TATA CONSULTANCY SERVICES
Experience certainty.

J&J Clinical Trial Landscape Analysis : Progress Update

DESS Artificial Intelligence

TCS **Understanding :** Landscape Analysis

- Clinical trials are enormously expensive in terms of cost and time.
- It's cost effective to run a hypothetical, mathematical model of the clinical trial rather than the actual trial and to run mathematical models input data is necessary
- Currently, agency personnel manually go through each and every study, identify the data of relevance (to the clinical trial), extract the data and prepare a structured report
- The data is mostly present in unstructured text, charts and tables in various studies.
- Data extraction is mostly manual
- Limitations of current process:
  - ✓ Time consuming
  - ✓ Error prone
  - ✓ Expensive
- The average duration for each cycle is around 4-6 months an hence limits the number of models validated in a year

Landscape Analysis : Proposed Solution

1. Information Sources

- Identify the literature sources
- Identify the key words for search
- Decide inclusion/exclusion criteria
- Perform Search
- Review search results and Identify relevant documents

2. Information Ontology

- Define Attributes
- Define Attribute hierarchy

3. Information Extraction

- Upload the Information documentss
- Define the Information Ontology
- Automate the attribute extraction

Landscape Analysis : Information Sources

- Clinical Trial Registries
  - ClinicalTrials.gov
  - EU Clinical Trials Register
  - WHO ICTRP
- Literature Database
  - PubMed
  - Embase
  - Cochrane
  - Scopus
  - SEER

- Conference Abstracts*
  - ASCO
  - ESMO
  - Oncologists Meet
- Health Regulatory Websites
  - FDA
  - EMA

*** The conference abstracts will be therapeutic area specific. Currently a representative list of conferences for oncology is indicated.**

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such	
	3b	Important changes	reasons
Participants	4a	Eligibility	
	4b		
Interventions		allow replication, including how and when they were	
Outcomes		primary and secondary outcome measures, including how and when they	
		to trial outcomes after the trial commenced, with reasons	
Sample size		How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

Do we use this to explain the process of defining an ontology for a therapeutic area? Or do we do it differently?

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

Example: "Heart attack" AND "Los Angeles"
Search for studies: Search
Advanced Search | Help | Studies by Topic | Glossary

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

Brief Title ICMJE
Official Title ICMJE
Brief Summary
Detailed Description
Study Type ICMJE
Study Phase
Study Design ICMJE
Condition ICMJE
Intervention ICMJE
Study Arms

Recruitment Status ICMJE
Enrollment ICMJE
Completion Date
Primary Completion Date
Eligibility Criteria ICMJE

Sex/Gender
Ages
Accepts Healthy Volunteers
Contacts ICMJE
Listed Location Countries ICMJE

Participant Flow: Overall Study

STARTED
COMPLETED
NOT COMPLETED
Withdrawal by Subject
declined any surgery
partial cystectomy vs radical cystectomy

Measured Values

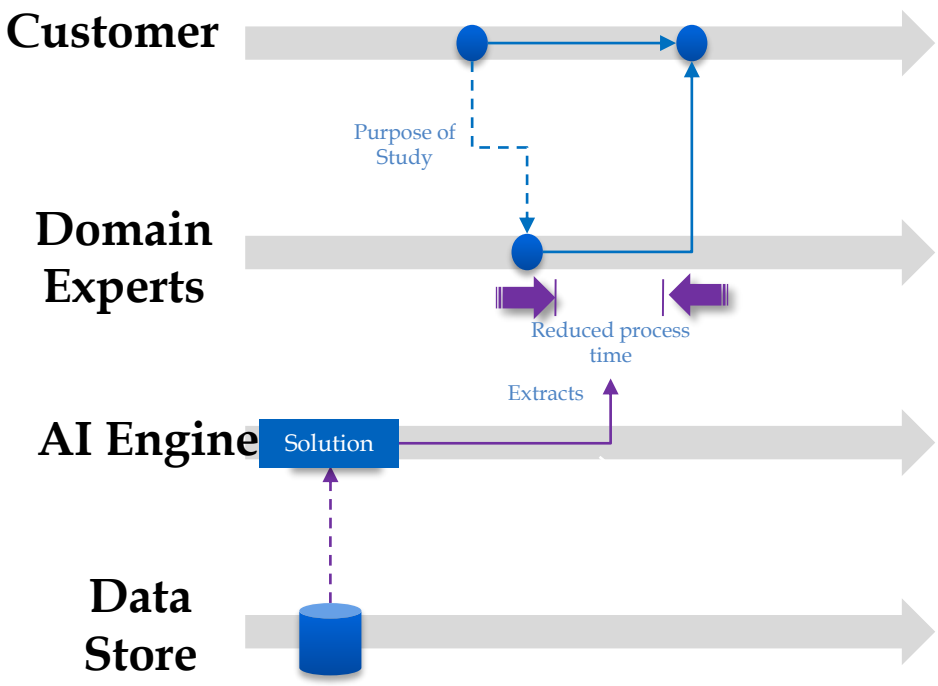
Participants Analyzed
[Units: Participants]

