Narrativa

A Natural Language Artificial Intelligence and Data Augmentation Platform for Life Sciences and Healthcare

Confidential

Narrativa

Powering the data stories that transform the world

Narrativa is an internationally recognized **business intelligence company** that uses its **artificial intelligence** and machine learning platform to build and deploy **natural language content solutions for enterprises**.

Its proprietary technology suite, comprised of **data extraction**, **data analysis**, **natural language processing** (NLP) and **natural language generation** (NLG) tools, all work together seamlessly to power a lineup of smart content creation, reporting automation and process optimization products.

Narrativa's platform enables Life Sciences companies to speed up the clinical trial process by rapid data interpretation and report automation, reducing time, cost and risk.

We take data like this…

{
"_class_Study": {
"study_id": "XXXX-XX-XXXX",
"rid": "#study_XXXX-XX-XXXX"
},
"_class_Subject": [
{
"id": "XXXX-XX-XXXX-XXXXX",
"rid": "#s_R668-XXXX-XX-XXXX",
"_relation_took": "#dd_300 mg xxxxxxxx yyy",
"sex": "F",
"age_range": ">=18-<40", "finished_treatment": false, "is_early_withdrawal_reason": "OTHER", "is_early_withdrawal_reason_other": "OTHER: PREGNANCY", "is_early_withdrawal_ae": false, "is_early_withdrawal_efficacy": false, "is_early_withdrawal_other": true, "is_early_withdrawal": true, "_relation_first_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_0", "_relation_last_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_16", "_relation_suffered": ["#sae_XXXX-XX-XXXX-XXXXXXXXX_1", "#sae_XXXX-XX-XXXX-XXXXXXXXX_2", "#sae_XXXX-XX-XXXX-XXXXXXXXX_3", "#sae_XXXX-XX-XXXX-XXXXXXXXX_4", "#sae_XXXX-XX-XXXX-XXXXXXXXX_5"], "_relation_randomized": "#dd_300 mg xxxxxxxx yyy" }, { "id": "XXXX-XX-XXXX-XXXXXXXXX", "rid": "#s_XXXX-XX-XXXX-XXXXXXXXX", "_relation_took": "#dd_300 mg xxxxxxxx yyy", "sex": "M", "age_range": ">=18-<40", "finished_treatment": true, "is_early_withdrawal_reason": null, "is_early_withdrawal_reason_other": null, "is_early_withdrawal_ae": false, "is_early_withdrawal_efficacy": false, "is_early_withdrawal_other": false, "is_early_withdrawal": false, "_relation_first_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_0", "_relation_last_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_35", "_relation_suffered": ["#sae_XXXX-XX-XXXX-XXXXXXXXX_1", "#sae_XXXX-XX-XXXX-XXXXXXXXX_2", "#sae_XXXX-XX-XXXX-XXXXXXXXX_3", "#sae_XXXX-XX-XXXX-XXXXXXXXX_4", "#sae_XXXX-XX-XXXX-XXXXXXXXX_5", "#sae_XXXX-XX-XXXX-XXXXXXXXX_6"], "_relation_randomized": "#dd_300 mg dupilumab q4w" }, { "id": "XXXX-XX-XXXX-XXXXXXXXX", "rid": "#s_XXXX-XX-XXXX-XXXXXXXXX", "_relation_took": "#dd_Placebo", "sex": "M", "age_range": ">=40-<65", "finished_treatment": true, "is_early_withdrawal_reason": null, "is_early_withdrawal_reason_other": null, "is_early_withdrawal_ae": false, "is_early_withdrawal_efficacy": false, "is_early_withdrawal_other": false, "is_early_withdrawal": false, "_relation_first_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_0", "_relation_last_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_34", "_relation_suffered": ["#sae_XXXX-XX-XXXX-XXXXXXXXX_1", "#sae_XXXX-XX-XXXX-XXXXXXXXX_2", "#sae_XXXX-XX-XXXX-XXXXXXXXX_3", "#sae_XXXX-XX-XXXX-XXXXXXXXX_4", "#sae_XXXX-XX-XXXX-XXXXXXXXX_5"], "_relation_randomized": "#dd_Placebo" }, { "id": "XXXX-XX-XXXX-XXXXXXXXX", "rid": "#s_XXXX-XX-XXXX-XXXXXXXXX", "_relation_took": "#dd_300 mg dupilumab q8w", "sex": "F", "age_range": ">=40-<65",
"finished_treatment": true,
"is_early_withdrawal_reason": null,
"is_early_withdrawal_reason_other": null,
"is_early_withdrawal_ae": false,
"is_early_withdrawal_efficacy": false,
"is_early_withdrawal_other": false,
"is_early_withdrawal": false,
"_relation_first_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_3",
"_relation_last_exposure": "#se_XXXX-XX-XXXX-XXXXXXXXX_33",

