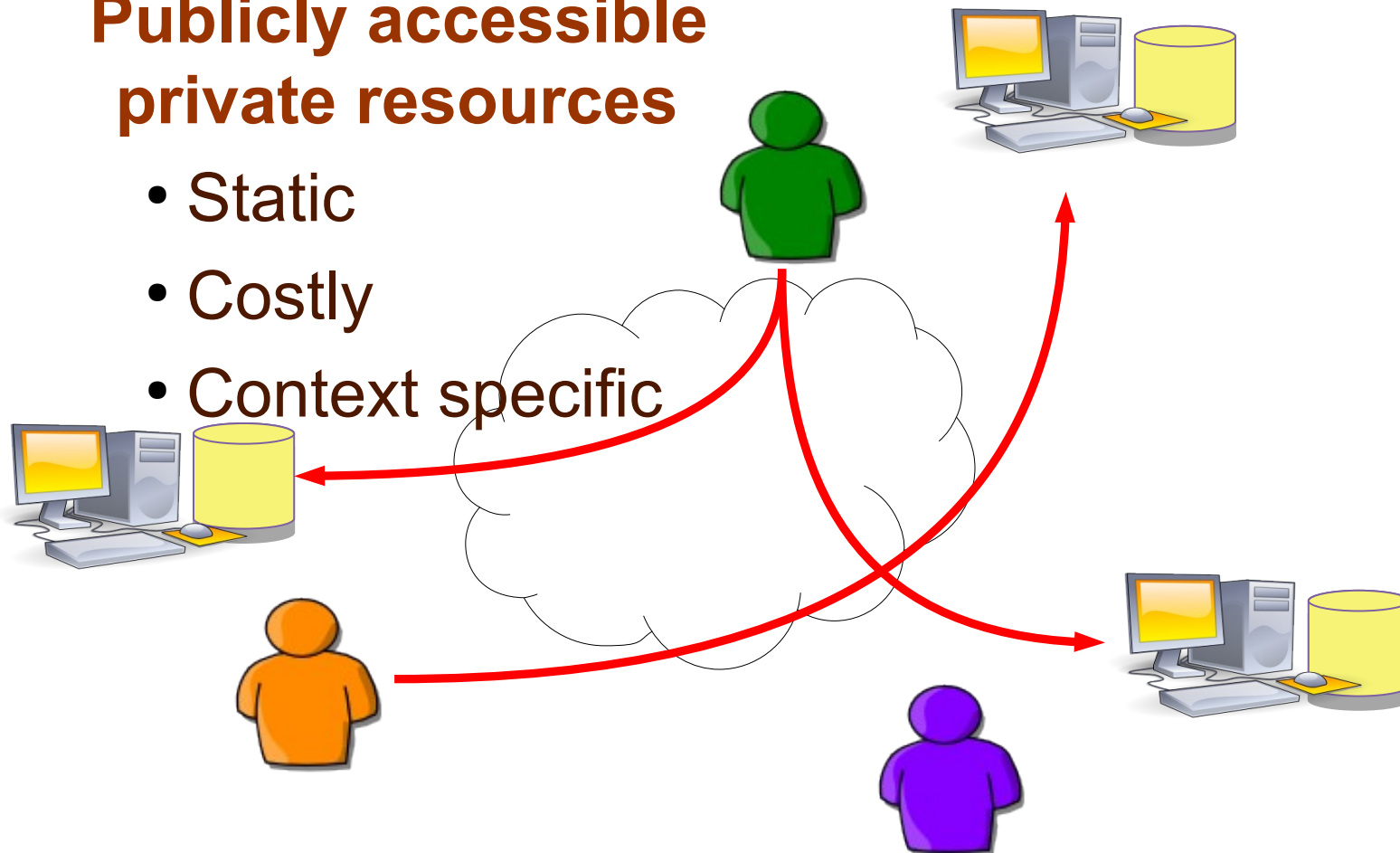
- ◉ Experimental DATA
 - Describe it
 - Make it available
 - Question – dispute – challenge – extend
- ◉ Experiments
 - Describe them
 - Make them available
 - Reproduce them
 - Question – adapt – modify – extend