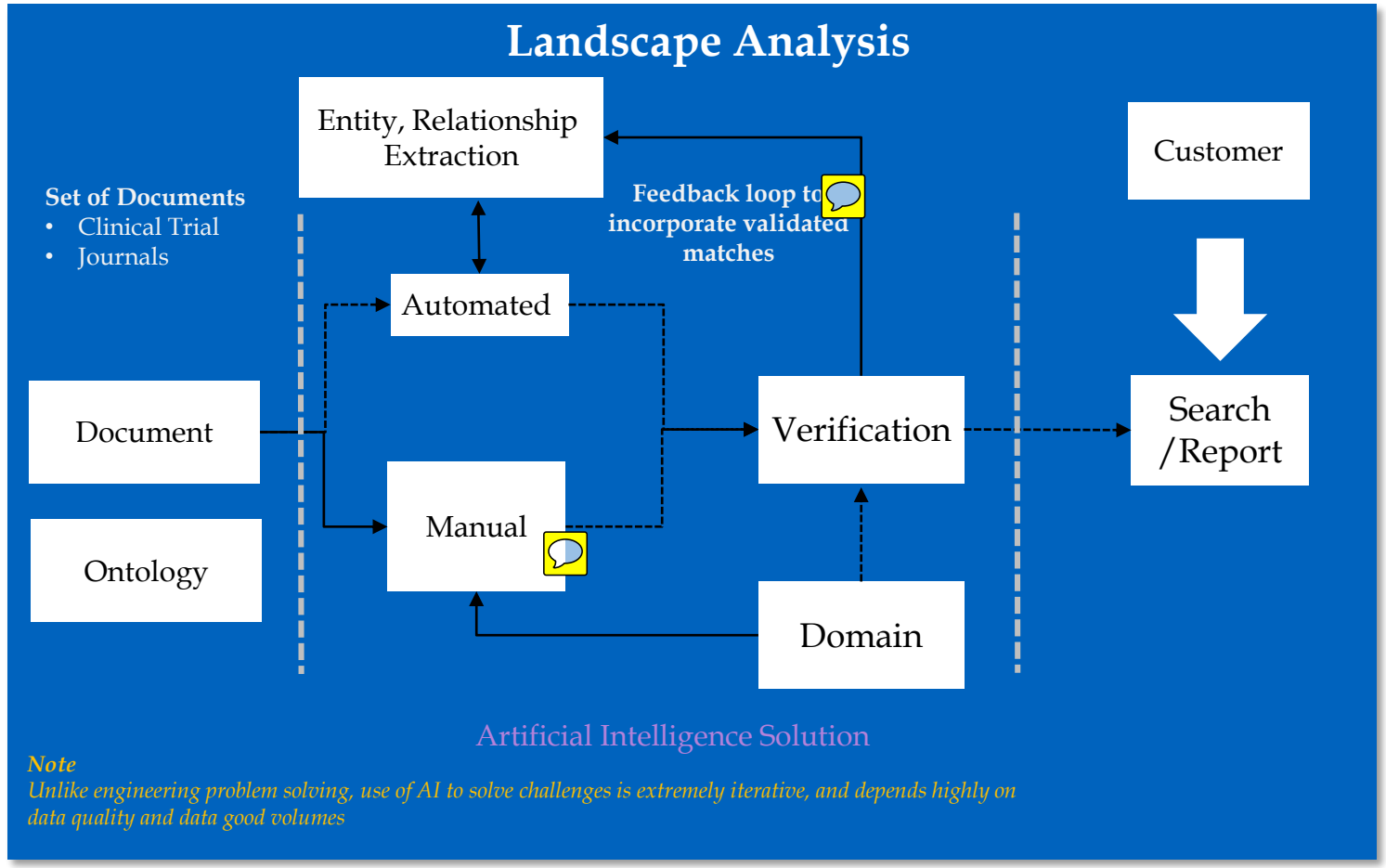
The Pathologic Complete Response Rate (<pT0) of Neoadjuvant GCS Regimen in Patients With Muscle-invasive Bladder Cancer.
[Units: Percentage of participants]
Number (95% Confidence Interval)

