|  |  |
| --- | --- |
| **Slide #** | **Audio Script** |
| 1.1 | Welcome to Understanding the Recorded Future Platform. Click Next to begin. |
| 1.2 | Take a moment to review the icons on this screen to help you navigate the course. |
| 1.3 | This course is designed to give you an introduction to what Recorded Future does and how it works, as well as give you an introduction to software navigation and major features. When you’re ready, click next to continue. |
| 2.1 | In this first lesson we will discuss what Recorded Future does to give you a better understanding of what this platform is used for. |
| 2.2 | Let’s first take a look at the Recorded Future solution and how it fits into your environment. The circles on the left depict our access to open, technical and closed/dark sources - the broadest, largest set of sources of any threat intelligence company.  Examples of Open sources are: mainstream media, blogs, social media, and white hat researches. Technical sources include: databases of vulnerabilities, hashes, DNS information. Close/Dark sources are: Tor, darknet, criminal forums, special access.  The Recorded Future platform applies machine learning to structure, analyze and deliver this information as threat intelligence. The intelligence can be accessed via our own user interface where users can search, monitor and alert on the intelligence.  Recorded Future is a unique combination of machine learning and human threat analysis to automate analysis at scale and enable complex analytics.  We also have a robust API that enables access to the intelligence from third-party products including: SIEMs, or Security Information and Event Management systems,Ticketing Systems, Incident Response, Vulnerability Management, and TIPs or Threat Intelligence Platforms.  These integrations provide threat intelligence directly into the systems you are already using to secure and monitor your organization. |
| 2.3 | Recorded Future enables you, the analyst, to access Open Source Intelligence safely and without having to know exactly where to start on the internet. We harvest the web at scale, processing millions of documents per day, and structuring that data to make it searchable by topic, event, time, and/or source type, thus reducing the burden of collection and processing of data for the analyst, and leaving more time for actual analysis. Analysts conduct this research from within the secure User Interface and even have the ability to request additional collection or data processing. |
| 2.4 | There is a lot of information out there on the web, and Artificial intelligence can connect the dots… Here are some of the ways we do that. These will be explained in more detail later in the course.  Natural language processing is what transforms unstructured text… The machine processes unstructured text and extracts understanding and meaning from it by classifying into events and entities.  Classification of events and entities helps human analysts decide to perform deeper investigation… Parsing of structured text into event classifications and entities helps an analyst to research, triage, and understand large amounts of data.  Ontologies and events are representations of structured knowledge of the world…  And Predictive models from historical data helps you anticipate events and entity properties. Based on previous behavior (event) or relationship to other malicious entities...For instance, if an IP address (an entity) is in a malicious CIDR block (an ontological relationship with a larger entity), there's a greater chance that IP address could be used maliciously as well.  Essentially, Artificial Intelligence helps you to make sense of the massive amounts of data you’re dealing with from the web. All done at scale, dynamically, and in real time. |
| 2.5 | How do we do all of this? The graphic show here depicts the big picture. We will dive more into collecting, analysis, analytics, and indexing in this course in just a moment. ~~Analytics will be covered in later courses.~~ |
| 2.6 | Now that we’ve explained what we do, we’ll move on to the next lesson and understanding our data, sources, and structure. When you’re ready, click next to learn about collection. |
| 3.1 | In the second lesson we are going to discuss and understand how the Recorded Future collects, analyzes, and indexes data, sources and how the platform structures that information. |
| 3.2 | Let’s talk about the basics of how our data is structured. We source data in different languages and it is then index into entitles, events, language, source type, and time. Click on each one to learn more. |
| 3,2a | Entities are people, places, or things--the subject about which you are researching (if you have one). Examples of entities include companies, organizations, facilities, cities, malware, vulnerabilities, etc...the list goes on and on. Entities can be large, encompassing many smaller entities (example of the United States which is comprised of states, neighborhoods, facilities, etc), or small--down to the minute level (such as a singular building or IP address). We also have entity *types* which help when you know generally what you want to research (such as “Any Malware”) vs a specific piece of malware (like Dridex) or “Any Military Base” vs a specific one (like Stuttgart). Along the same lines, we also have entity categories, which further break down entity types--for instance, you know you want to research malware, but you know it’s in the Botnet family--you can search by the entity category called Botnet (which is comprised of all the known Botnets in Recorded Future). Finally, entities are created by the machine, but can be hand-curated by Data Scientists (indicated by a star in the entity picker). And if there is no entity for what you’re looking for, you can always use the free text term, indicated by italics, and always found at the bottom of the options in the entity picker. |