Before

Publicly accessible private resources

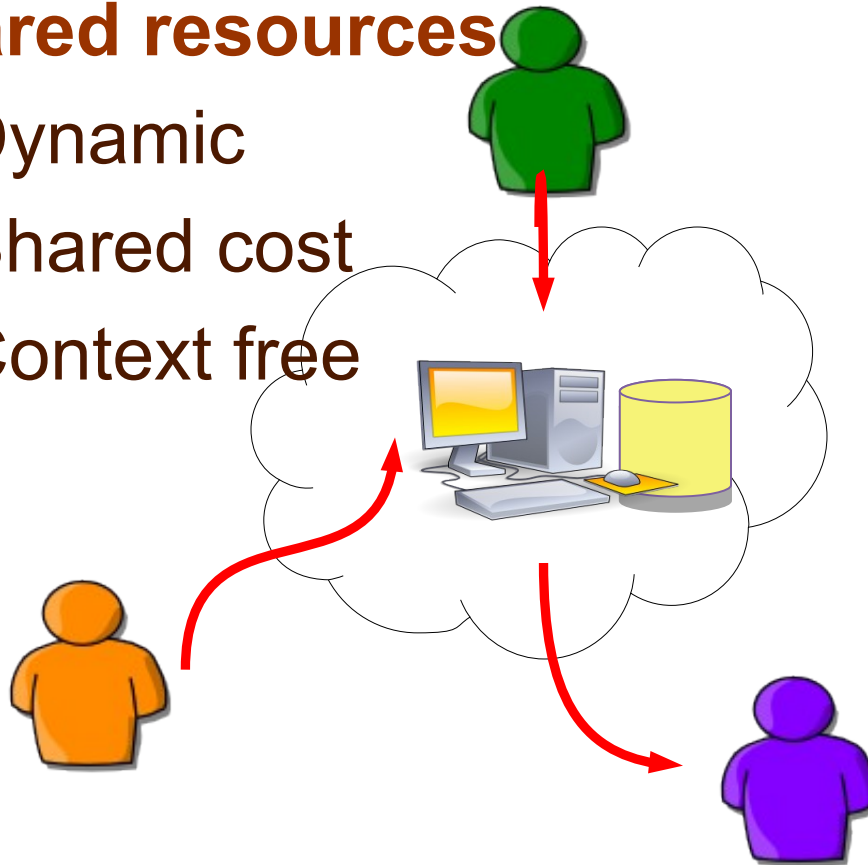
- Static
- Costly
- Context specific



DAE

Publicly accessible shared resources

- Dynamic
- Shared cost
- Context free



Outline of Talk

- **General Context**

Challenges of reproducible and traceable experimental research

- **The DAE Platform**

A way of addressing the challenges, technical architecture and design choices

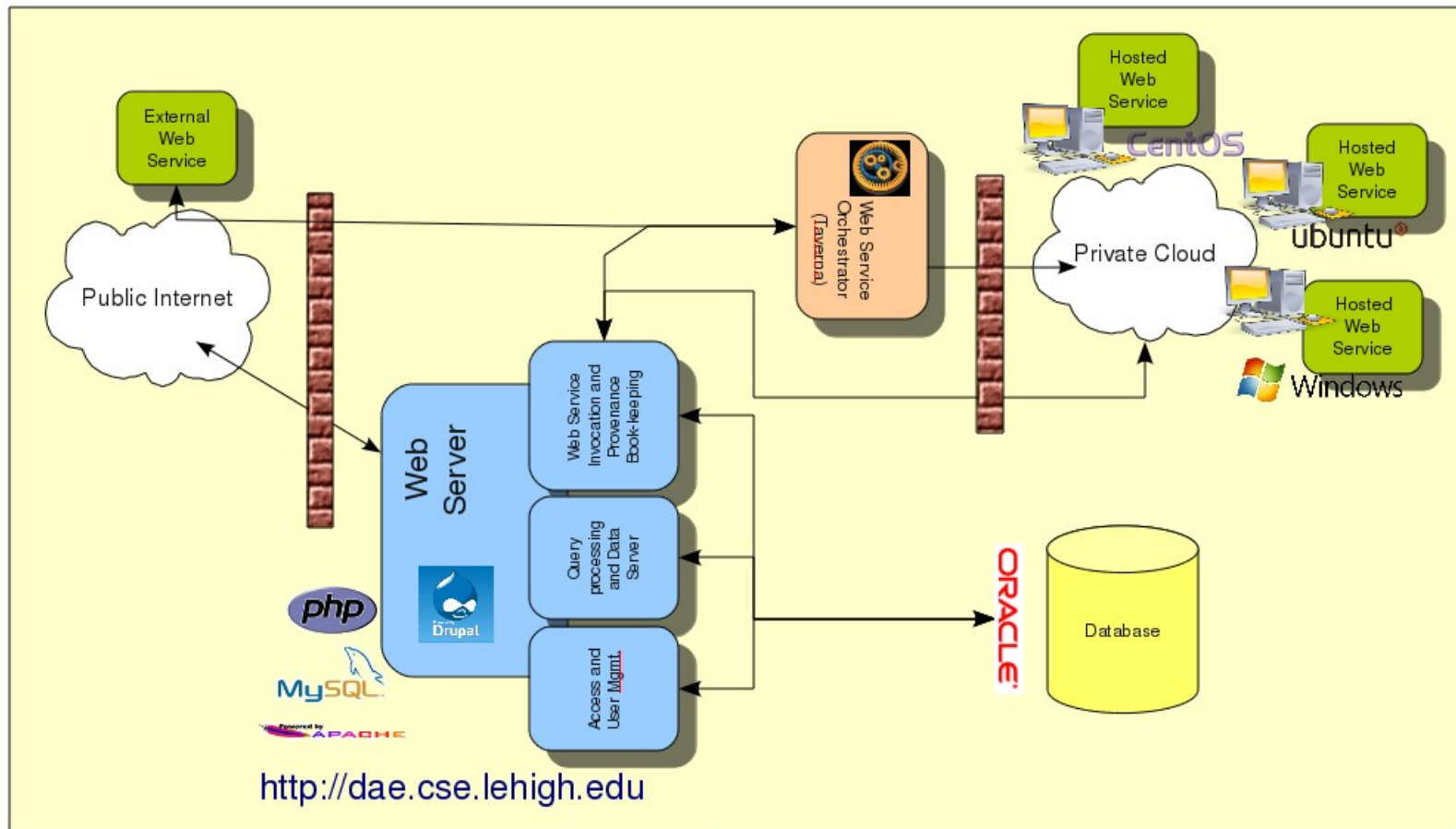
- **First Experiments**

Reports of actual uses of the platform

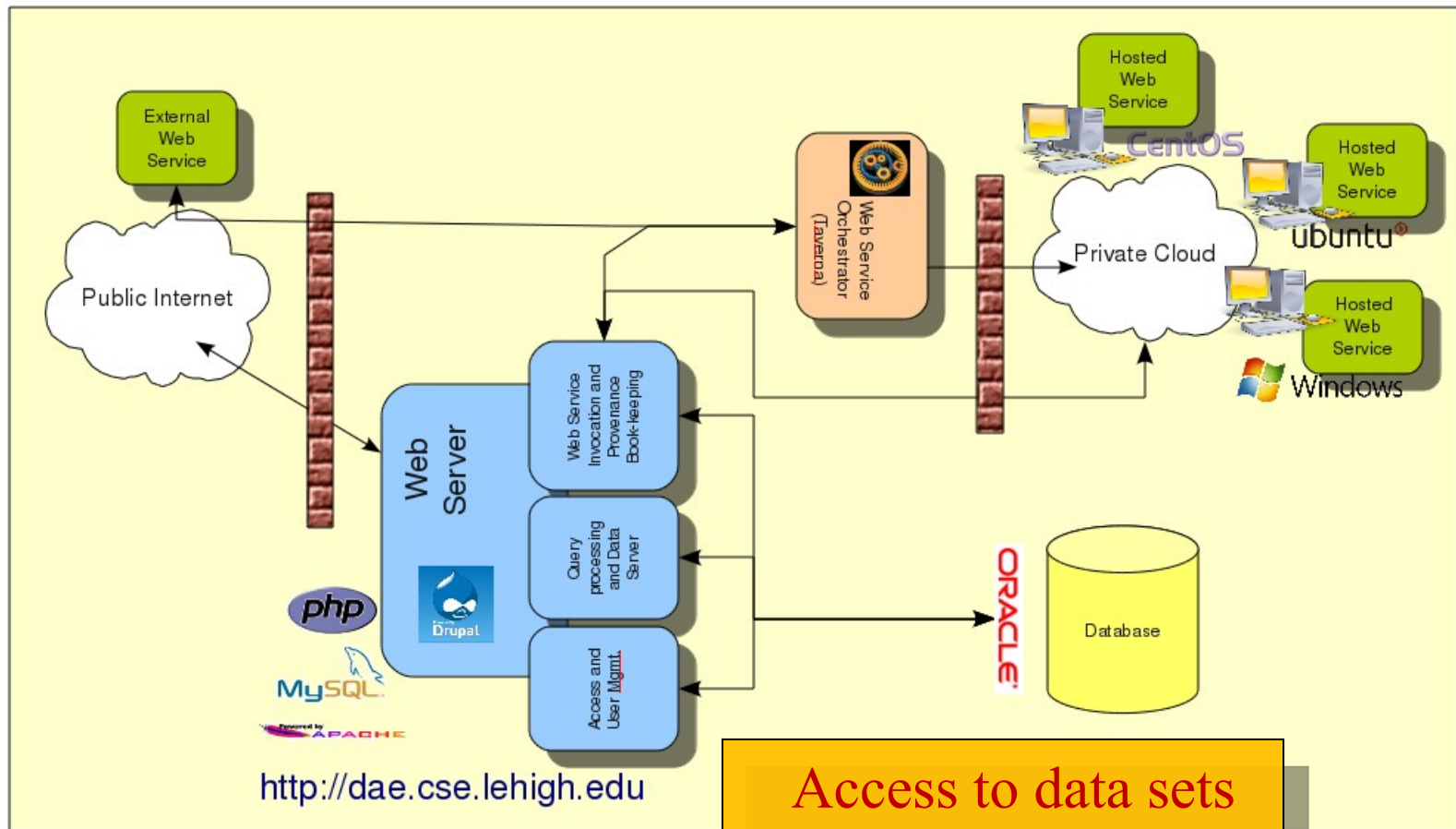
- **Looking Beyond and New Challenges**



DAE Platform - dae.cse.lehigh.edu



DAE Platform - dae.cse.lehigh.edu



Access to data sets and associated interpretations

Browsing Data Sets

The screenshot displays the DAE platform interface. At the top, a navigation bar includes 'Home', 'Goals', 'People', and 'Wiki'. The main header reads 'Document Analysis and Exploitation'. A search bar on the left prompts 'Search this site:' with a 'Search' button. The central content area shows a data set titled 'Minnesota Challenged Ballots' submitted by 'John Appleseed'. A left sidebar lists a hierarchical structure of folders: 'back pages', 'front pages', 'diverse front pages', 'Waseca front pages', 'waseca_Waseca_...00.tif' (highlighted), 'waseca_Waseca_...00.tif', 'Watsonwan front pages', 'Carver front pages', 'Clearwater front pages', 'Douglas front pages', 'Goodhue front pages', 'Kandiyohi front pages', and 'Koochiching front pages'. A red arrow points from the 'Hierarchical structuring within data sets' callout to the 'waseca_Waseca_...00.tif' file. To the right of the file list, a preview of a document page is shown. Further right, technical details for the selected file are listed: 'Name waseca_Waseca_3A_challengedballot1-000.tif', 'Kind image/tiff', 'Size 277.15 KB', 'Width 2480 pixels', 'Height 4112 pixels', 'Color Model Unknown', and 'Color Depth bits' (with sub-items 'VDPI' and 'HDPI'). A red circle highlights the 'Browsing features' callout near these details. Below the file list, a comment section shows a five-star rating by 'admin' with the text 'This data set is very big. Very, very big.' and a 'Was this comment helpful?' prompt with 'Yes' and 'No' options.

Author identification

Hierarchical structuring within data sets

Browsing features



Document Analysis and Exploitation

Document part annotation

Document tagging

Rating & commenting

Newsletters and discussion groups



Home Goals People Wiki

Search this site:

Search

admin

- Browse Data
- ▷ My Uploads
- Algorithms
- My Runs
- My account
- ▷ Create
- ▷ Admin
- Log out
- Repository

Home

1_3.JPG

Submitted By: John Appleseed



Show All

Improving the Quality of Degraded Document Images

Ergina Kavallieratou and Efstathios Stamatatos
Information and Communication Systems Engineering,
University of the Aegean
83200 - Karlovassi, Greece
{ergina.stamatatos}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e., foreground) of the document image. Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents

✓ Paragraph

Element 1

✓ Column

Element 2

Element 3

✓ Title

Element 3

✓ Name

Element 4

✓ Heading

Element 5

Element 6

Document Analysis and Exploitation newsletter

Stay informed on our latest news!

User:
admin

Unsubscribe

Previous issues

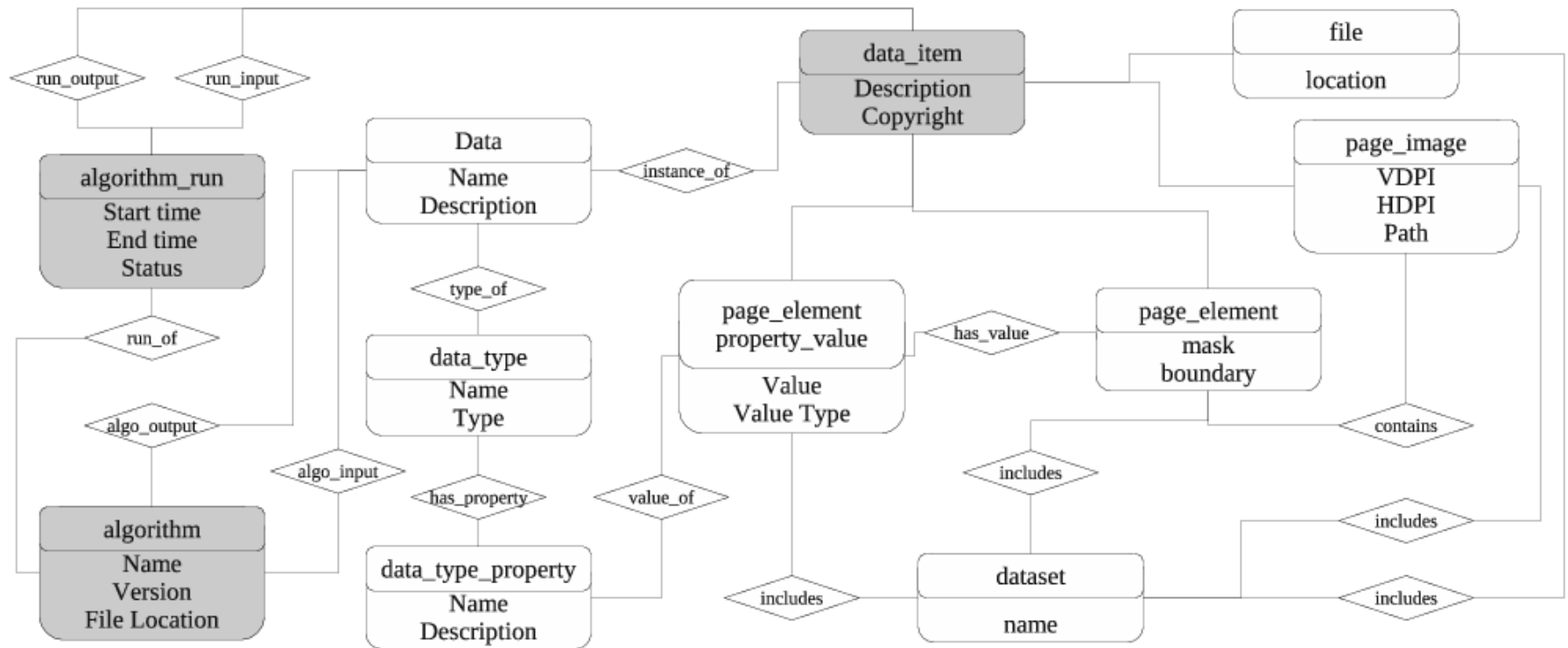


This document is quite nicely annotated and shows embedded and overlapping annotations. However, bounding boxes are not very precise.

CKEditor: the ID for excluding or including this element is browse/dataitem/149394.edit-dae-data-comment.

