



---

# Module 12

# TEAMWARE

**A Collaborative,  
Web-based Annotation Environment**

---



# Hands-on Preparation

---

- Open the Teamware URL in your browser.
  - Log on with the provided user name and password.
  - Click on the link “Annotation Editor” to download and prepare the software for our first hands on. (This will get all the jars into your Java Web Start cache.)
  - When it opens, leave it as is, till we need it
  - [Click here](#) for more information about JWS
-



# Outline

---

- Why Teamware?
- What's Teamware?
- Teamware for annotation
- Teamware for quality assurance and curation
- Teamware for defining workflows, running automatic services, managing annotation projects
- Outlook

# From Annotation Tools to Collaborative Annotation Workflows

GATE

We have lots and lots of tools and algorithms for annotation; what we need is

1. methodological instead of purely technological
2. multi-role instead of single role
3. assistive instead of autonomous
4. service-orientated, not monolithic
5. usable by non-specialists

## GATE Teamware

- Research users in several EU projects
- External users at IRF and Matrixware
- Interest from other commercial users as well

# GATE Teamware: Annotation Workflows on the Web

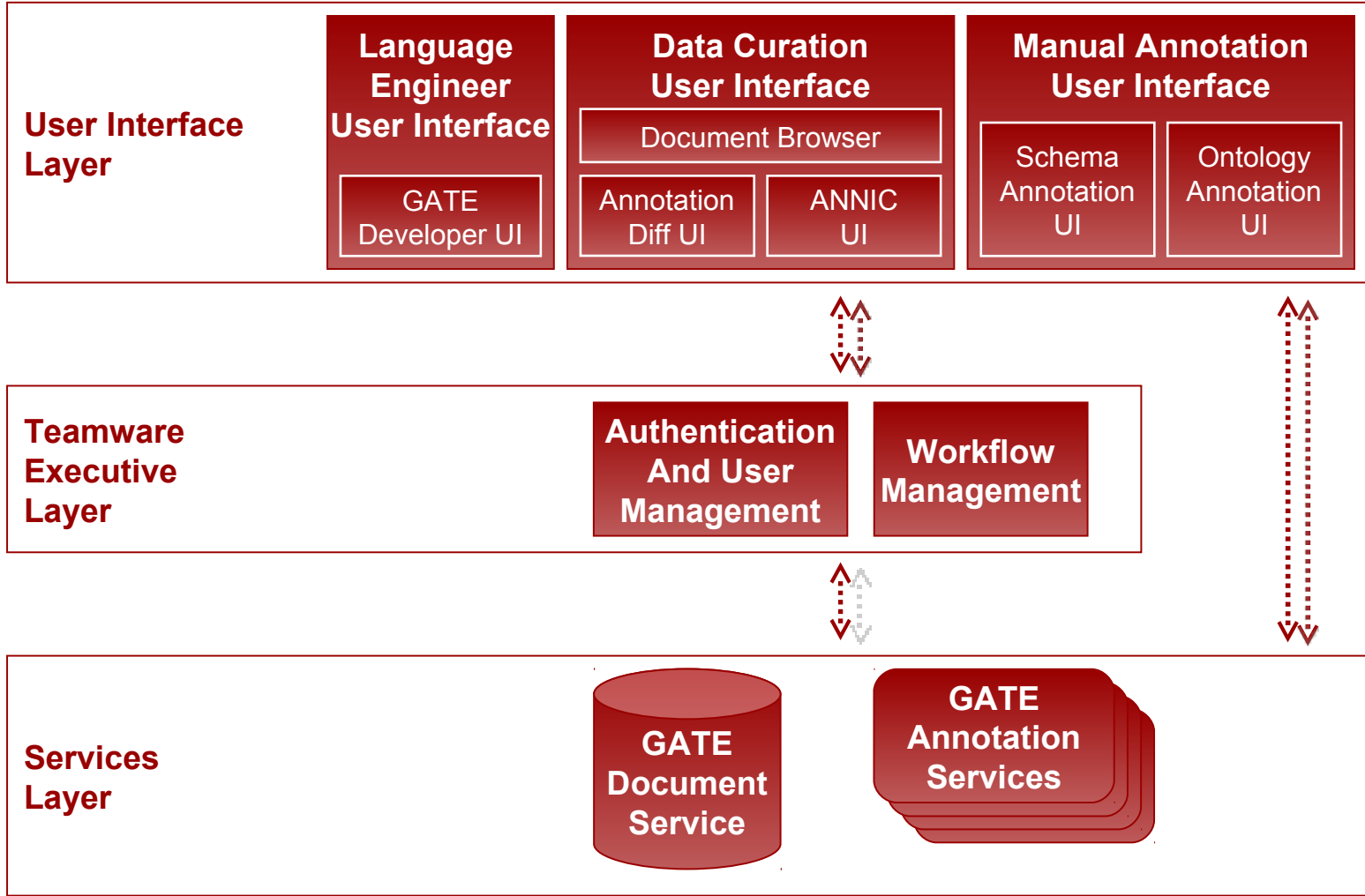
The GATE logo consists of the word "GATE" in a bold, red, sans-serif font. The letters are contained within a green, rounded rectangular border that has a slight 3D effect, with the top and bottom edges being slightly thicker than the sides.