...and automatically transform it into a narrative like this:

CSR Section 12

EXTENT OF EXPOSURE

The majority of subjects were exposed to the IMP for the whole treatment period. 11 subjects withdrew from the study during treatment period. 6 subjects in the Placebo group (after 7, 8, 27, 11, 19, and 19 weeks respectively), 4 subjects in the 300 mg dupilumab q4w group (after 16, 2, 17, and 24 weeks respectively), 1 subject in the 300 mg dupilumab q8w group (after 31 weeks), 0 subjects in the 300 mg dupilumab q2w group, and 0 subjects in the 300 mg dupilumab qw group.

ADVERSE EVENTS

Brief summary of adverse events

All subjects in Placebo and dupilumab groups experimented TEAE. The percentage of subjects with a TEAE suspected to be related to the treatment were similar in Placebo and dupilumab groups. Overall, 6 out of 19 subjects (31.6%) in Placebo group and 17 out of 45 subjects (37.8%) in dupilumab groups reported at least one TEAE suspected to be related to the treatment during the study.

Nine (9) subjects (5 receiving Placebo, 3 receiving 300 mg dupilumab q4w, 1 receiving 300 mg dupilumab q8w, 0 receiving 300 mg dupilumab q2w, and 0 receiving 300 mg dupilumab qw) permanently discontinued IMP secondary to TEAE.

During the study, there were 23 severe TEAEs reported, affecting 14 subjects, and 24 treatment emergent SAEs, affecting 18 subjects. There was 1 death, although it was not related to the treatment.

The most frequently reported TEAEs by primary SOC were in the Infections and infestations SOC, followed by the Skin and subcutaneous tissue disorders SOC. The most frequently reported TEAE by PT were Dermatitis atopic and Nasopharyngitis. The incidence of individual PTs between the dupilumab and Placebo groups were:

- Dermatitis atopic: 19 (100%) in Placebo group and 35 (77.8%) in dupilumab groups (15 [100%] subjects in the 300 mg dupilumab q4w, 8 [80%] subjects in the 300 mg dupilumab q8w, 7 [63.6%] subjects in the 300 mg dupilumab q2w, and 5 [55.6%] subjects in the 300 mg dupilumab qw)
- Nasopharyngitis: 4 (21.1%) in Placebo group and 12 (26.7%) in dupilumab groups (5 [55.6%] subjects in the 300 mg dupilumab qw, 3 [20%] subjects in the 300 mg dupilumab q4w, 2 [20%] subjects in the 300 mg dupilumab

AI-Assisted Medical Writer and Insights Miner

Data to Text Writing

Automatic Generation of:

- CSR
- Individual Patient Narratives
- Other documents

Text-to-Text Writing

Verb-tense update (Protocol to CSR)

Automatic extraction of bibliographic data

Tagging of medical documentation

Topic extraction and classification

Summarization

QA

Automatic consistency check of data and text within documents.