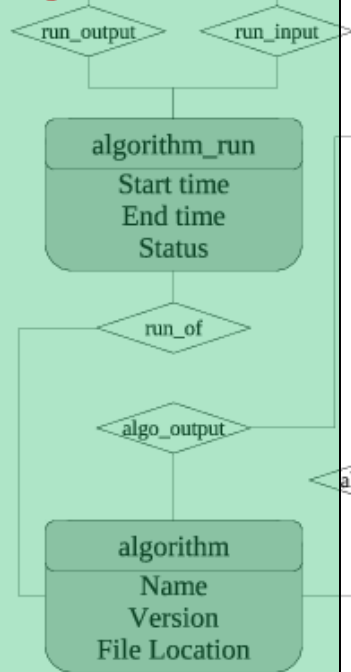
Add Comment

Underlying Data Model

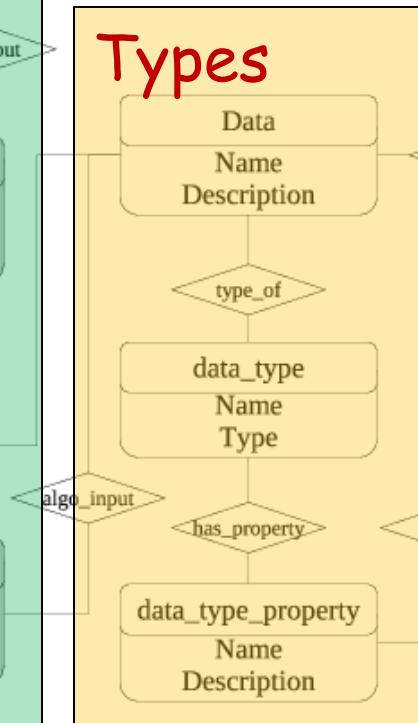


Underlying Data Model

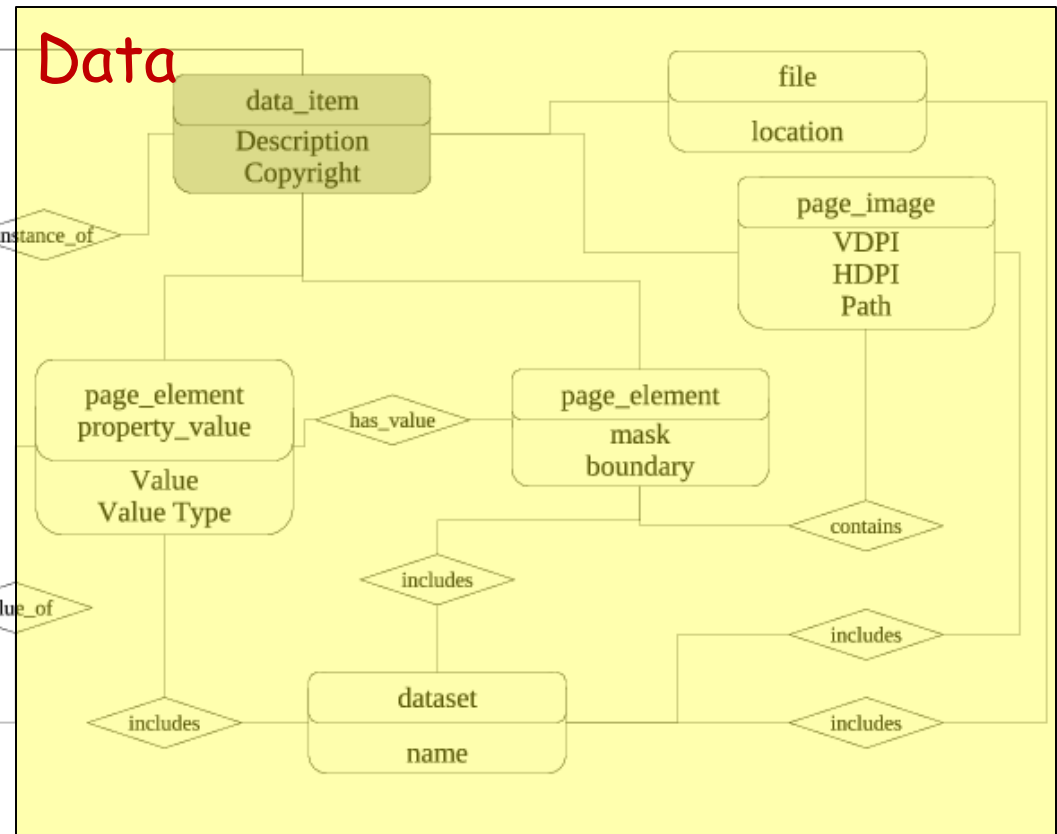
Algorithms



Types

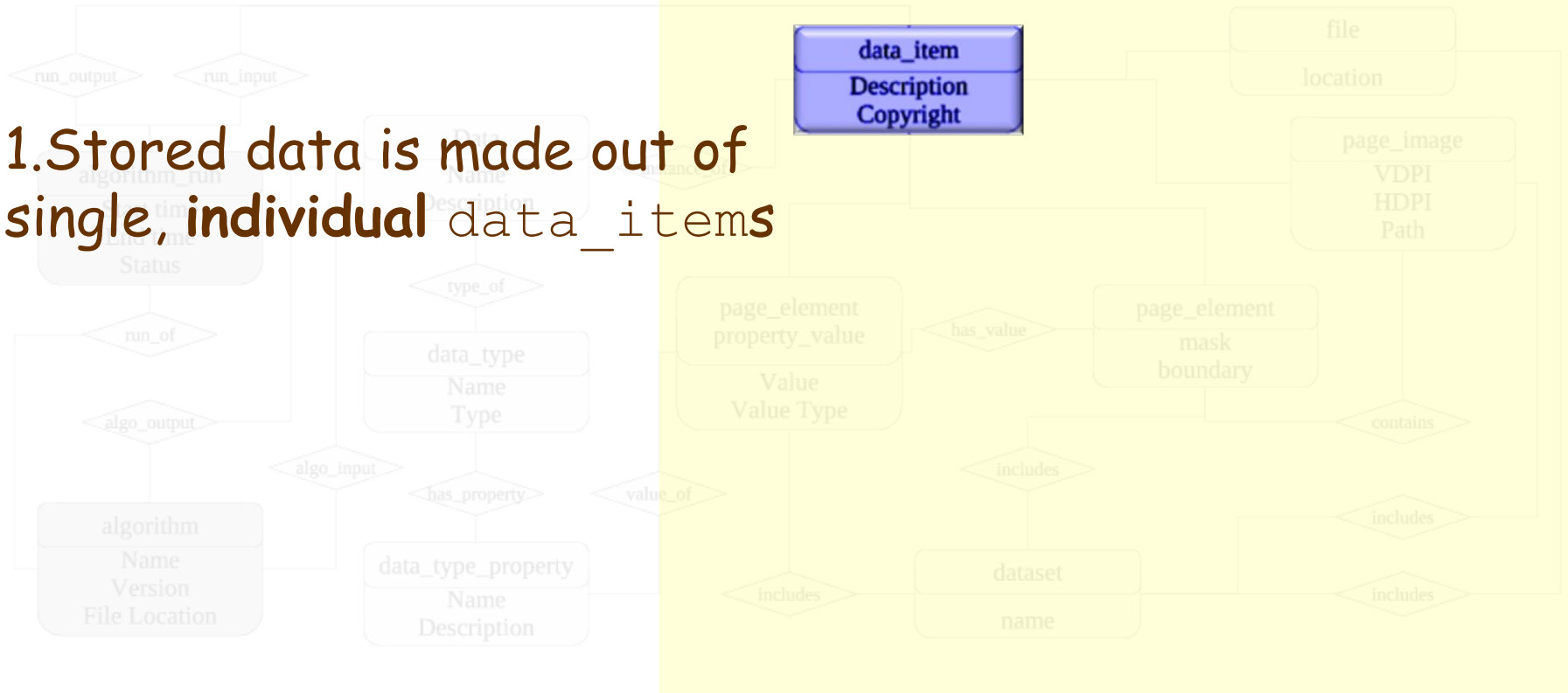


Data



Data Model : data_items

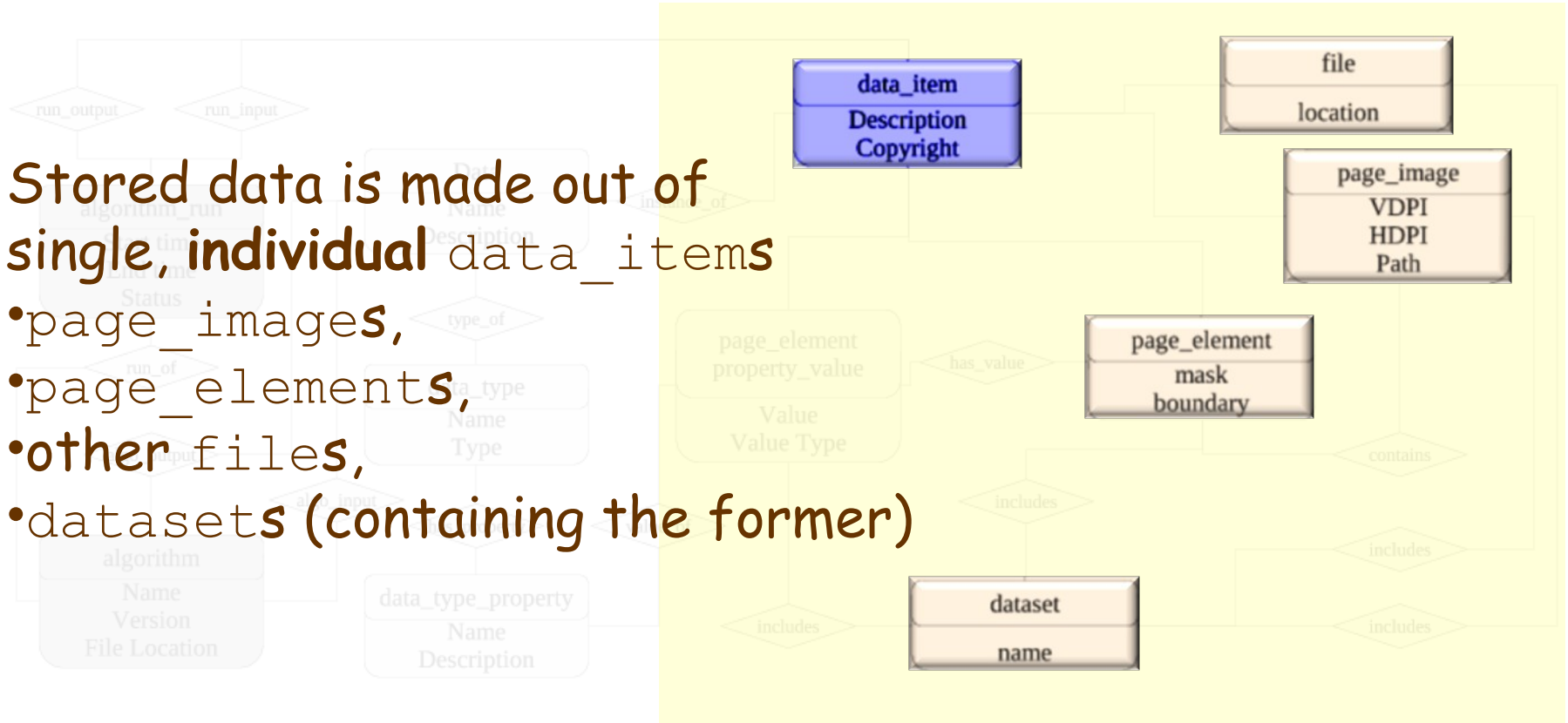
1. Stored data is made out of single, individual data_items



Data Model : data_items

Stored data is made out of single, individual data_items

- page_images,
- page_elements,
- other files,
- datasets (containing the former)