GATE Teamware is:

- Collaborative, social, Web 2.0
- Parallel and distributed (using web services)
- Scalable (via service replication)
- Workflow based with business process integration



# Teamware - Layer Cake



# Division of Labour: A Multi-role Methodology

---



- **Annotators**
    - carries out manual annotation: from scratch or (faster) by correcting automatic annotation (generated by rule-based tools, machine learning, or a combination)
  - **Manager**
    - Defining **annotation guidelines and schemas**
    - **Choose relevant automatic services to pre-process**
    - Toolset including **performance benchmarking, progress monitoring tools, small linguistic customisations**
    - Define **workflow, manage annotators**, liaise with language engineers and sys admins
    - **Reconcile differences** between annotators, using IAA, AnnDiff
  - **Admin**
    - Setup the Teamware system, users, etc.
  - **Language engineer**
    - Uses **GATE Developer** to create bespoke services and deploy online
-

The screenshot displays the GATE Annotator GUI, which is connected to POOL mode: kalina. The interface includes a menu bar (File, View, History, Delicious, Tools, Help), a toolbar with icons for file operations and help, and a main document editor window.

The document editor shows a table of annotations with the following columns: Type, Set, Start, End, Id, and Features. The table contains the following data:

Type	Set	Start	End	Id	Features
Organization		4	19	159	{}
Date		23	31	170	{kind= date, rule1= GazDate, rule2= DateOnlyFinal}
Percent		104	119	157	{}
Percent		159	169	169	{rule= PercentBasic}
Date		212	222	138	{kind= date}
Organization		276	297	145	{orgType= unknown, rule1= OrgXBase, rule2= OrgFinal}
Organization		344	359	139	{}

Below the table, the document text is displayed with annotations. The text includes: "The Bank of England on Thursday surprised the financial markets and delighted British businesses with a 0.25 percentage point cut in its main interest rate to 5 per cent. Interest rates have not been lower since early 1964. But the mood inspired by the cut was marred by the European Central Bank's refusal to cut rates, and by concern at the Bank of England's growing pessimism about the global outlook. Forecasters had expected the Bank's monetary policy committee to leave rates unchanged, but fears more cuts would fuel consumer demand and threaten higher inflation. But the MPC said the unexpected severity of the global slowdown means that inflation on track to hit its 2.5 per cent target in two years. "Monetary policy needs to balance the weaker external environment but growth," it said. The MPC believes a two-speed economy of weak international demand and strong consumer demand is better than an economy in recession. Since last month's MPC meeting, which voted eight to one to leave rates unchanged, signs domestic demand is cooling. Retail sales, household borrowing and car sales have all fallen." The text is annotated with various tags such as Organization, Date, Percent, and Person.