Verification of correct use of medical terms

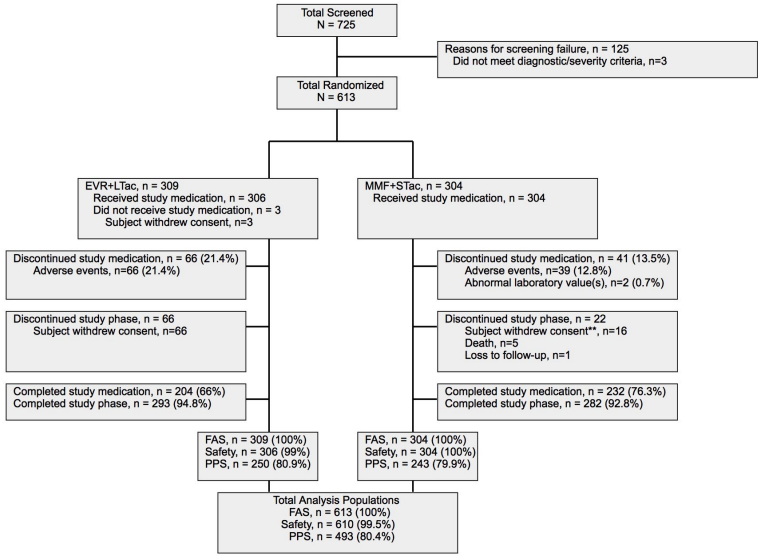
Document format verification

AI-Assisted Authoring that Transforms Data into Narratives


```
{
 "disposition": {
  "total_screened": 725,
  "total_randomized": 613,
  "screening_failure": [
   {
    "reason": "Did not meet diagnostic/severity criteria",
    "n": 3
   }
  ],
  "fas": 613,
  "safety": 610,
  "pps": 493
 },
 "drugs": [
  {
   "drug_name": "Everolimus Plus Low-Dose Tacrolimus",
   "drug_acronym": "EVR+LTac",
   "total_population": 309,
   "overall_incidences_aes": 303,
   "aes_suspected_drug_related": 225,
   "severe_aes": 155,
   "adverse_events": [
    {
     "name": "Leukopenia",
     "value": 27,
     "type": "other",
     "is_severe": false
    },
    {
     "name": "Diarrhea",
     "value": 76,
     "type": "other",
     "is_severe": false
    },
```


Patient disposition

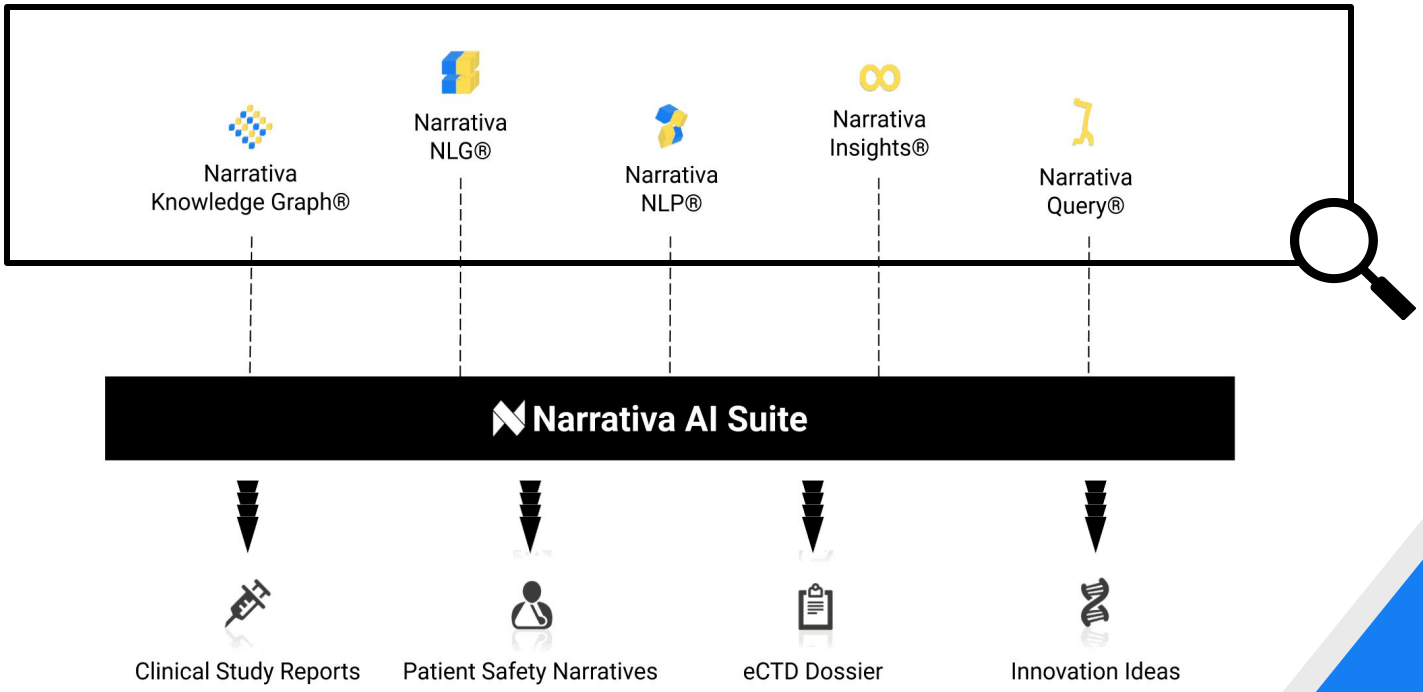
In total, 613 patients were randomized to EVR+LTac (N = 309) or MMF+STac (N = 304) groups. Of these, 575 (EVR+LTac: 293 [94.8%]; MMF+STac: 282 [92.8%]) completed the study. A total of 107 patients (EVR+LTac: 66 [21.4%]; MMF+STac: 41 [13.5%]) discontinued study medication, with the most common reason being Adverse events (EVR+LTac: 66 [21.4%]; MMF+STac: 39 [12.8%]; Figure 1)