| 3.2b | Events are the things that happen--in Recorded Future we classify them generally into cyber events, military activity, personnel travel, civil violence, and even corporate activities. This helps us to make sense of “big data;” for instance, if you are searching for Russian military exercises, you wouldn’t want to just search for Russia (as the entity)--you’d want to apply an event type (think of it as a filter) such as “Military Maneuver” to pare down the data that fits your intelligence requirement. Further, we collapse references into events--so if there are hundreds of references about a particular event, those will generally be collapsed into a few event clusters, again to make that data more manageable. Here are some event types shown in the graphic: |
| 3.2c | The ability to research by time--whether relative or by specific date range, is truly remarkable. For instance, you can search for things in the last day, week, month, year, etc (we have over eight years’ worth of indexed data in our system--that’s billions of structured data points!) or even something from September 15, 2013-January 3, 2014. This can really help you focus your research and identify specific things that either happened within or were published during certain dates. Our machine understands specific and relative time--so if somebody mentions they will bomb the Hoover Dam next Christmas, the machine understands that is an event where an author has threatened an attack upon the Hoover Dam on Christmas Day--this date will be recorded with that event, the entities (the location & author if available), and the date. |
| 3.2d | We collect in all languages and process natively in 12--meaning we extract entities, events, etc in the native language the reference is in, all prior to translation. This means the machine must essentially know these languages to understand what is being said. We also translate the reference within Recorded Future so that you can understand it on your end (without having to leave the environment or call a linguist). This allows you to not only understand the data, but you can also search by specific language. |
| 3.2e | Finally, we harvest over 750,000 sources--that might seem overwhelming, but it’s a wide variety of open, closed, dark, and technical sources (shown on the next slide) that allow you to paint an overall OSINT picture, or even research only by specific source types. Having visibility into closed or dark sources in all languages--all without having to leave the Recorded Future environment-- gives you an immense capability to identify and understand threats. |
| 3.3 | Shown here are various sources that Recorded future uses to gather data.  As an Recorded Future user, you also have the ability to task additional collection--if you don’t see a particular source that you regularly monitor or you want to try to collect on a source you haven’t been able to before, you can submit for that collection through the Collection Request Application, or work with your Intelligence Services Analyst.  Let’s look at an example. Click on the magnifying glass to zoom in on the image to see how Recorded Future has collected the data. |
| 3.4 | Let’s look at an example. Click the magnifying glass to see how information is collected. |
| 3.4a | The bolded text indicates extracted entities, each of which you can click on to pivot into further analysis. Also notice the title of the reference (here, “Ransomware Cyber Attack against National Health Service)--that is the extracted event. Further, the reference provides the date of the post (and if the reference had the date of attack, that would also be noted) as well as the original source (making attribution easy). |
| 3.5 | What sources does the platform collect from? Well it sources data from everywhere…some examples of places the platform collects data from are:   * Dark Web / Special Access * Forum - Underground * Social media * Paste sites * Mainstream News and * Code Repositories   The ability to select which source type to query is also important here--especially if one isn’t looking for a specific source, but wants to stay within a category of sources, like Forums or Paste Sites. This is a powerful capability for getting insight into closed or hard-to-access sources, and helps to focus one’s research. |
| 3.6 | How do we collect from all the sources? We use RSS Feeds, Providers, Custom Harvesters, Forum Reading, and Special Access. This gives us an increasing complexity of access.  As such, some sources are quite easy to add and maintain, others can be more challenging or time consuming. We’re constantly working to add more sources and increase our depth and breadth with the hard-to-reach or closed sources. We highly encourage you to reach out to your Intelligence Services analyst if you need additional collection. |
| 3.7 | Let’s take a minute before we move to the next topic to check in with our progress. Read the question and choose the best answer. |