An "Annotation Editor Dialog" is open, showing a list of annotation types (Address, Date, Location, Money, Organization, Percent, Person, Phone) and a "Features" section. The dialog also includes checkboxes for "Case", "Regex", "Whole", and "Highlights", and a text input field containing "Bank of England".

The bottom status bar shows: Task: 162 Document: ft-bank-of-england-02-aug-2001.xml





# Manual Annotation

---

- Annotator: log into Teamware
  - Click on “Open Annotation Editor”
  - Request an annotation task (first button)
    - Annotate the assigned document
    - Save work in progress occasionally
    - Click on the “Finish task” button when done
  - If you want to save work and return to this task later – “Save” button, then close the UI. Next time you request a task, you get the same document back to finish it.
  - Depending on the project setup, you may be able to reject a document and then ask for another one to annotate (Reject button)
-

# Hands-on



- 
- Open a web browser and Teamware
  - Login using your annotator user name
  - Open the annotation UI
  - Try requesting tasks, editing annotations, saving your work, asking for another task
    - You need to annotate names of people and organisations
  - This is what Teamware looks like for manual annotation tasks
-

# Teamware for Managers

---



- Adjudication and measuring IAA
- Defining workflows
- Running annotation projects
- Tracking progress



# Adjudication

---

- Only available to the manager role
  - Identify differences between annotators using IAA
  - Inspect differences in detail using AnnDiff
  - Edit and reconcile differences if required
  - More sophisticated adjudication UI (the Annotation Stack View) in GATE Developer
-



# IAA: Recap

---

- The IAA on IE tasks, such as named entity recognition, should be measured using f-measure across all annotators
- For classification tasks, use Kappa to measure IAA
- For details, see Module 2 slides and the GATE user guide

# IAA: Do my annotators agree?



[Home](#)
[Resources](#) →

[Projects](#) →

[Support](#) →

## IAA Caculation on the document ft-airtours-08-aug-2001.xml

Please note you can NOT do caculation if there is no common annotation type on the two selected annotation sets or more.

Available Annotation Sets

Key  
Original mark  
safe.preproce  
<Default>



Annotation Sets to Compare

annotator1  
annotator2  
annotator3



Annotation Type

Date

Algorithm

all-ways-f-measure

Feature

Caculate

Cancel

# IAA: Results

## IAA Calculation on the document ft-airtours-08-aug-2001.xml

Please note you can NOT do calculation if there is no common annotation type on the two selected annotation sets or more.

Available Annotation Sets

Key  
annotator2  
Original mark  
annotator1  
annotator3  
safe.preproc  
<Default>



Annotation Sets to Compare



Annotation Type

Date

Algorithm

all-ways-f-measure

Feature

Calculate

Cancel

### Overall F-Measure

Precision	Recall	F1	Precision Lenient	Recall Lenient	F1 Lenient	Correct	Partially Correct	Missing	Spurious
1.0	0.9411765	0.96875	1.0	0.9411765	0.96875	16.0	0.0	1.0	0.0

### F-Measure values against the Key annotator1

Response	SP / SR / SF	LP / LR / LF	C / P / M / S
<a href="#">annotator2</a>	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	17.0 / 0.0 / 0.0 / 0.0
<a href="#">annotator3</a>	1.0 / 0.88235295 / 0.9375	1.0 / 0.88235295 / 0.9375	15.0 / 0.0 / 2.0 / 0.0

# AnnDiff: Finding the differences



Annotation Diff GUI [Connected in DIRECT mode: ft- airtours-08- aug-2001.xml\_\_1248214749368\_\_610]

Annotation Diff GUI

Document: ft- airtours-08- aug-2001.xml | Annotation Set: annotator1 | Annotation Type: Date | F-Measure Weight: 1.00

Response: ft- airtours-08- aug-2001.xml | Annotation Set: annotator3 | Features:  All  Some  None