e Measures

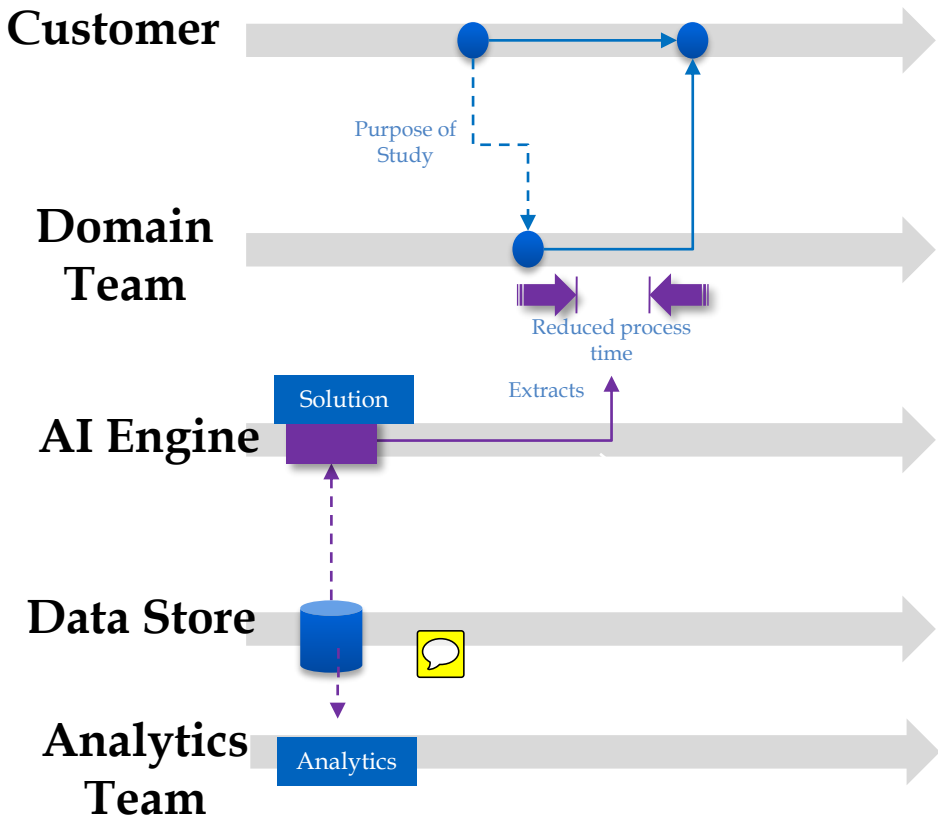
Participants Analyzed
Participants]
Participants]
3 years
ween 18 and 65 years
5 years
Participants]
ale
of Enrollment
Participants]
ed States