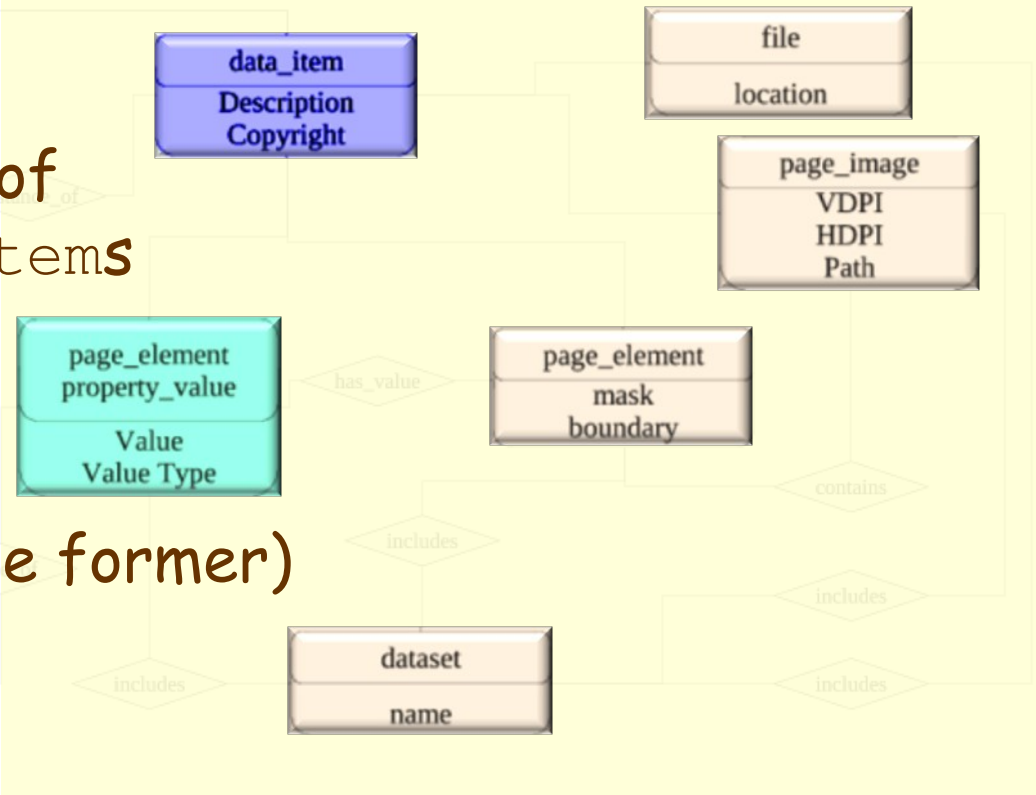


Data Model : data_items

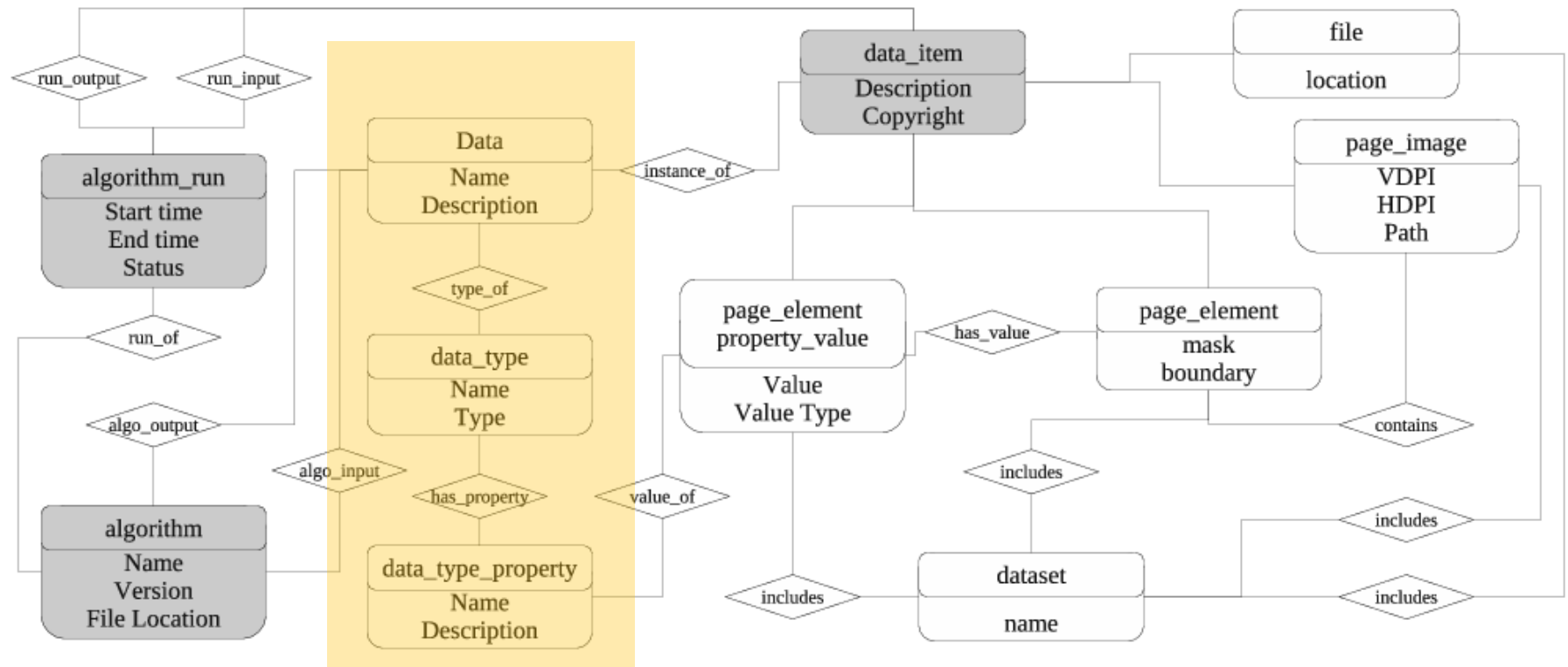
Stored data is made out of single, individual data_items

- page_images,
- page_elements,
- other files,
- datasets (containing the former)

Even interpretations !



Data Model : datatypes



1. All data is typed



Result: Fully Query-able Data

Tobacco800 Dataset:

- **Finding all name labels of identified signatures**

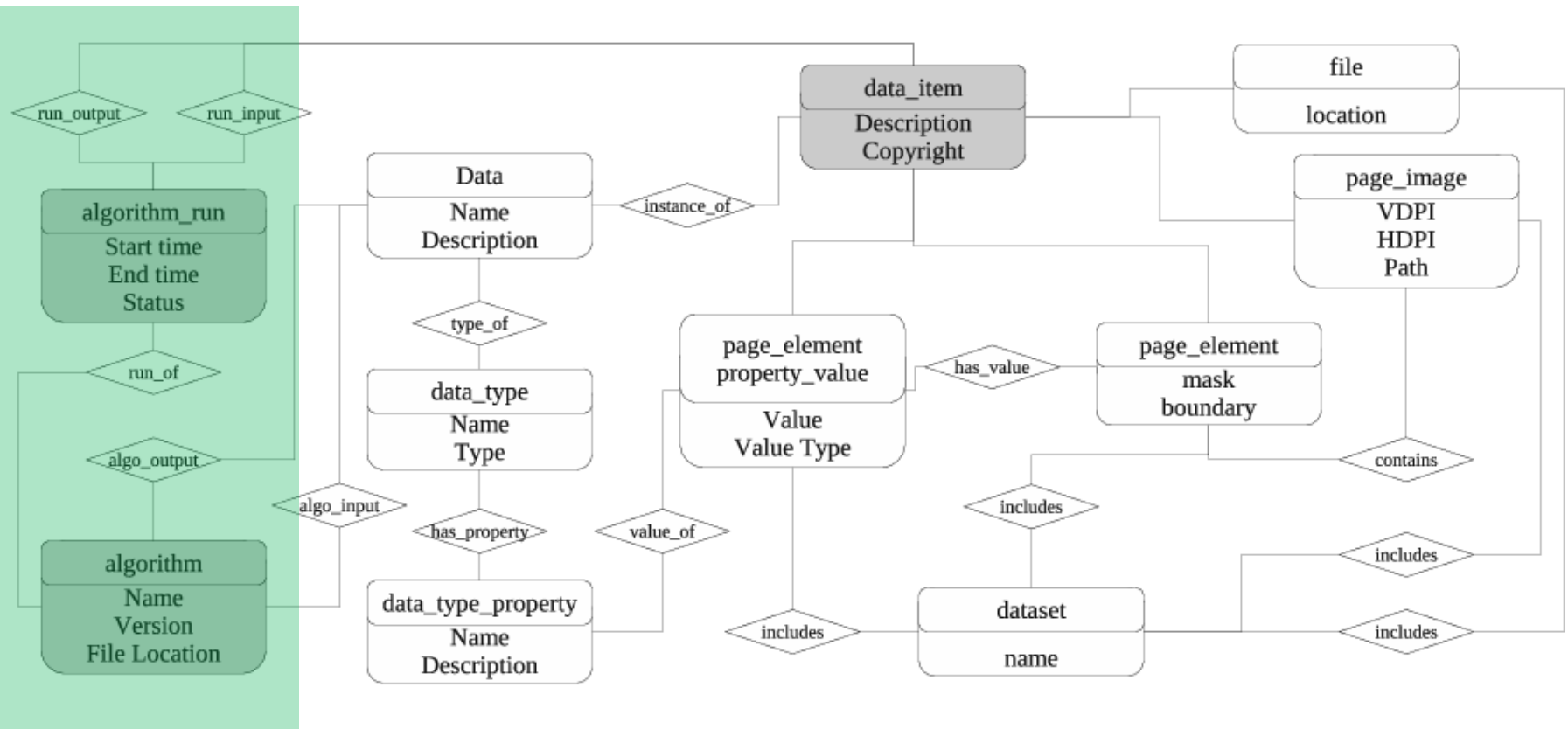
```
select distinct to_char(pepv.value)
from datatype_property dp, value_of vo, page_element_property_value pepv
where
    vo.data_type_PROPERTY_ID = dp.id
    and dp.name = 'gedi_type_DLSignature'
    and pepv.id = vo.page_element_property_value_id;
```

- **Finding all occurrences of a person's signature**

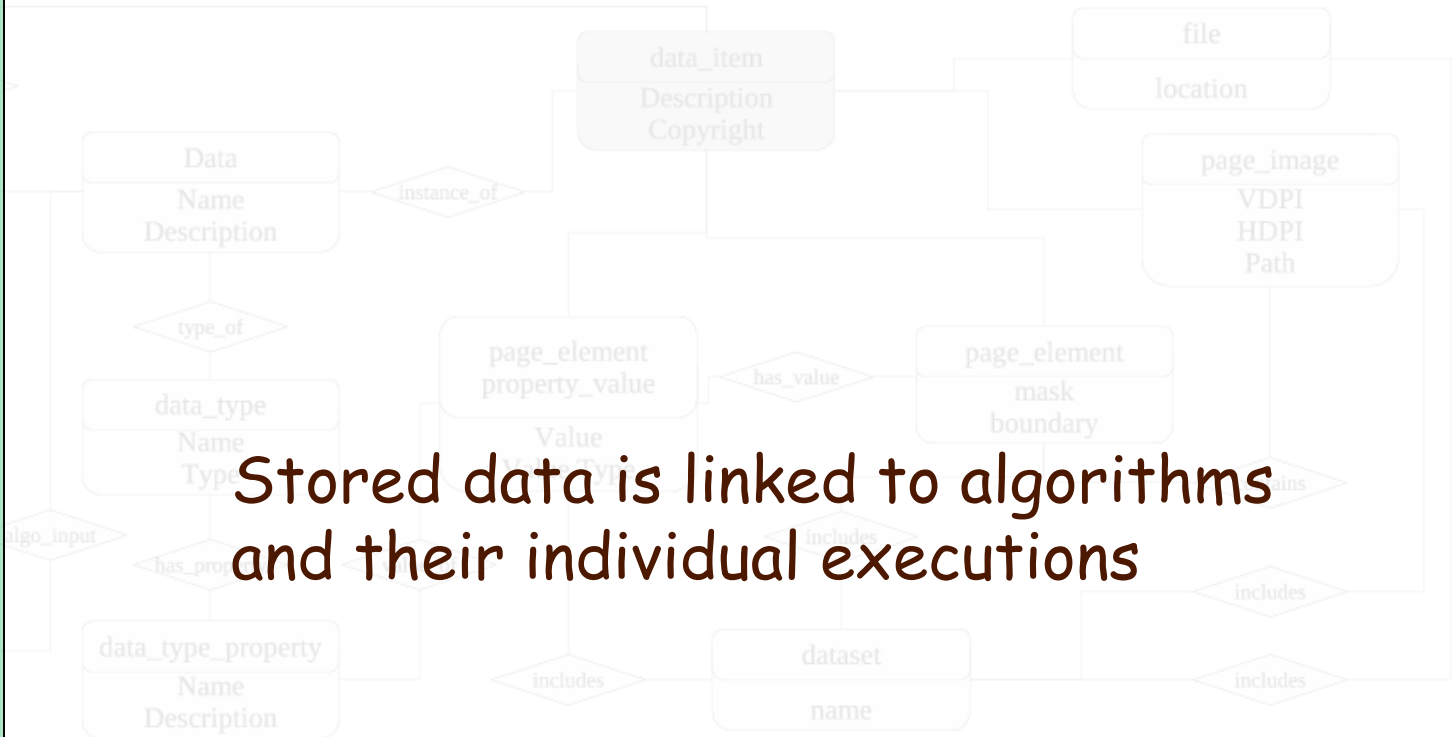
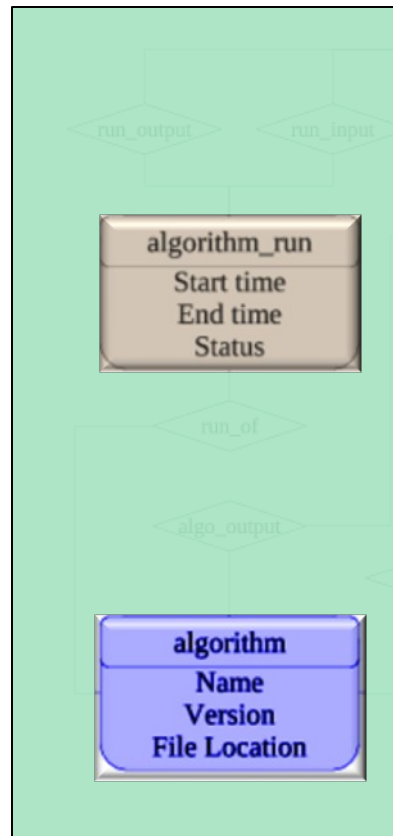
```
select hv.page_element_id
from datatype_property dp, value_of vo, page_element_property_value pepv, has_value hv
where
    vo.data_type_PROPERTY_ID = dp.id
    and dp.name = 'gedi_type_DLSignature'
    and pepv.id = vo.page_element_property_value_id
    and to_char(pepv.value) = 'John Doe'
    and hv.page_element_property_value_id = pepv.id
```