Start	End	Key	Features	=?	Start	End	Response
123	136	third quarter	{kind= date}	-?			
180	219	first nine months of its financial year	{kind= date}	-?			
2473	2484	summer 2001	{kind= date, rule1= DateName, rule2= DateOnlyFinal}	=	2473	2484	summer 2001 {kind= date, rule1=
2121	2130	this year	{kind= date}	=	2121	2130	this year {kind= date}
237	255	quarter to June 30	{kind= date}	=	237	255	quarter to June 30 {kind= date}
1204	1220	end of this year	{kind= date}	=	1204	1220	end of this year {kind= date}
2797	2824	second quarter of next year	{}	=	2797	2824	second quarter of next year {}
2409	2418	last year	{kind= date}	=	2409	2418	last year {kind= date}
2270	2281	summer 2001	{kind= date, rule1= DateName, rule2= DateOnlyFinal}	=	2270	2281	summer 2001 {kind= date, rule1=
337	346	last year	{kind= date}	=	337	346	last year {kind= date}
2167	2176	next year	{kind= date}	=	2167	2176	next year {kind= date}
1560	1584	nine months of this year	{kind= date}	=	1560	1584	nine months of this year {kind= date}
2326	2332	winter	{kind= date, rule1= GazDate, rule2= DateOnlyFinal}	=	2326	2332	winter {kind= date, rule1=

Correct: 15      Recall   Precision   F-Measure  

Partially Correct: 0      Strict: 0.8824 1.00 0.9375

Missing: 2      Lenient: 0.8824 1.00 0.9375

False Positives: 0      Average: 0.8824 1.00 0.9375

Document: ft- airtours-08- aug-2001.xml\_\_1248214749368\_\_610



# Where are these in Teamware?

---

- Only visible to managers
  - Resources/Documents menu
  - Select the corpus worked on
  - Iterate through each document
  - Run IAA and AnnDiff, as required
  - These are clumsy, as are on a document by document basis
-

# GATE Developer adjudication facilities

---



- Download the corpus as a zip file
  - Corpus QA Tool
    - A corpus-level view of IAA (F-measure or Kappa)
  - Extended AnnDiff to allow easy reconciliation of the differences between 2 annotators
  - Annotation Stack View to reconcile difference visually, of 2 or more annotators
-



# Adjudication in AnnDiff

**Annotation Diff Tool**

Key doc: ft-pirelli-10-aug-2... Key set: annotator1 Type: Date Weight: 1.0

Resp. doc: ft-pirelli-10-aug-2... Resp. set: annotator2 Features:  all  some  none 1.0 Compare

Start	End	Key	Features	K=?	R	Start	End	Response
590	597	July · 28	{kind= date, rule1= Da...rule2= DateOnlyFinal}	<input type="checkbox"/> =	<input checked="" type="checkbox"/>	590	597	July · 28 {kind= date, r
1313	1321	Thursday	{kind= date, rule1= Ga...rule2= DateOnlyFinal}	<input type="checkbox"/> =	<input checked="" type="checkbox"/>	1313	1321	Thursday {kind= date, r
1262	1266	2002	{kind= date, rule1= Ye...rule2= DateOnlyFinal}	<input type="checkbox"/> ~	<input checked="" type="checkbox"/>	1254	1275	between · 2002 · and · 2004 {kind= date, r
1271	1275	2004	{}	<input type="checkbox"/> -?	<input type="checkbox"/>			

Target set: consensus Copy selection to target set

2 documents loaded Show document Export to HTML

Statistics Adjudication

# Adjudication in AnnDiff



- 
- Select the annotations which are correct by ticking the boxes (see previous screen shot)
  - Provide the name of the target consensus set
  - Click on the button to copy them into that set
  - Once copied, they are removed from the list of annotations to adjudicate, so the curator can focus on the remaining ones
  - Adjudication works one annotation type at a time and only for 2 annotators, whose results are stored into 2 different annotation sets
-



# Annotation Stack View

Annotation Sets | Annotations List | Annotations Stack | Co-reference Editor | OAT | Text |

But when the Telecom Italia payment of E15.5m was proposed last week, its board rejected it. Pirelli and its ally, the Benetton family, gained control on **July 28** by buying 23 per cent of Olivetti from Bell, an investment group headed by Mr Colaninno and others. The E31m payouts would be in lieu of options and any other compensation due to him. Mr Colaninno was also chief executive of Olivetti.