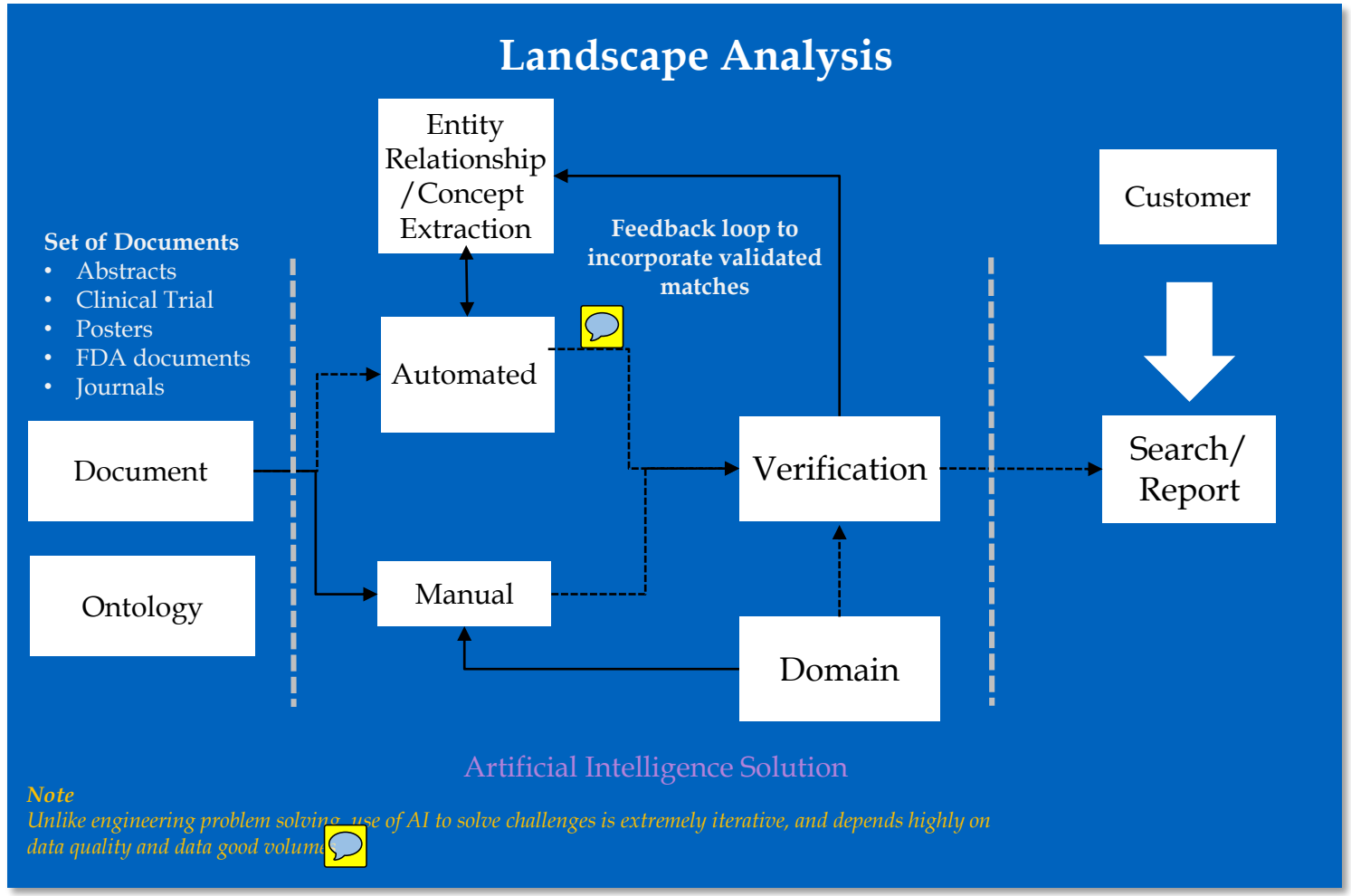
Landscape Analysis : Initial Phase





Landscape Analysis : At Maturity





TATA CONSULTANCY SERVICES
Experience certainty.

Pilot – Landscape Analysis Bladder Cancer

PoC Demo: URL: http://ai.tcsmobilitycloud.com/jj-clinical-trial/

Data sources for Pilot

- Two main data sources have been identified for the pilot phase for the data extraction:
  1. **Registry and results database (Clinicaltrial.gov):** Studies are generally submitted to the Web site (that is, registered) when they begin, and the information on the site is updated throughout the study. Results of the study are submitted after the study ends. Information available on clinical trials registries:
     - **Summaries of Clinical Study Protocols** that includes summary of the purpose of the study, recruiting status, disease or condition and medical product under study, study design, phase of the trial, inclusion/exclusion criteria, location of the trial and contact information
     - **Summaries of Clinical Study Results** that include description of study participants (e.g., number enrolled, demographic data), overall outcomes of the study, summary of adverse events experienced by participants