Data Model: Algorithms



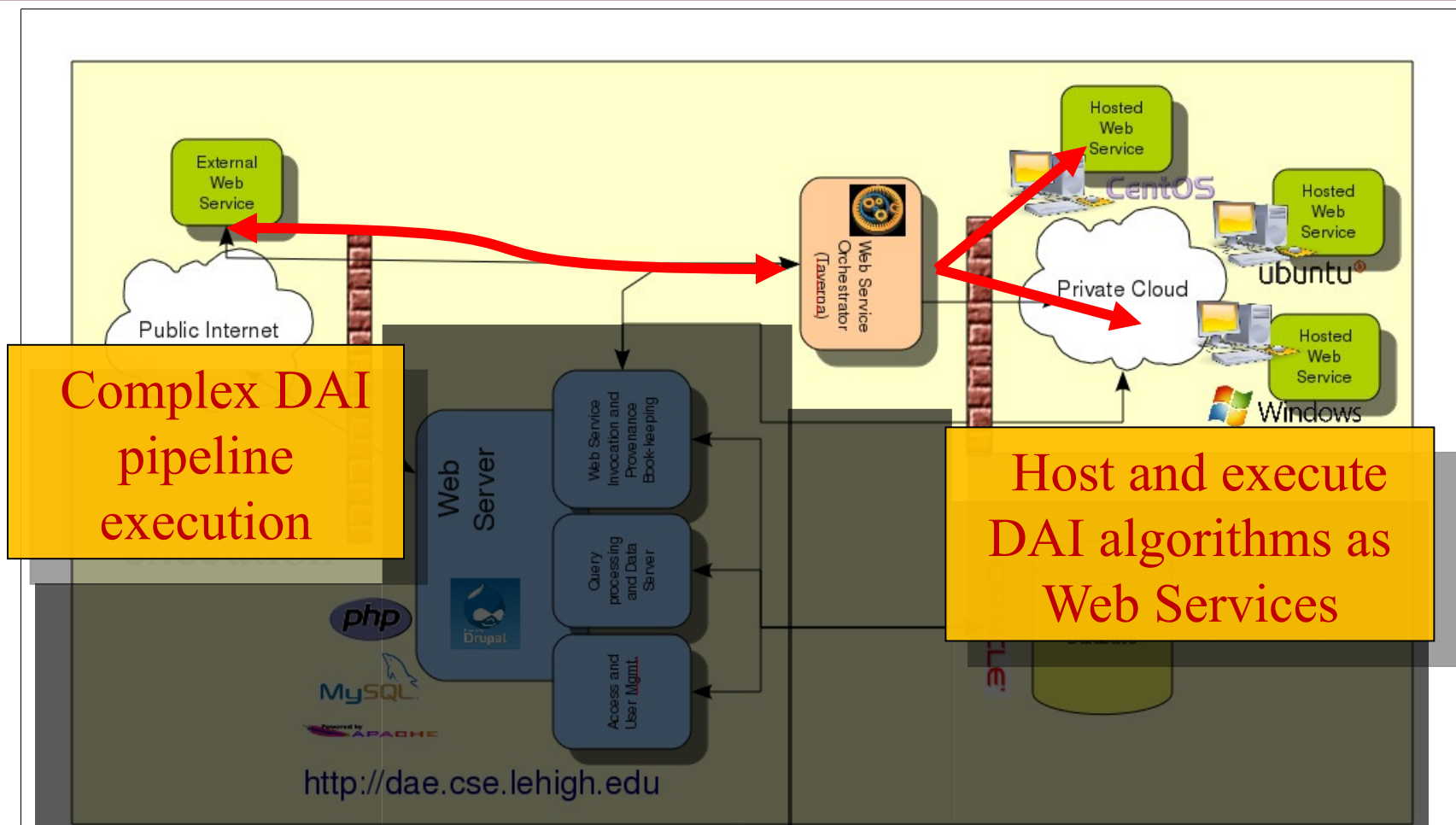
Data Model: Algorithms



Stored data is linked to algorithms and their individual executions



DAE Platform - dae.cse.lehigh.edu



Pitfalls

- **Coercion into**
 - programming languages
 - data structures
 - operating systems
 - releasing control
- **Obsolescence**
- **Home-brew solutions**
(high maintenance costs)
- **Steep learning curve**



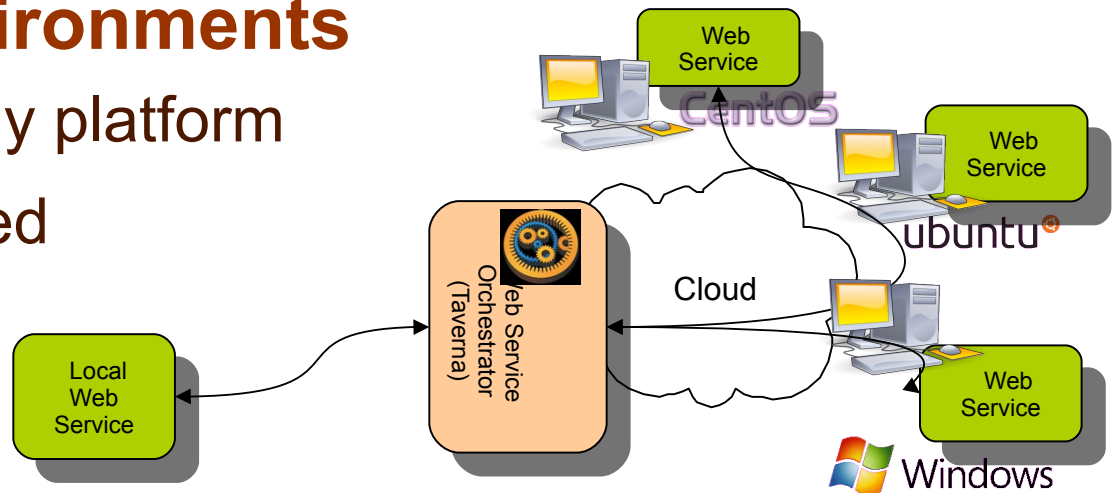
Solution

◉ DAI Algorithms as Web Services

- WSDL open standard
- Not bound to any technology
- Formalized I/O

◉ Virtualized Environments

- Not bound to any platform
- Easily maintained
- Secure
- Cloud ready



Pitfalls

~~No~~ Coercion into

- ~~x~~ programming languages
- ~~x~~ data structures
- ~~x~~ operating systems
- ~~x~~ releasing control

~~No~~ Obsolescence

~~No~~ Home-brew solutions

(high maintenance costs)


? Steep learning curve ?

- No change in habits
- No change of existing code
- No need to release code
- High resilience and maintainability
- High level of interoperability



WSDL & Learning Curve?

- **Plug-and-Play wrappers**

- No change in existing code
- For quick-win testing
(15 lines of code in PHP)
- Taverna  click-and-play

- **More complex integration possible**

- Can be progressive
- Supported by all major programming environments



Benefits

- ◉ **Pipelines = Experiments**
 - Are formalized
 - Can be made available (published) for
 - ▶ Verification
 - ▶ Replay – Re-use
 - ▶ Extension – Modification
 - ▶ Debate
- ◉ **Legacy code can be re-applied and maintained**
- ◉ **No requirement to integrate within the DAE platform**



Bonus for Integration within DAE

◉ Full Provenance

- Find images on which algorithm A disagrees with algorithm B;
- Find images on which no algorithm has found the initially provided interpretation;
- Find interpretations with the strongest level of disagreement;
- Find all data seen by a user (or processed by an algorithm run by the same user);

Etc.



Outline of Talk

- ◉ **General Context**

Challenges of reproducible and traceable experimental research

- ◉ **The DAE Platform**

A way of addressing the challenges, technical architecture and design choices

- ◉ **First Experiments**

Reports of actual uses of the platform

- ◉ **Looking Beyond and New Challenges**



ICDAR 2011 Contest

- ◉ **Domain:**

Document Analysis

- ◉ **Challenge:**

Extract Named Entities from scanned document

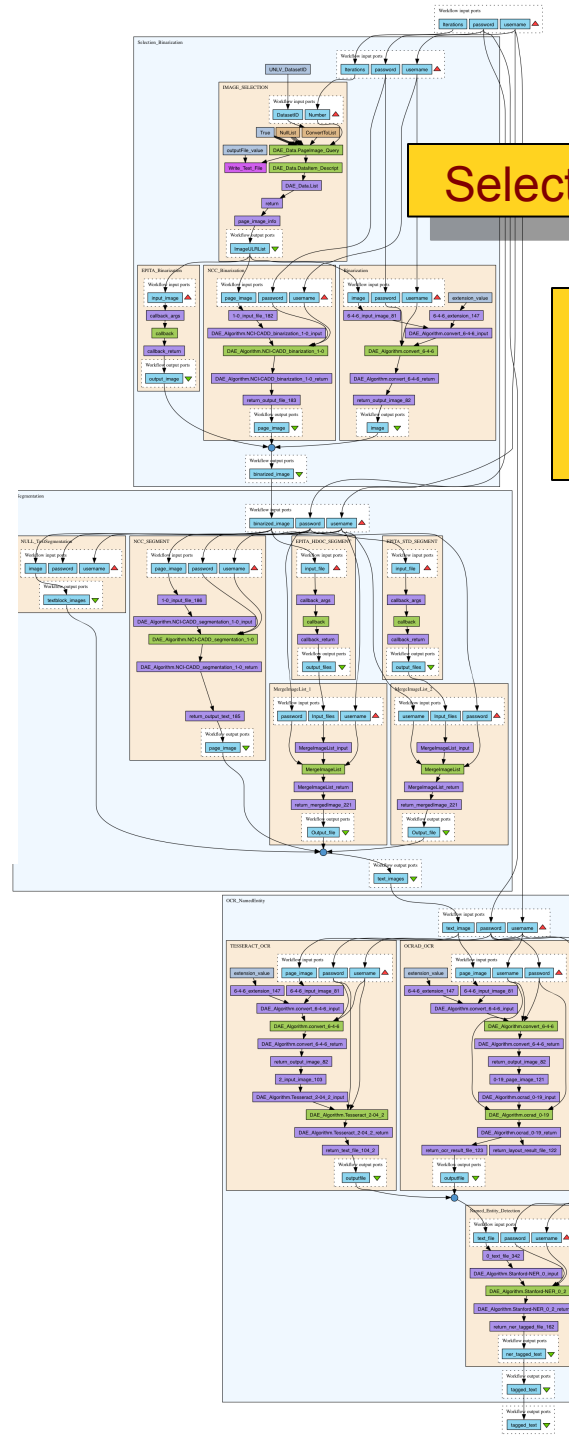
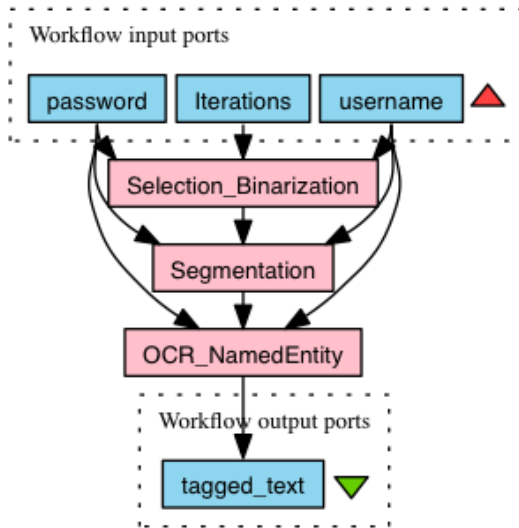
- ◉ **Tools:**

- DAE platform
- Binarization – Segmentation – OCR – NLP
- Web Service wrapper code

- ◉ **Task:**

Contribute 1 component





Selection of Images on Image Server

Binarization Phase:

- ✓ 2 contributed algorithms
- ✓ 1 provided

Text Segmentation Phase:

- ✓ 3 contributed algorithms
- ✓ 1 provided

OCR Phase:

- ✓ 2 provided algorithms

Named Entity Detection

Result

- 3 Binarization – 4 Segmentation – 2 OCR
= 24 pathways to compute same data.