Several board members objected that Telecom Italia shares had not risen since Mr Colaninno took control and that the figure was too high.

Mr Colaninno had a salary last year of E3.5m from Telecom Italia and E819,000 from Olivetti. He also had cash bonuses of E3.6m from Telecom Italia and E700,000 from Olivetti and other companies controlled by Olivetti.

In addition, he received options to buy 12.6m Olivetti shares **between 2002 and 2004** at E2.81 a share. Olivetti shares on **Thursday** closed at E1.94 compared with the E4.17 – an 80 per cent premium to the market – paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

Pirelli is expected to bring in a new board this autumn but several members are likely to object again to the payment. The Olivetti board might approve its E15.5m payout but minority shareholders in both companies may

Previous boundary | Next boundary

Context	6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thu
annotator1#Date	<div style="display: inline-block; width: 20px; height: 15px; background-color: #f4a460; margin-right: 10px;"></div> <div style="display: inline-block; width: 20px; height: 15px; background-color: #f4a460; margin-right: 10px;"></div> <div style="display: inline-block; width: 20px; height: 15px; background-color: #f4a460;"></div>
annotator2#Date	<div style="display: inline-block; width: 150px; height: 15px; background-color: #f4a460; margin-right: 10px;"></div> <div style="display: inline-block; width: 20px; height: 15px; background-color: #f4a460;"></div>

Original markups

- ▼ annotator1
  - Date
  - Money
  - Organization
  - Percent
  - Person
- ▼ annotator2
  - Date
  - Money
  - Organization
  - Percent
  - Person

# Annotation Stack View

---



- Open the Document Viewer
  - Select the “Annotations Stack” button
  - Select the annotation types you’d like to reconcile, from as many annotation sets as needed
  - The Stack view window is at the bottom
  - Context shows the part of the document which we are working on now
-



# Annotation Stack View

Mr Colaninno had a salary last year of E3.5m from Telecom Italia and E819,000 from Olivetti. He also had cash bonuses of E3.6m from Telecom Italia and E700,000 from Olivetti and other companies controlled by Olivetti.

In addition, he received options to buy 12.6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thursday closed at E1.94 compared with the E4.17 - an 80 per cent premium to the market - paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

- Date
- Money
- Organization
- Percent
- Person

Copy annotation to another set

- Percent
- Person
- consensus
- Date

In addition, he received options to buy 12.6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thursday closed at E1.94 compared with the E4.17 - an 80 per cent premium to the market - paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

Previous boundary

Context

annotator1: Previous boundary Next boundary Target set: consensus

annotator2: Context 6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thu

annotator1#Date			
annotator2#Date			
consensus#Date			

# Annotation Stack View



- 
- The Previous/Next boundary buttons position the text we're working on (the Context) on the previous/next annotation of the chosen type
  - Note that such annotation may only exist in one of the sets, i.e., the other annotator might have missed it!
  - Hover the mouse over the coloured annotation rectangles to see further details
  - Right click would open the annotation editor
  - Double click copies the annotation to the target consensus set (you choose which one it is)
-