| 3.8 | Now that you’re familiar with how and from what sources we collect data, we’ll look at the analysis of the data. Click through to the next slide to get started. |
| 3.9 | When we harvest content from a source, the text content is inspected first to identify the natural language. If the content is identified as a Deep Analysis language, then the content is processed.  First we start with;  Dividing the content into paragraphs and sentences, then  The sentences are analyzed to detect entities – this uses advanced natural language processing.  Then implied context is used to resolve references to time, locations, and entities.  Next, sentence level event detectors are applied and finally document level event detectors are applied.  Deep Analysis transforms the original content into a collection of reported events. This includes both semantic events and Mention events that highlight sentences that involve specific entities.  Deep Analysis supports dozens of entity detectors and almost 100 event detectors. The Query Builder provides a comprehensive list of the current entity and event detectors.  The Recorded Future Platform collects and translates from many languages, but Deep Language Analysis is only performed on a few. click on the information button to learn about deep analysis languages. |
| 3.10 | This is an example of how machine and human interaction work together…in this example the platform has identified the words “zbot”, “zeus”, and “zeus” as malware. This is a simulation of how the machine + human data curation process happens. The machine has identified references with the words Zbot and and Zeus. The machine knows Zbot is a kind of malware, and recognizes Zeus as a kind of malware too based on the first reference (Zeus Malware), so it maps both Zeus references to the Zeus malware. |
| 3,11 | When the human steps in to review, it tells the machine that Zbot and Zeus are synonyms for the same malware. |
| 3.12 | The third reference actually refers to a company--so it must disambiguate the two based on new rules it teaches the machine. |
| 3.13 | This is an iterative process and constantly improves our machine learning algorithms. |
| 3.14 | Shown here is a multilingual example, where the machine processes the russian reference in its native language first, extracting the Sberbank term in Russian, and mapping it to the entity. |
| 3.15 | Now that you’re familiar with how the data we collected is analyzed we will move on to how the machine indexes this data. . Click next to continue. |
| 3.16 | As we have discussed, data is structured and indexed a few different ways. The first one is an entity. Recorded Future uses Entities to structure intelligence information. Entities are people, places, and things in the world. Some entities exist in the physical world, such as a City or a Person. Other entities like Companies exist in the legal world. Many entities exist in the information security world: Malware, Vulnerabilities, IP Addresses, and so forth.  Depending on the entity, specific actions may have been completed. Those actions can be disambiguation, gazetteering, aliases, and linking associated terms (hashtags, people, products etc.) the process that was just demonstrated in the previous slides. Entities that have had human interaction are referred to as “curated” and are indicated by a star next to it. You can see the level of curation by checking out the Entity Structure Viewer found in the Mega Menu in the top right-corner (graphic shown below).  Entities without a star next to it indicate they have been extracted and structured by the machine, with no additional human interaction. You can always request Data Review for entities when conducting research in the UI.  Free text or text match terms should be used when you are looking for a particular string of words and letters and/or when you don’t want any additional ontological data included in your results (described in more detail in the next slide). For instance, if you were to use the entity United States, in your results you would get tons of data mentioning the United States, all 50 states, thousands of cities, companies, regions, facilities, airports, etc--because those are all linked to that entity. To get just mentions of the “United States” use the text match term  Click the information icon for more information. |
| 3.17 | Entities in Recorded Future have ontology relationships. “Ontology” simply means meaningful links between entities. Many ontology relationships are used by default in searches, because these are relationships where in general we consider one entity to be “part of” or “contained in” another entity. Searching for events involving the “container” entity will also implicitly match the “contained” entity.  Ontologies can be very useful in research because you’ll spend less time coming up with search terms--these entities will likely already contain the terms you’re looking for. Other times, you’ll want to be more specific--you have the power to try both approaches now that you understand the differences.  Click on each example to learn more. |
| 3.18 | Let’s take a look at an example of an ontology. Here, the platform is searching for Rockwell Automation, an industrial equipment company. |
| 3.19 | Here you see how the levels of curation work on the platform.  Drag the slider on the screen to learn more. |