Safety

The EVR+LTac and MMF+STac groups had both similar levels of total AEs, and AEs believed to be related with the investigational medicines. Hyperlipidemia was significantly ($p<0.05$) higher with EVR+LTac versus MMF+STac and Leukopenia and Diarrhea were significantly ($p<0.05$) higher with MMF+STac versus EVR+LTac. The proportion of patients reporting SAEs was 50.2% and 46.7% ($p = 0.45$) in EVR+LTac and MMF+STac groups respectively. Peripheral edema was the most commonly reported SAE (EVR+LTac 37.5% and MMF+STac 29.6%; $p = 0.07$).

Our Technology and Use Cases



How Our NLG Solution Works

Input

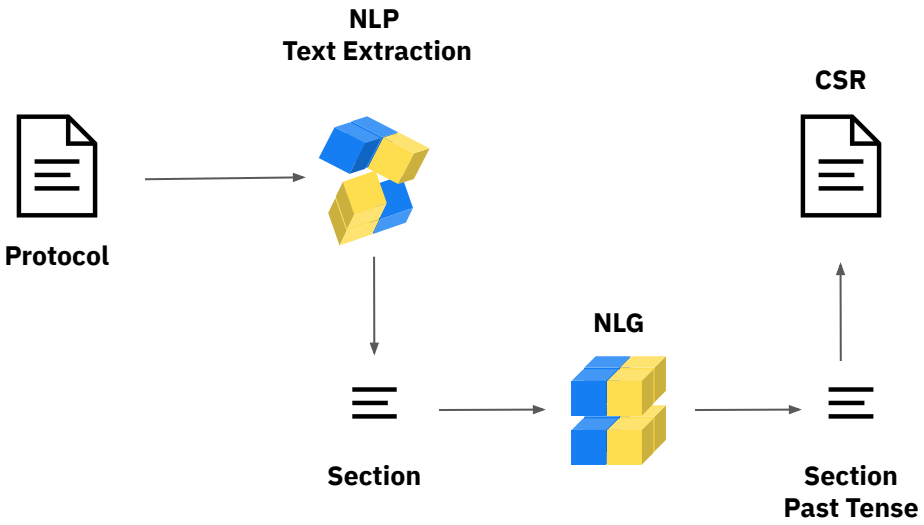
- Medical Trial Data
- Business Rules/Text Structure/Knowledge Graph
- Corpus of Examples

Output

- Medical Trial Content API

Automated Generation of Clinical Study Reports (CSRs)

CSR Automation (Sections 1-9)

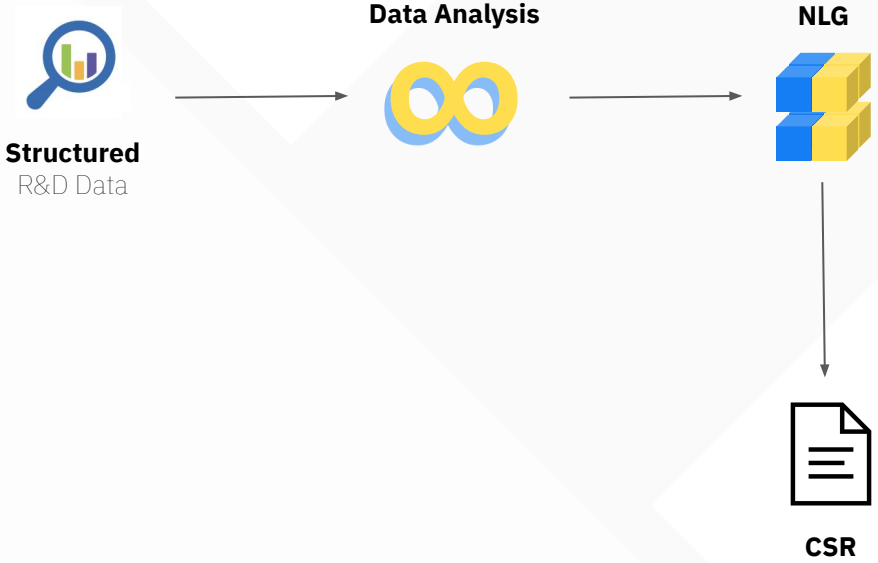


- Protocol specific sections are extracted
- Verb tense is updated
- Text is added to the CSR
- Full 1st Draft auto generated