Contributions were

- ▶ DAE hosted (Sun/Solaris, Linux, Windows)
- ▶ Remotely hosted
- Full Experimental Pipeline available and reproducible
- All Intermediate Data available and verifiable
(including provenance: who, what, when)
- Full comparison possible by third party



Outline of Talk

- ◉ **General Context**

Challenges of reproducible and traceable experimental research

- ◉ **The DAE Platform**

A way of addressing the challenges, technical architecture and design choices

- ◉ **First Experiments**

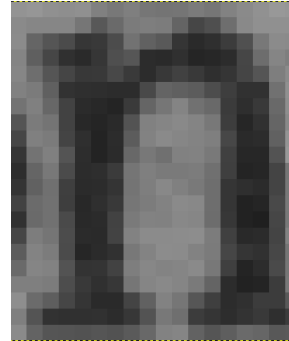
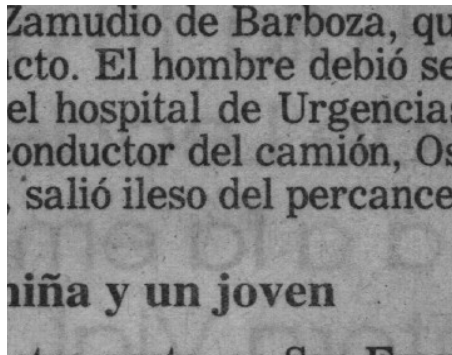
Reports of actual uses of the platform

- ◉ **Looking Beyond and New Challenges**

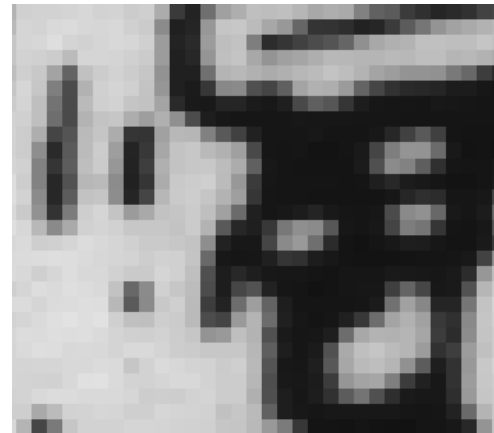
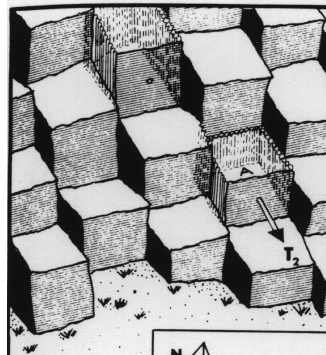


Common Misconceptions

Separating Signal from Background is Easy



Shapes are Obvious



Common Misconceptions

◦ Experimental Validation Data is

- Unique
- Unambiguous
- Universal

NO !

◦ Experimental Validation Data

- Depends on context
- Can be open to controversy
- Can contain errors

YES !



Truth vs. Interpretation

Multiple Interpretation Contexts for Same Data

- Notion of “ground truth” is obsolete
- Data interpretation depends on
 - “intent” of the author (quasi-impossible to obtain)
 - “expression” of the author (usually undisputable)
 - “interpretation” of the user(s)
 - Human users
 - Algorithms

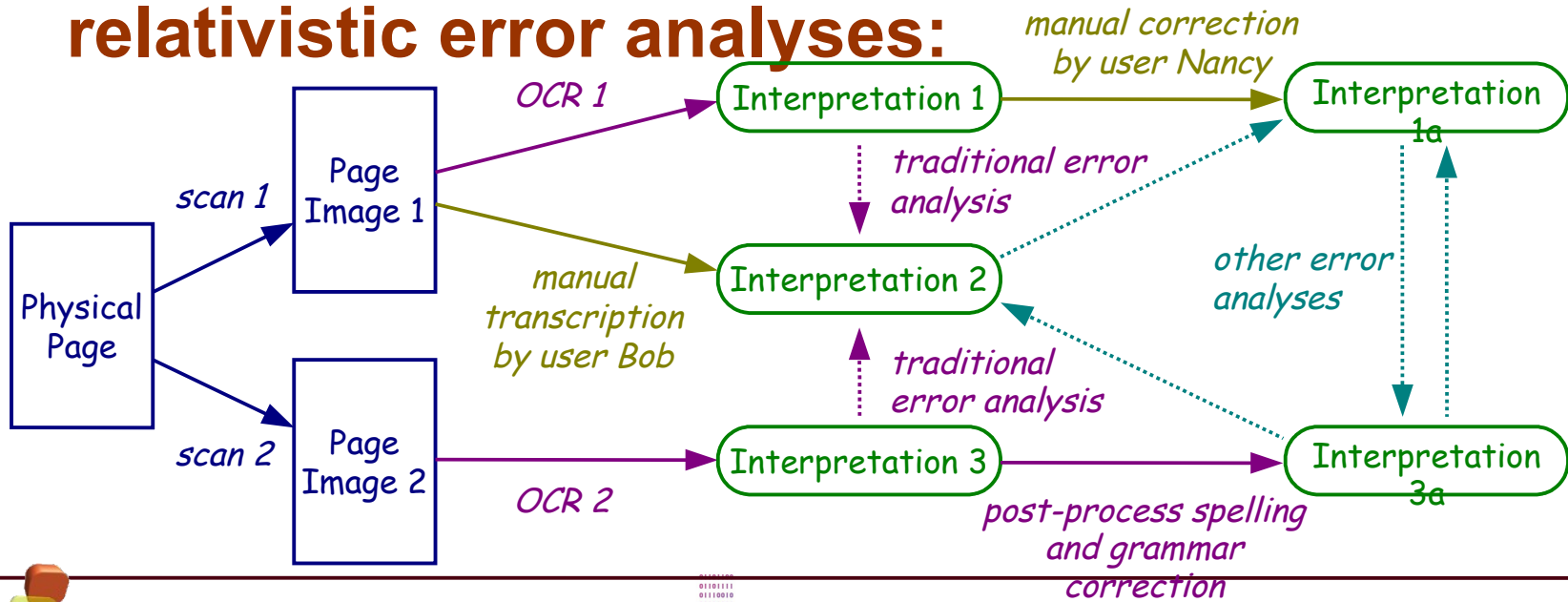


Alternate Interpretations

Entities on a page are located and interpreted:

- by humans interacting with the system
- by algorithms invoked by “others”

Possibility of alternate interpretations and relativistic error analyses:



Interpretations and Reputation

- **Which Interpretation is “the best”?**
- **Which Interpretation is incompatible with my context?**
 - Who has interacted with similar data?
 - Do I trust some more than others?
 - What is their interpretation context?
 - What is their reputation?



Kinds of Reputation

- **Algorithm**

How well does it address a stated task?

- **Implementation**

Buggy or robust?

- **Dataset**

Is it representative for a particular problem?

- **Interpretation**

Is it trustworthy?

- **Publication**

Is it highly cited?

- **Researcher**

Amalgamation of contributions to above



Further Work

- **User Interaction**

- Querying
- Semantics Discovery
- Multiple Interpretation Selection

- **Promotion and Dissemination**

- Multiple Platform Interaction
- Community Development



Further Work

- **User Interaction**

- Querying
- Semantics Discovery
- Multiple Interpretation Selection

- **Promotion and Dissemination**

- Multiple Platform Interaction
- Community Development



Semantics Discovery

- **Context**

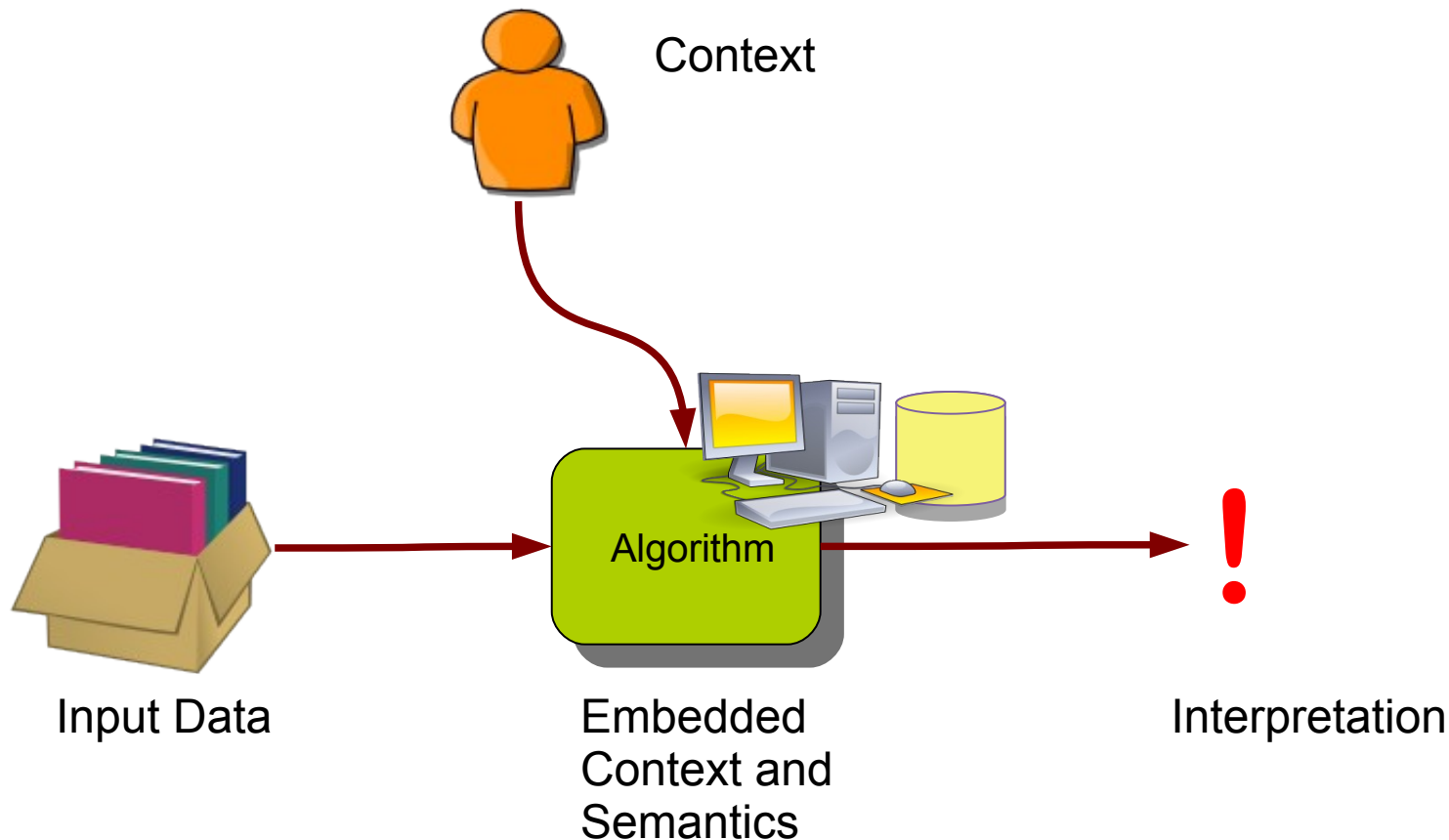
- The semantics of “*Semantics*” ill defined
- Interpretation and Context discovery