# Teamware Workflows



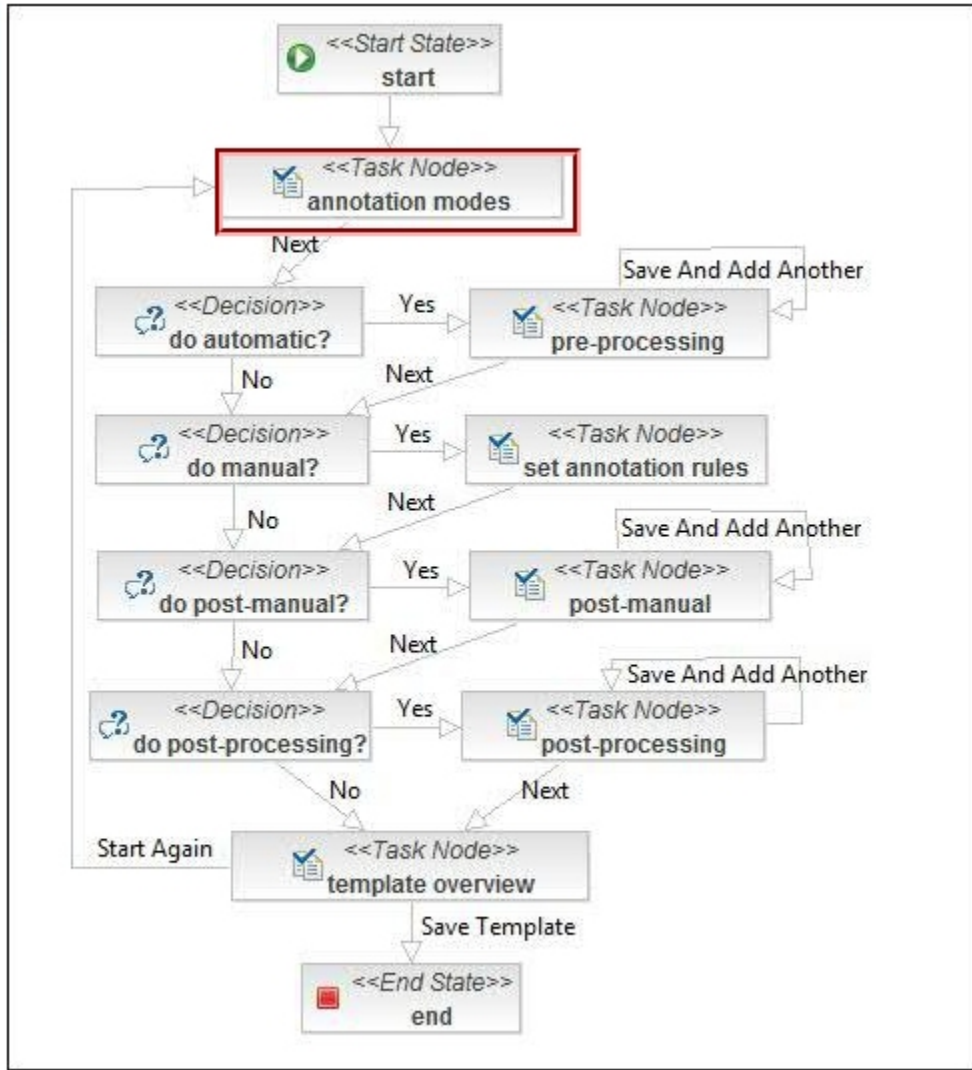
- 
- Whole process is controlled by a **workflow manager**
  - Workflow may be simple:
    - Give the document to  $n$  human annotators
  - or more complex
    - Invoke one or more web services to produce automatic annotations
    - Pass each document to 2 annotators



# Workflow Templates

Automatic Annotation:	<input type="checkbox"/>
Manual Annotation:	<input type="checkbox"/>
Post-Manual:	<input type="checkbox"/>
Review:	<input type="checkbox"/>
Post-Processing:	<input type="checkbox"/>