Demo: Verb Tense Update

Book Demo

CSR Automation (Section 10)

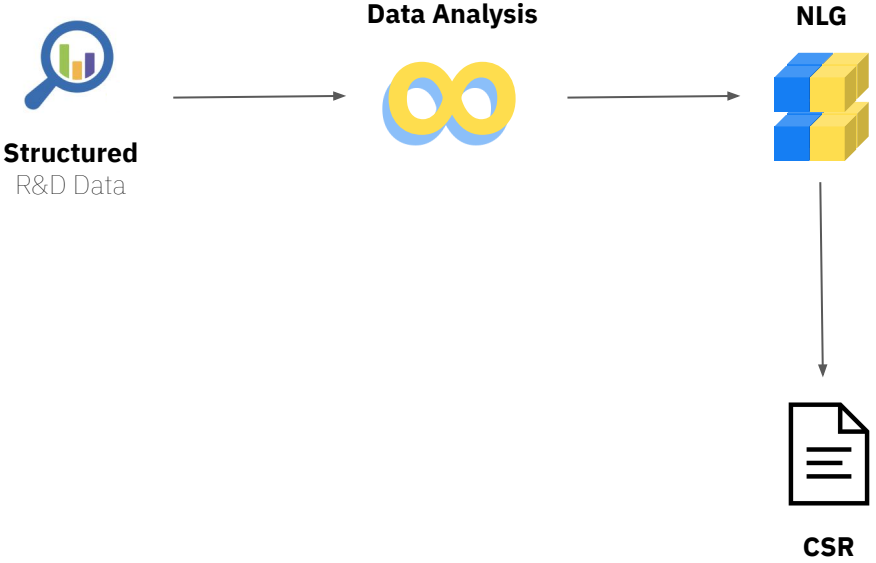


- Data is extracted and analyzed
- Text is generated
- Text is added to the CSR
- Full 1st Draft auto generated

Demo: Patient Disposition

Book Demo

CSR Automation (Sections 11-12)

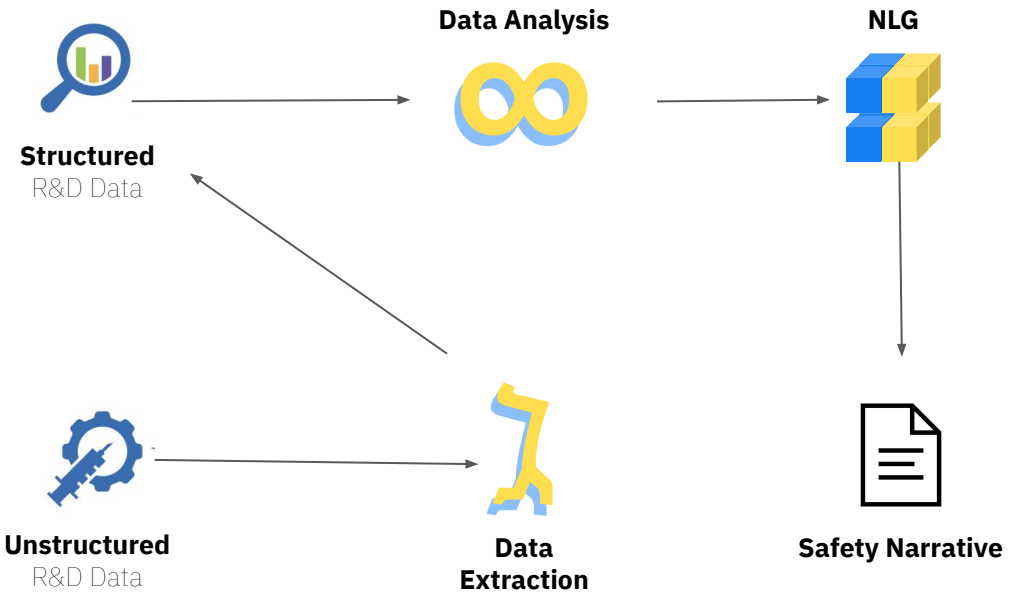


- Data is extracted and analyzed
- Text is generated
- Text is added to the CSR
- Section 11: Partial draft auto generated
- Section 12: Full 1st Draft auto generated