- **Tools**

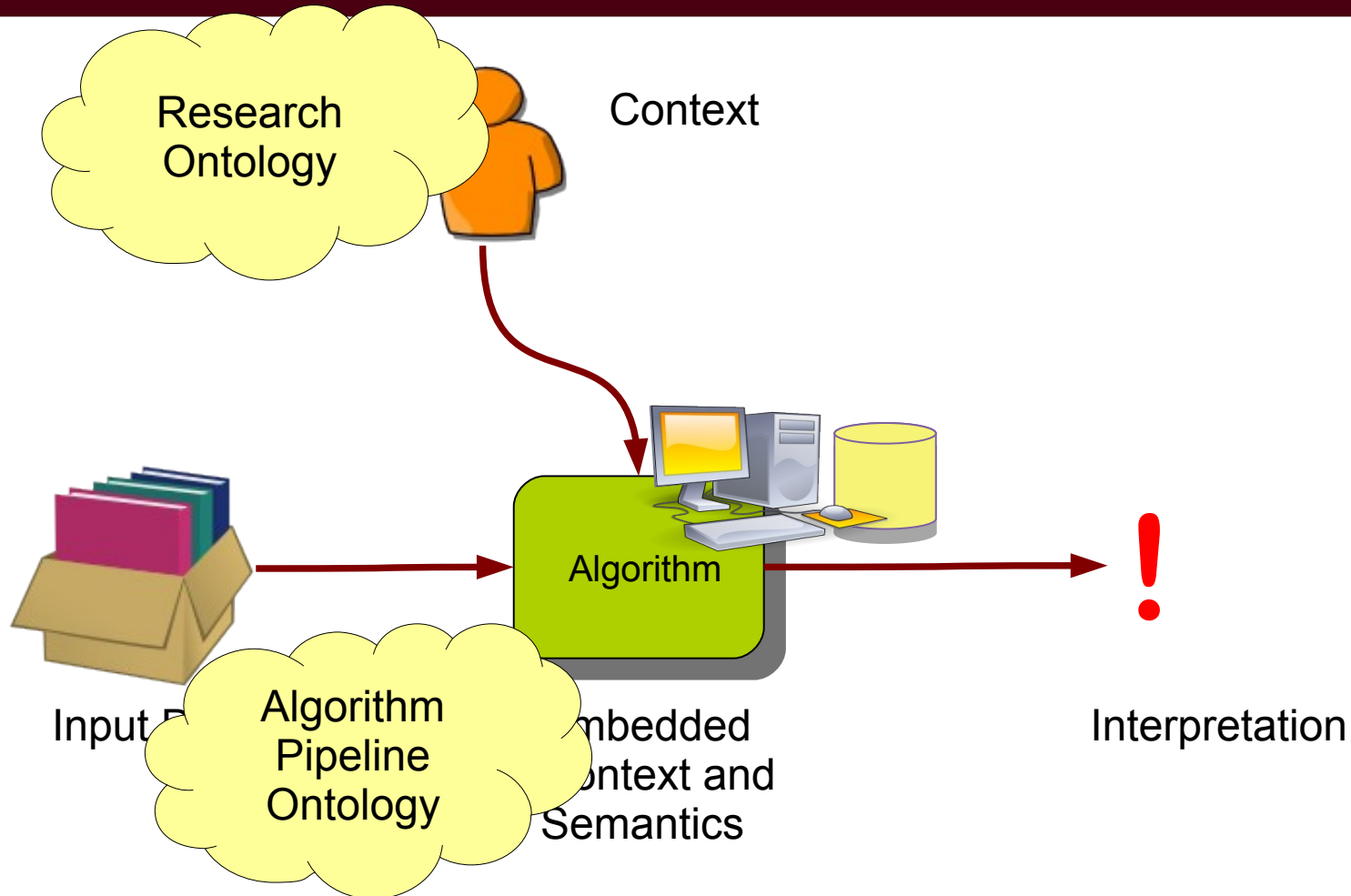
- DAE Provenance data
- Data interaction ontology
- Researcher ontology



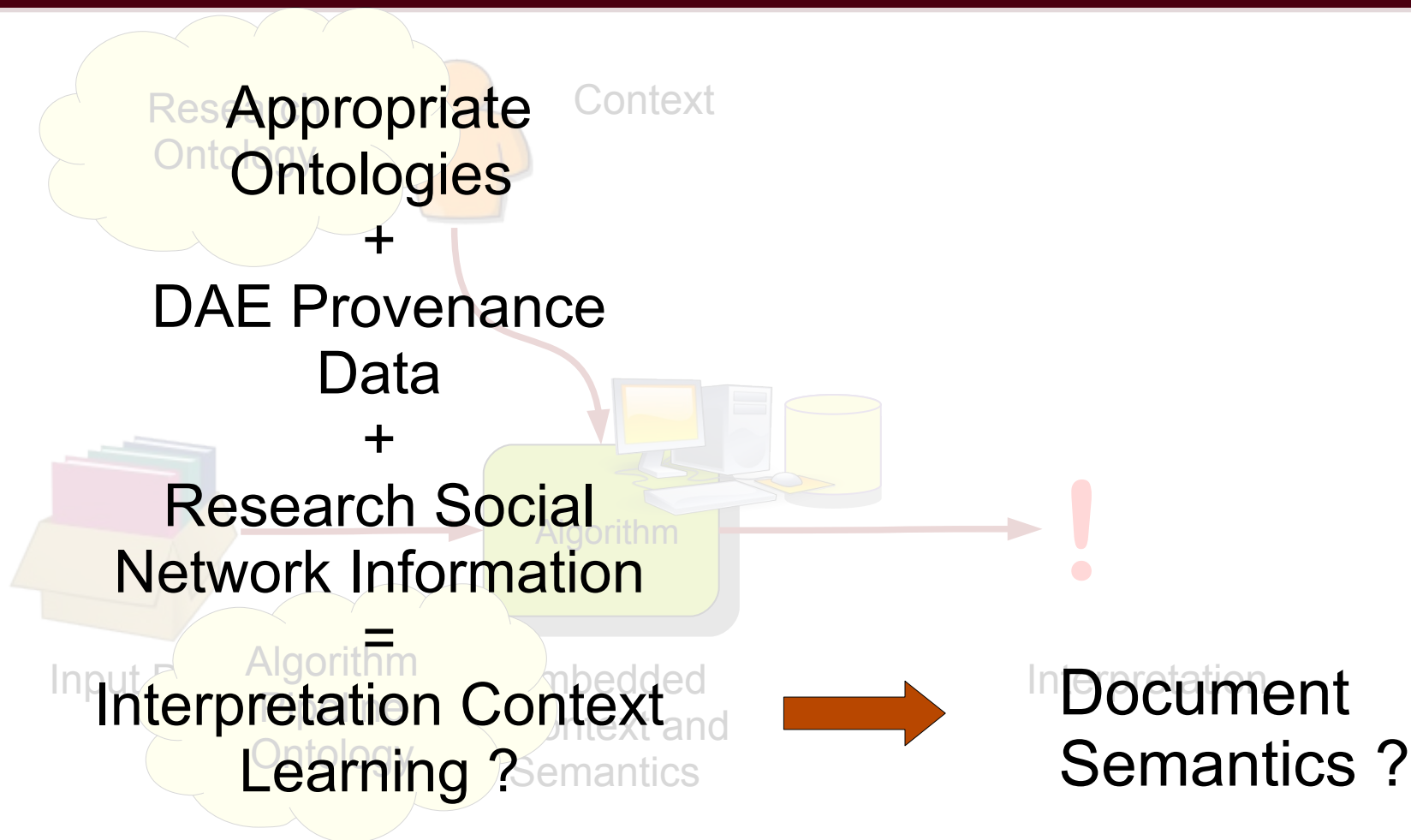
Semantics Discovery



Context Learning



Open Question



Ontology1320681377361 (http://www.semanticweb.org/ontologies/2011/10/Ontology1320681377361.owl) - [/Users/lamiroy/ontologies/Ontology1320681377361/Ontology1320681377361.owl]

Ontology1320681377361 (http://www.semanticweb.org/ontologies/2011/10/Ontology1320681377361.owl)

Active Ontology Entities Classes Object Properties Data Properties Individuals OWLViz DL Query OntoGraf

Class hierarchy Class hierarchy (inferred)

Class hierarchy: Thing

- Thing
 - Domaine_Scientifique
 - Evenement_Scientifique
 - Conférence
 - Personne
 - Scientifique
 - Publication

Object property hierarchy Data property hierarchy Individuals by type

Object property hierarchy:

- topObjectProperty
 - Applies_to
 - Participe_à
 - is_Relevant_to
 - Couvre
 - Traite_de
 - Travaille_sur
 - may_be_Relevant_to
 - pourrait_Travailler_sur
 - shares_Domain
 - Organise
 - est_Auteur_de
 - est_Ecrit_par
 - est_Présenté_à
 - is_CoAuthor_of

Individual Annotations Individual Usage

Annotations: Olivier_Gaucher

Annotations +

Description: Olivier_Gaucher

Types +

- Scientifique

Same individuals +

Different individuals +

Property assertions: Olivier_Gaucher

Object property assertions +

- Travaille_sur Circle_Detection
- Travaille_sur Segmentation
- est_Auteur_de Robust_Circle_Detection
- is_CoAuthor_of Olivier_Gaucher
- is_CoAuthor_of Laurent_Fritz
- is_CoAuthor_of Bart_Lamiroy
- shares_Domain Yassine_Guebbas
- shares_Domain Olivier_Gaucher
- shares_Domain Laurent_Fritz
- shares_Domain Robust_Circle_Detection
- shares_Domain Robust_Circular_Arc_Detection
- shares_Domain Bart_Lamiroy
- is_Relevant_to Circle_Detection
- is_Relevant_to Segmentation