  2. **Pubmed:** provides access to information in MEDLINE as well as additional life science journals, integrated molecular biology databases (NCBI), and journals/manuscripts deposited in PMC, Both MEDLINE and other PubMed citations may have links to full-text articles or manuscripts in PMC, NCBI Bookshelf, and publishers' Web sites

Criteria taken in consideration while performing literature search (For Internal Use only)

- Indications: e.g. Hematuria
- Type of Bladder Cancer: e.g. TCC, Adenocarcinoma etc.
- Correlation of cancer prevalence: Age, Gender, Race, Geographical regions
- Lifestyle modifications: Smoking, non-smoking
- Type of studies: Pre-clinical, In-vitro, Ex-vivo, Diagnostic etc.
- Disease condition: Metastasis, Recurrence, Co-morbidity, Pregnancy
- Treatment type: Chemotherapy, Surgery, Immunotherapy, Gene therapy, Radiation therap etc.
- Endpoints: e.g. Overall survival, Disease free survival, Progression free survival etc.
- Study design and phase: Randomization, Blinding, Phase of trial etc.

Do we include this?

DATA SOURCES & STATISTICS

Data Sources	URINARY BLADDER CANCER		
Clinicaltrial.gov	Filters	Studies	
		1094	
	With Results	99	
	Without Results	995	
	In 2015	25*	
PubMed			
	Filters	Links	In Humans
		58268	
	Free Full Text	12368	10584
	Last 5 Years	3859	3088
	In 2015	955	783*

* Number of publications, articles, studies with the citation of bladder cancer data considered for extraction for pilot phase of Landscape analysis

JOURNALS

Name of the Journal	No of bladder cancer articles
PLOS ONE	58
Oncotarget	51
BioMed Central	42
International Journal of Clinical and Experimental Pathology	27
Japanese Journal of Clinical Oncology	24
Nature Research Journal	22

AddtionalJournals _BladderCancer

* Names listed above are the journals with the maximum number of articles published on 'Urinary Bladder Cancer'. Attached is the list of other journals which has also published urinary bladder cancer related articles (in lesser numbers).

BLADDER CANCER : ONTOLOGY

Delete Add Edit

- Source
  - *Source Number*
  - *Title*
  - *Authors*
  - *Publication Year*
- Trial
  - *Trial Identifier Number*
  - *Registry Name*
  - *Status*
  - *Notes*
  - *Location*
  - *Health Authority*
  - *Sponsor*
  - *Collaborator*
  - *Investigator*
  - *Phase*
  - *Study Purpose*
  - *Trial Design*
  - *Enrollment*
  - *Randomization*
  - *Multicentric*
  - *Trial Blinding*
  - Trial Period
    - *Trial Year*
    - *Start Date*
    - *Completion Date*