Book Demo

Safety Narratives

CSR Automation (Patient Safety Narratives)



- Source data sets are analyzed, and applicable datasets are extracted
- Structured Data is analyzed and statistical models are used (if required).
- Text is generated using the analyzed structured data.
- Full 1st Draft auto generated

Book Demo

Word Plugin
Demo Video

eCTD Dossier Automation Solution

eCTD Automation Solution

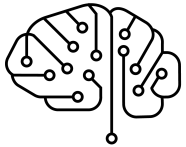
The goal of the CSR NLG Solution is to **automatically generate** different regulatory reports part of the eCTD.

Study Protocol/ SAR

Corpus Reports

- Medical Trials Data (from databases, xml, json, sas7dbat, rtf, etc.)
- Corpus of CSR to extract implicit knowledge, structured and writing style.

NLG Automation Platform



AI

- Extraction and normalization of data.
- Knowledge Graph (MedDRA, ICH guidelines, etc)
- Automatic Generation Report

eCTD

- Module 1: SmPC
- Module 2: Quality, Non-Clinical Study Reports and Clinical Study Reports
- Module 3: Quality
- Module 4: Toxicity analysis, ADME, etc.
- Module 5: Protocol, Efficacy, Safety, etc.

Other Automation Solutions

Other Automation Solutions

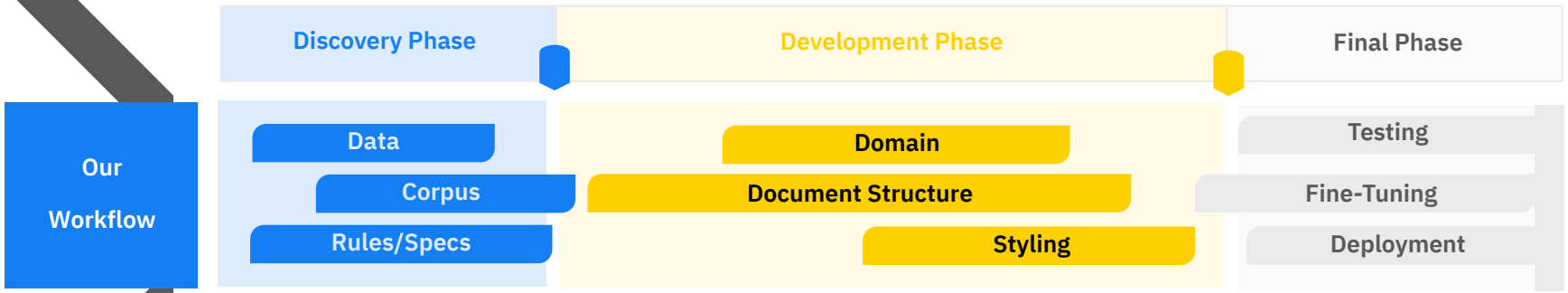
Summarization and Tagging
Public Health Data Reports
Anonymization
IT Auditing Reports for compliance obligations
GMP Reports in Drug Manufacturing

Extraction of Social Media Insights

Automation of communication with the public to survey trends in the use of medication and home remedies.

Our Approach to Service Delivery

Our Approach and Timelines



Pilot for 1 CSR Type (e.g. Section 10 or 12) **Timeline: 4-5 Weeks***

Development & Deployment **Timeline: 3-4 Months***

*Depending on therapeutic area.

Quality Control

- **Discovery: Definition of the scenarios to cover.**
- **Development: During the NLG model adaptation. Narrativa QA team to use synthetic data to test each scenario and check the generated text:**
  - **Contains all required information accurately.**
  - **Does not contain incorrect or not fact-based statements.**
- **Final Phase: Client's acceptance tests (UATs). Client's team to use real or synthetic data to validate the solution.**
- **Deficiencies: Any and all deficiencies in each phase will be fixed by Narrativa's automation team and retested by the corresponding team**

40%

Expected cost reduction on medical writing

65%

Reduction on time spent on documentation generation

We look forward to changing the future with you.

Jennifer Bittinger, President
jennifer.bittinger@narrativa.com
730 Arizona Avenue
Santa Monica, CA 90401
www.narrativa.com