Next Help Quit



# Defining new workflows

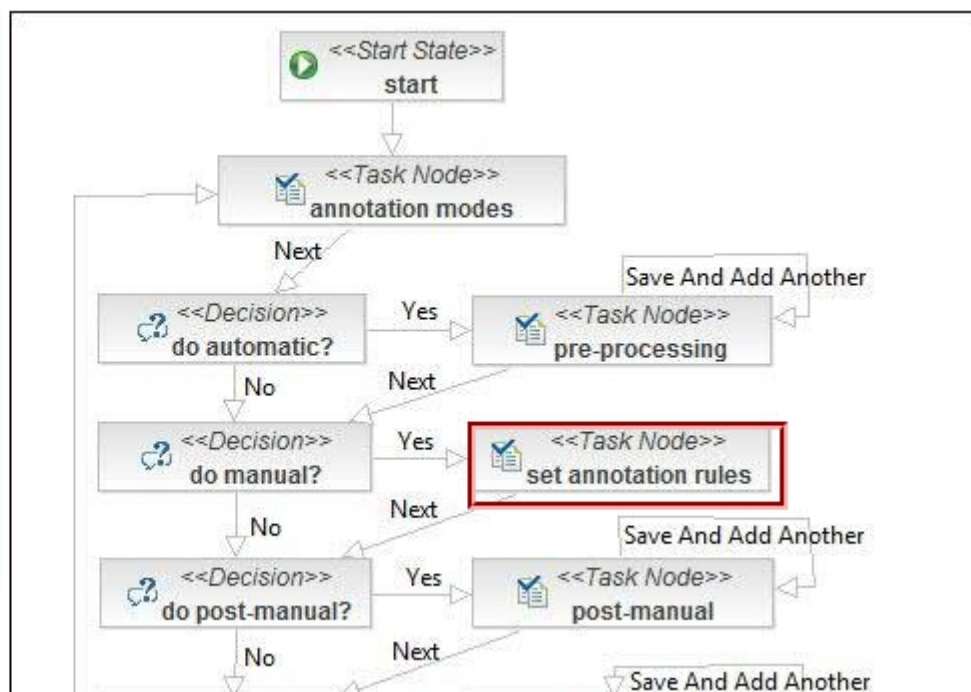


- 
- Select Projects/WF Templates
  - Opens the WF wizard
  - Automatic annotation:
    - Runs one or more web services to pre-annotate
    - These need to be GATE Annotation Services (GAS), not just any service
    - The default Teamware installation includes ANNIE
  - Manual annotation
  - Post-manual: post-processing/merging service
  - Post-processing: Finalise and export corpus

# Defining Manual Annotation WF

- Select only the Manual Annotation box in the WF template
- Configure further the number of annotators per document, whether they can reject documents, whether annotation sets are anonymous (annotator1, annotator2, ...)
- Select or upload annotation schemas

Annotators per Document:	2
Cancel Task Allowed:	<input checked="" type="checkbox"/>
Anonymous Annotation:	<input checked="" type="checkbox"/>
Annotation Schemas:*	<ul style="list-style-type: none"> <li>OrganizationSchema.xml</li> <li>PersonSchema.xml</li> <li>AddressSchema.xml</li> <li>Non-FunctionalRequirem</li> <li>ConcordiaDataGroupsSc</li> <li>ConcordiaRequirementsS</li> </ul>
Add Schema:	<a href="#">Add Schema</a>
Pre-Manual Service:*	None



# Annotation Schemas



- 
- GATE's annotation schemas can be used to constrain manual annotation for consistency
  - You can use the standard ANNIE schemas or develop your own
  - Schemas should be consistent with your annotation guidelines
-



# Annotation Schemas

- Define types of annotations and restrict annotators to use specific feature-values
  - Person
  - Person.gender = male
  - Person.gender = female
- Uses the XML Schema language supported by W3C for these definitions

```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
  <element name="Person">
    <complexType>
      <attribute name="gender" use="optional">
        <simpleType>
          <restriction base="string">
            <enumeration value="male"/>
            <enumeration value="female"/>
          </restriction>
        </simpleType> ...
```



## Running a Manual Annotation WF

---

- The WF template defined in the previous steps can be saved and the project started or revised, as necessary
  - To actually run a manual annotation WF, you need to create a project, which defines:
    - Which corpus is going to be annotated
    - Who are the managers and annotators
  - The corpus can be uploaded via the link provided and is a ZIP file of documents
    - can be in any format supported by GATE
-



# Running a Manual Annotation WF

<b>Project Name *</b> Manual WF project	<b>Managers</b> kalina	<b>Curators</b> ac4sa59 agaton ishrar-cur-man kalina kalina-curator matthew-cur-man milan	<b>Annotators</b> ac4sa59 adam agaton angus ayrin-ann danica diana hamish ian ishrar-ann
--	---------------------------	--	--

[Add Corpus](#)