Delete Add Edit

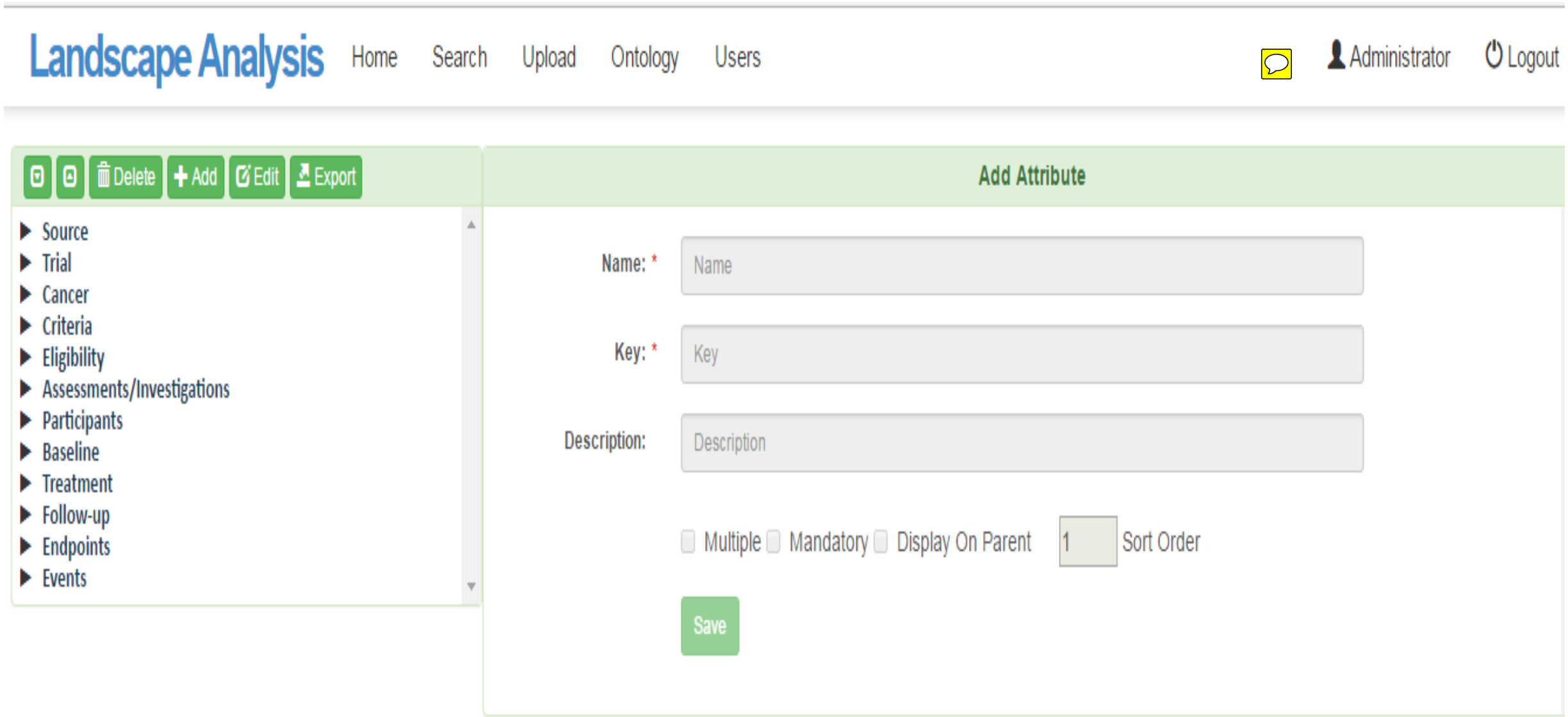
- Trial
- Cancer
  - *Cancer Site*
  - *Cancer Type*
  - *Cancer Stage*
  - *Grade*
  - *Recurrence*
  - *Prognosis*
  - TNM Classification
    - *Solitary Tumor (T)*
    - *Cancer in Lymph Node (N)*
    - *Metastasis (M)*
- Criteria
  - Smoking History
    - *Smokers*
    - *Non-smokers*
  - Strata
    - *Count*
    - *Type*
- Eligibility
- Assessments/Investigations
  - Investigation
    - *Name*
- Participants
- Baseline

Delete Add Edit

- Participants
  - Participant
    - *Period*
    - Arm
      - *Arm Label*
      - Count
        - *Count*
        - *Type*
      - Incomplete Count
        - *Count*
        - *Type*
- Baseline
- Treatment
- Follow-up
  - *Frequency*
  - *Duration*
- Endpoints
  - Endpoint
    - *Classification*
    - Arm
      - *Participants*
      - *Arm Label*
      - Measurement
        - *Type*
        - *Sub Type*
        - *Parameter*

BladderCancer_Ontology

UTILITY HOMEPAGE



Landscape Analysis PILOT : Technology Stack

S.no	Software	Version	License
1	Java SE Development Kit (JDK1.7)	1.7	Sun License
2	Python	2.7	Python Software Foundation License Version
3	Tesseract	3.04.00	Apache License, Version 2.0
4	Apache Tomcat Server	7.0.69	Apache License, Version 2.0
5	Elastic Search	2.3.2	Apache 2
6	Apache Maven	3.2.3	Apache License Version 2.0
7	WebPlotDigitizer*	3.9	GNU General Public License Version 3
8	OntoText S4	1.0	Pay per Use
9	Medex	1.3.5	Apache License Version 2.0
10	Grobid	0.4.1	Apache License Version 2.0
11	Cermine	1.11	GNU Affero GPL 3