Data property assertions +

Reasoner active ☒ Show Inferences



Further Work

- **User Interaction**

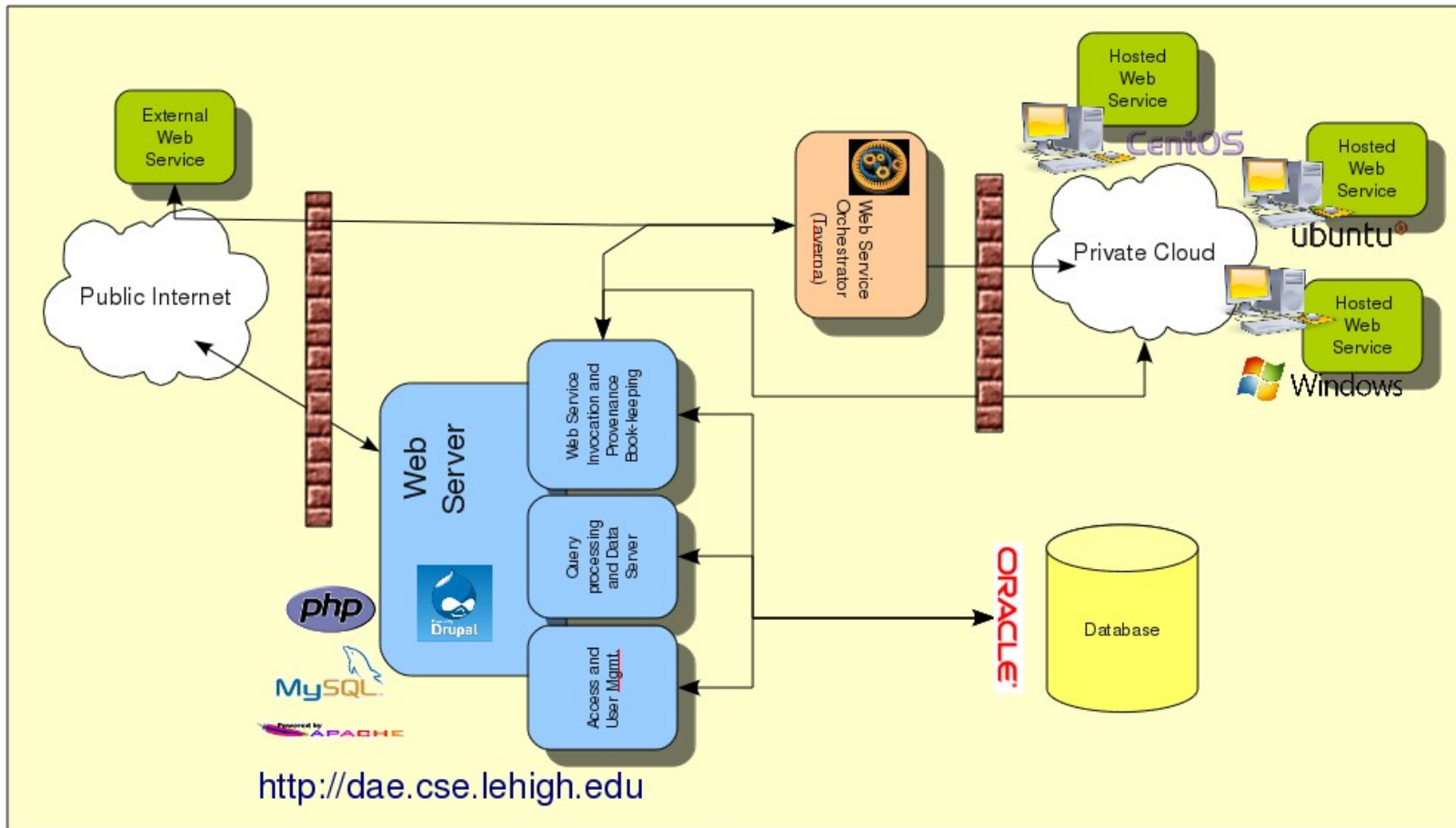
- Querying
- Semantics Discovery
- Multiple Interpretation Selection

- **Promotion and Dissemination**

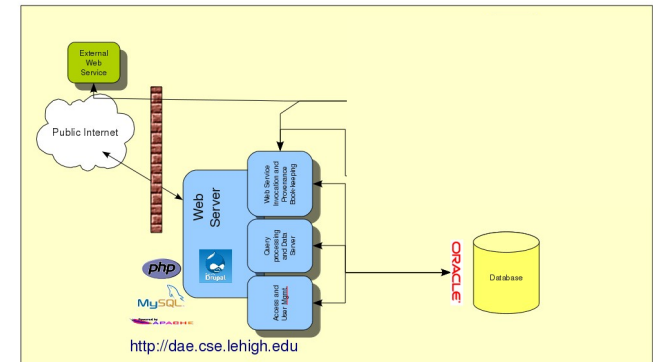
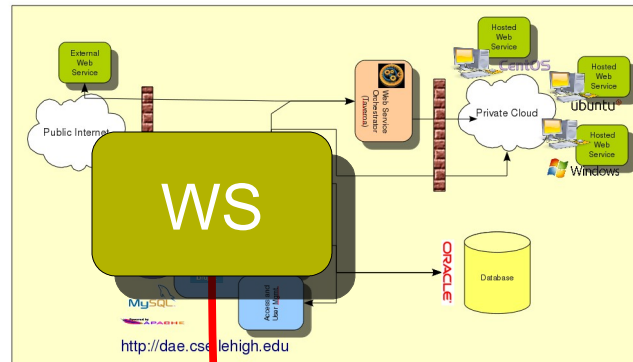
- Multiple Platform Interaction
- Community Development



DAE Architecture

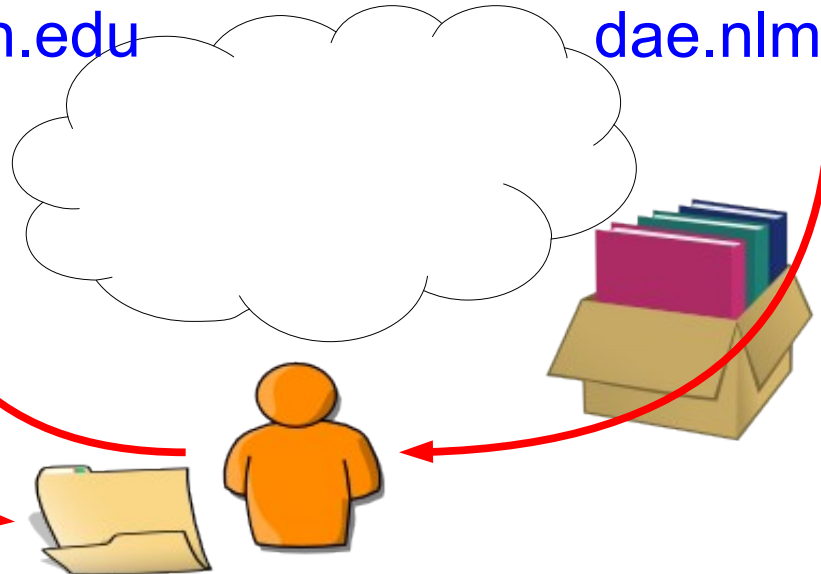


Multiple DAE Platforms

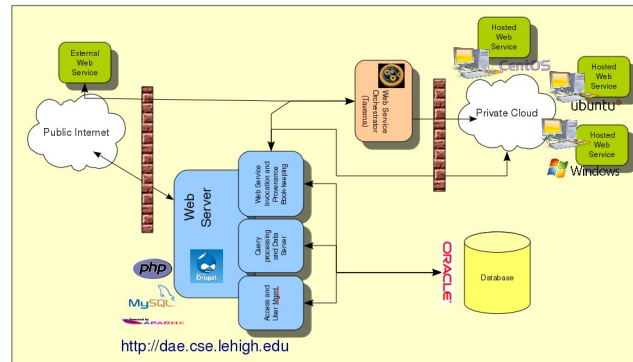


dae.cse.lehigh.edu

dae.nlm.nih.gov

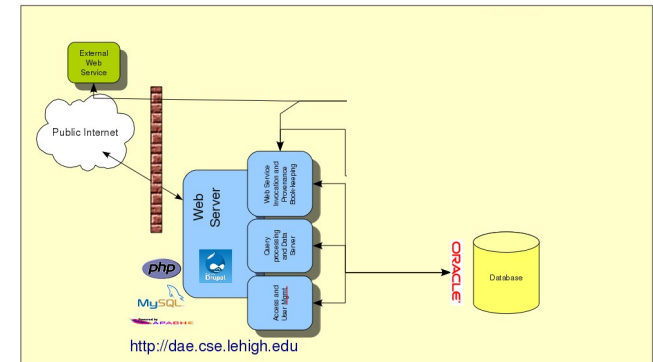


Extending Provenance Registry



dae.cse.lehigh.edu

- Algorithm WS was run by UserX using DocY from dae-nlm.cse.lehigh.edu
- Register provenance
- Register dae-nlm

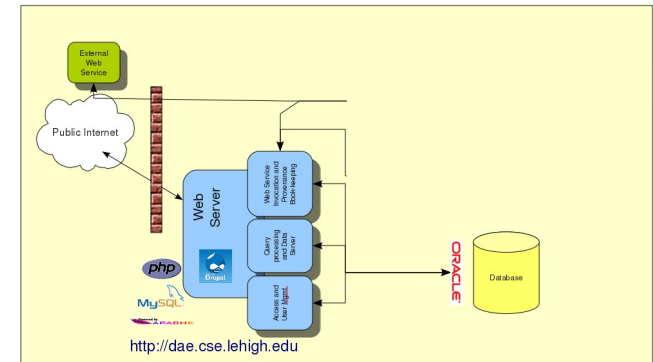
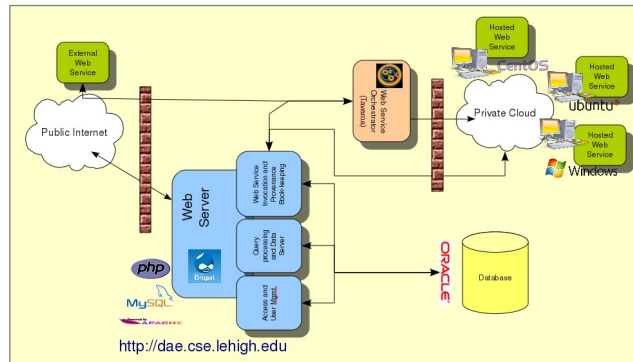


dae.nlm.nih.gov

- UserX has accessed DocY from dae.cse.lehigh.edu
- Register dae



Social Network Building



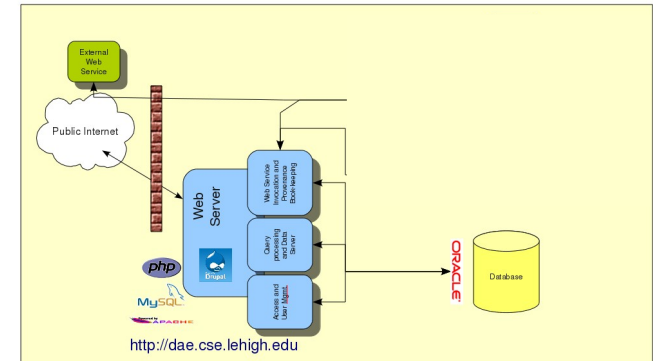
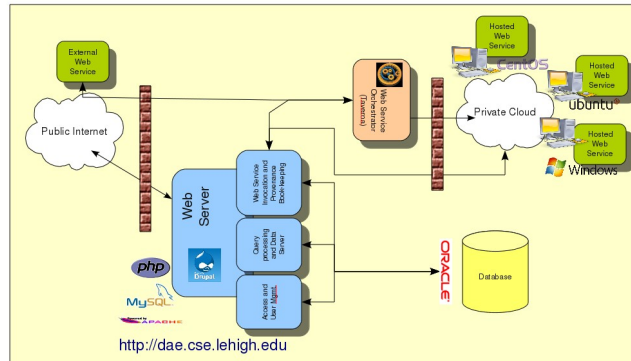
dae.cse.lehigh.edu

dae.nlm.nih.gov

Who are your DAE neighbors ?



Distributed Provenance



dae.cse.lehigh.edu

dae.nlm.nih.gov

Give DocY descendents



Technical Issues

- **Uniform User Identification**

- Drupal accepts OpenID
- Some integration required for web-services
- Taverna?

- **External data_items**

- OK in current model
- Handling resiliency (versioning) with caching
- Copyright issues of descendants?

- **Constructing Distributed Provenance Queries**

- Avoiding loops?



Conclusion

DAE Plaform is an Operational tool for:

- large, open and re-usable collections
- handling multiple interpretations of data
- user defined and extended new data types
- certifiable interaction with data
- certifiable algorithm benchmarking



Resources

- <http://dae.cse.lehigh.edu>
- <http://sourceforge.net/projects/daeplatform/>
- <http://tinyurl.com/DAE-Wiki>
- <http://tinyurl.com/DAE-WebServices>