Save & Start

My Workflow Templates

- Once project is started, annotators can login and ask for tasks





# Hands-On

---

- Login as a manager
  - Schemas:
    - In the resources directory you will find 2 annotation schemas: Person and Organization
    - Edit the Person one to include optional gender
    - To avoid clashes with other people's please rename these to <your-name>Organization and <your-name>Person
  - Define your own manual WF with your schemas
  - Start a project with your WF template
    - Choose yourself as a manager and add some annotators assigned to the project
    - For corpus: copy and save some text from a news site like the bbc (no more than 3-4 paragraphs), zip the saved document and upload it as your corpus
-



# Hands-On (Continued)

---

- Log out as manager and login as annotator
  - Open the annotation UI
  - You should now receive tasks from the projects of two other people, 2 documents in total (1 from each project)
    - If you have put more than 1 document in your corpus, please log back in as a manager and delete them from the corpus!
  - Annotate these documents for Person and Organization names
  - Make a note of any issues you had when annotating and let's discuss them
-



# Monitoring Project Progress

## Process Monitoring: Annotation Status

Detailed View

Back to Project

Status	#
Annotated	64
Canceled	1
Failed	0
In Progress	1
Not Started	7

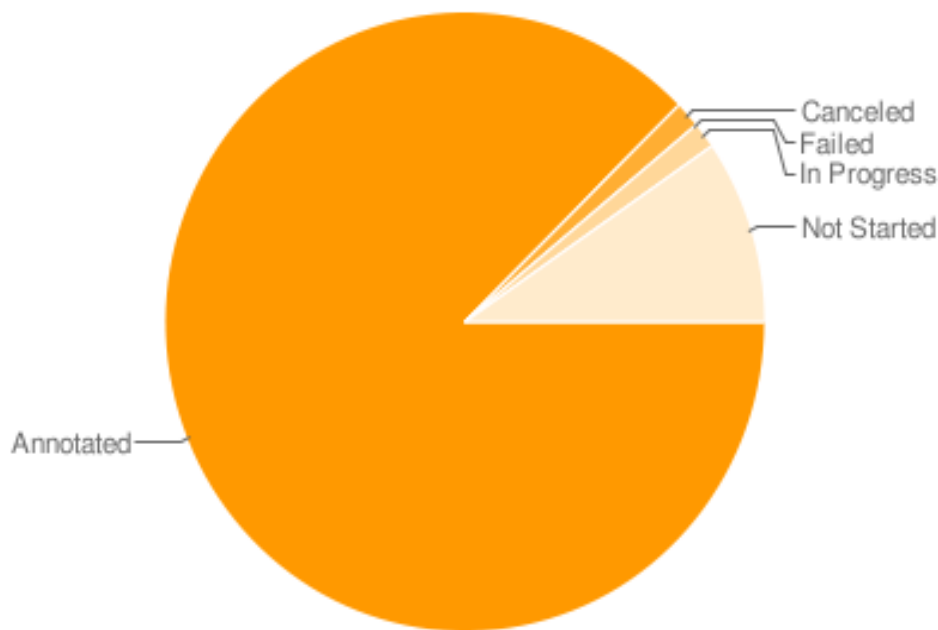
### Average Execution Time

627.609375

Detailed View

Back to Project

Annotation Status Chart



# Hands-On



- 
- Log back in as a manager
  - Go to Projects/My Projects
  - Select the Process Button
  - You will see the manual annotation task
  - Click on Monitoring to see the statistics inside
  - If both your annotators have completed their tasks, you should also receive a notification email
- 



# Setting up an Automatic Annotation Project

---



- Configure the web service(s)
- Define the Workflow template
- Run the project, choosing the corpus

# Semi-automatic Projects



- 
- Combine the automatic and manual workflow steps
  - You can preprocess the corpus with pipelines in GATE Annotation Services and then have people improve the annotations manually

# How can I use Teamware?

The logo for GATE (Global Architecture for Text Engineering) is displayed in red capital letters. It is enclosed within a green border that forms a rounded rectangle on the right side.

- 
- We are making it available as an on-line service later in the year
  - Please watch <http://gatecloud.net/> for updates or contact us for information