Hardware
4 cores , 16 GB RAM , 3.6 GHz, x64 and 1 TB SSD storage, - Cent OS/Ubuntu

* License to be procured

Landscape Analysis PILOT : Key Learnings & Focus Area

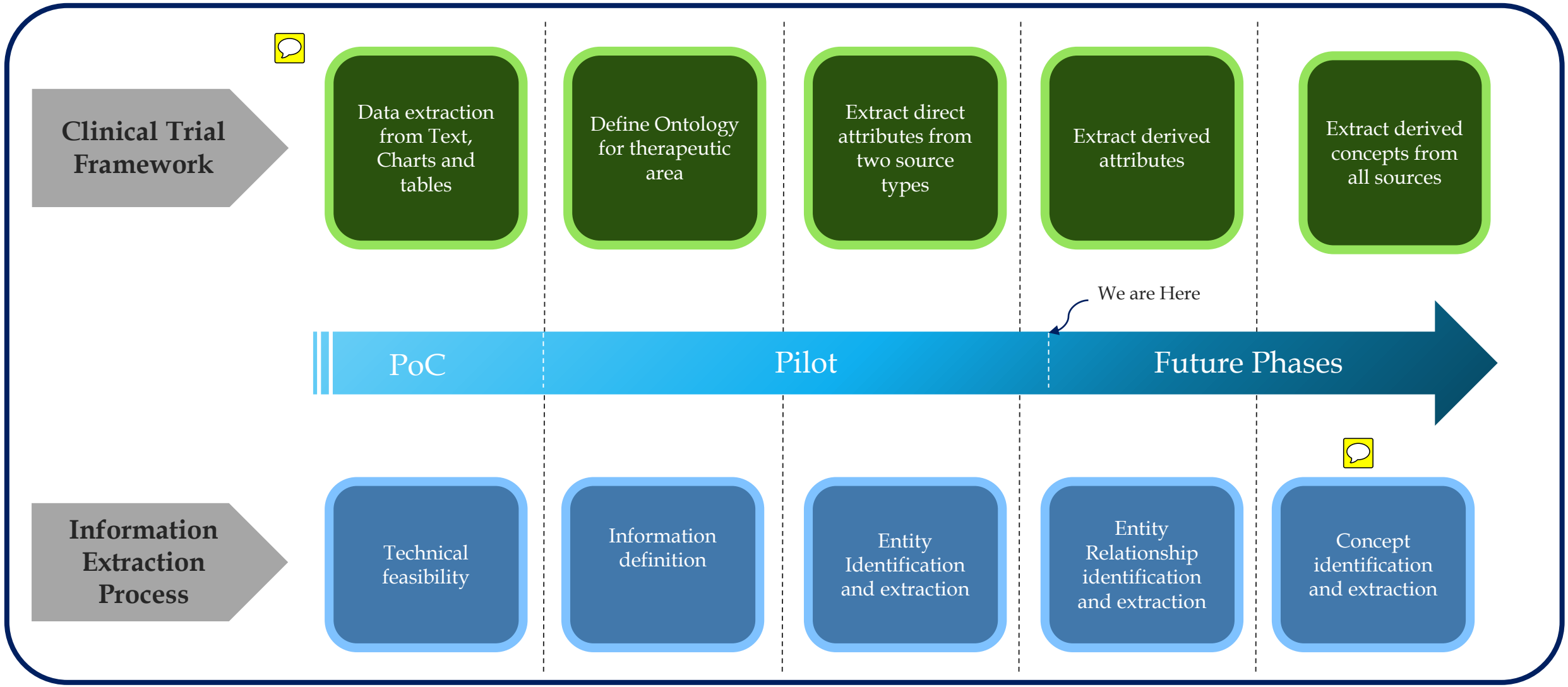
Key Learning

- Coverage of Information extraction from Journals can be extended with availability of sufficient labelled
- Attribute identification from tables and charts would be possible with adequate training data through deep learning approaches

Focus Areas

- Incorporate J&J feedback on improvising the solution approach
- Guidance and validation from J&J on the data sources/ontology
- Applying deep learning recurrent neural networks for improved accuracy

Landscape Analysis : Information Extraction Road Map



TATA CONSULTANCY SERVICES
Experience certainty.