| 3.20 | To check the curation and structure behind an entity (people, place, or thing) you can check the Entity Structure Viewer by clicking on the link on the screen.  If you need to request more work on an entity, use the Collection Request application or contact your Intel Services Rep. Both the Collection Request App and Entity Structure Viewer are found under Tools.  Click Collection Requests on the scren to see what happens. |
| 3.21 | Event Types can be selected in the Query Builder to refine your research--think of them as filters on big data. For instance, in the example shown, we used Russia as the entity and Military Maneuver for the Event Type (as well as selected only Russian news and blogs as our sources)--if we didn’t, we would get massive amounts of data about Russia--with no focus. Now, we are able to pull out key pieces of data regarding Russian military activity, both currently, 30 days in the past, and 30 days in the future (as set in our Query Builder). |
| 3.21a | In a second example, we’ve used Sergei Lavrov, the Russian Foreign Minister, and applied the Travel Event Type to extract events where Lavrov is mentioned to have travelled, or is reported to be in a certain location (note the future events and the locations that Lavrov is scheduled to visit): |
| 3.22 | The platform maps certain terms and phrases in various languages (NLP and basic) to Recorded Future Event Types.  Event Type categories include Cyber, Geopolitical, Person, Corporate, and Financial, with a number of event types within each category. These Event Types can be found in the Query Builder in the Events section. Note, you do not have to apply an Event Type if you don’t want to!  Think of Event Types as filters on large data sets to help make sense of data and begin to refine a query. The filters are not for identifying ‘needle in the haystack’ references.  To see a more in depth list of different Event Types, click on the link on the screen. |
| 3.23 | You can set the event time or publish time in the Query Builder if desired. You can set it by relative time (such as +/- 30 days, last year, last 3 months, etc--type in different numbers--as long as you see a green check mark you are correct.  You can also search by specific dates using the calendar feature, allowing you to pinpoint events that happened (or were published) in a certain date range--a powerful capability.  Click each feature to learn more. |
| 3.24 | Let’s stop here before we move to the next lesson to check in with our progress. Follow the instructions on the screen. |
| 3.28 | You’ve completed lesson 2! You now have an understanding of how our data, sources, and structure are collected, analyzed, and indexed. Click next to get started on lesson 3. |
| 4.1 | In lesson 4 we are going to give you a better understanding of how to navigate through the User Interface and understand how to use the various visualizations |
| 4.2 | Here you see what your workspace will look like and how it is organized. There are several types of items that you can add to your home screen to access quickly, like Favorites in an Internet Browser. Each one is depicted with a certain icon type. Click on one of each of the four icon types to learn more. We will also go into more detail in the coming slides.  Keep in mind that anytime you see blue in the system, it represents a link that goes to another place in Recorded Future. |
| 4.3 | A list of queries you have saved will appear at the bottom of your screen. The Recorded Future Platform allows you to modify your settings to give permission to others to view your lists, queries, collections, and alerts.  By selecting the pencil you can edit you can view your workspace. You can set a different permission for each item. For example you can set your alerts to be only viewable by you, but your lists to be viewed by one or more other people.  Click “Me” in the box in the image below. |
| 4.4 | Let’s take a moment to become familiar with the Mega Menu and how it works. There are five main headings.   * **Threat Research** section (Insikt Group is our own Threat Intelligence Research Group and you can access their notes in the product as well as their blog) * **Learn More** to access the Support Site (must be logged in to view articles) and the Change Log (to keep track of new changes to the product) * **Tools** to view the Collection Request app, the Entity Structure Viewer, and User Settings * **Alerting** to view Alert Rules and Manage Alerts * **Workspace** to go back to the home page or filter by specific item |
| 4.5 | In the workspace you can create a List of Entities to use in a Query, instead of typing in multiple Entities, one at a time, in the Query Builder.  You can manually type the Entities in, or you can upload a list via a CSV file.  The List will save to your home workspace. ~~to~~ To use in a Query, simply type the name of the list in the query builder involving field, or search for it in the lists category. |
| 4.6 | You can view the data you’ve collected using 5 different views. The platform has a timeline view, table, map, source map, and feed view.  Click on each one to learn more. |
| 4.7 | It’s time for another Knowledge Check. Read the question and choose the best answer. |
| 4.10 | Now that you are more familiar with the workspace, let’s move on to the next section. Click through to the next slide to get started. |