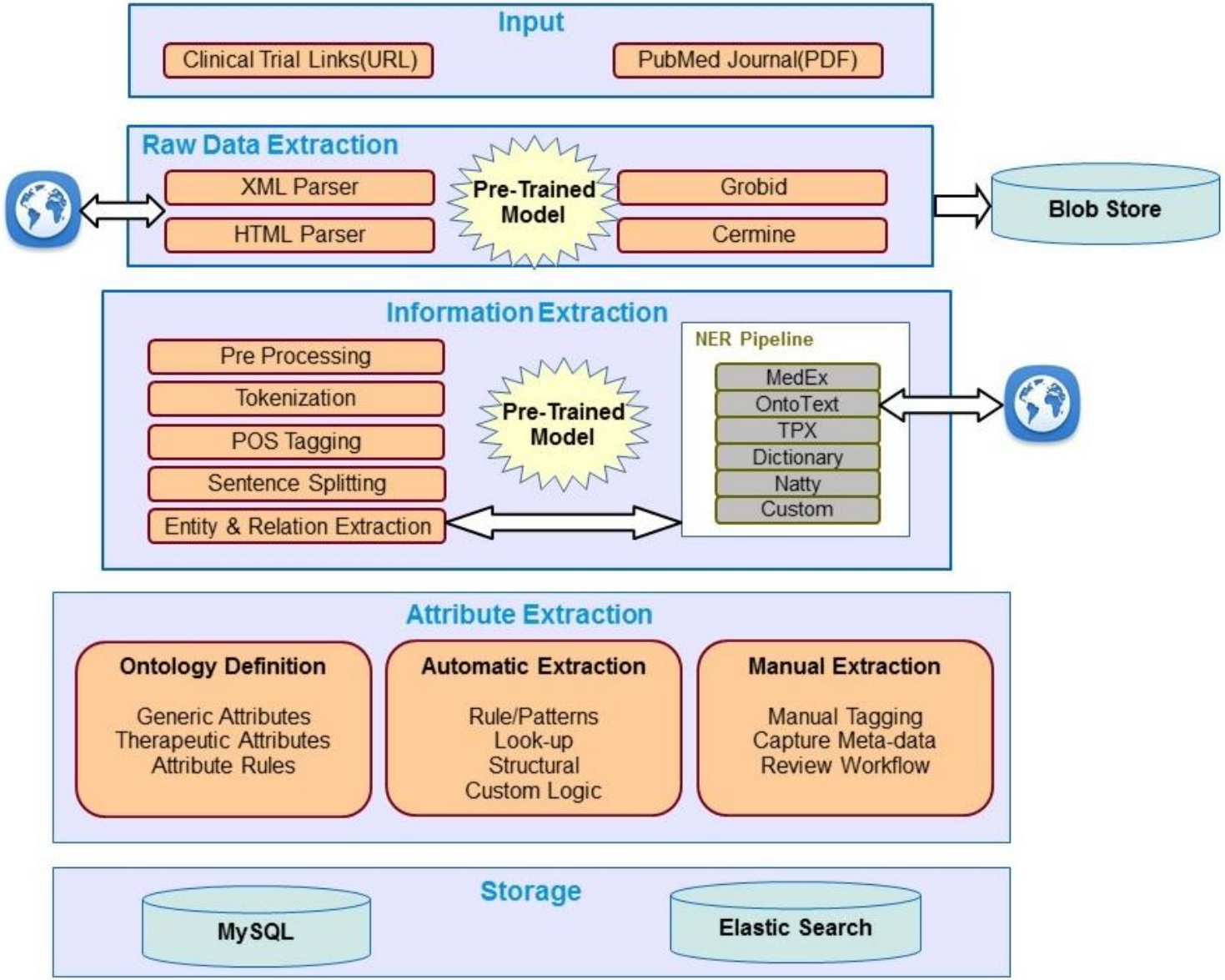
Thank You

IT Services
Business Solutions
Outsourcing

TATA CONSULTANCY SERVICES
Experience certainty.

Appendix

Landscape Analysis PILOT : **Components**



Clinical Trial Landscape Analysis PILOT : **Architecture**