

NILE-Metadata Framework Design

Curated Catalog	2
Metadata Requirements for Using Resources in NILE	2
Resources	3
Competency Metadata	3
Relevance	3
Engagement	4
Efficacy	4
Rigor	5
Transcript, Summary, and Phrase Cloud	5
Plot & Exposition Similarity	5
Educator Signalling by Organization Resources	7
Signature Assessments and Collections	8
Future Work	8
Appendix A: Navigator Glossary	10
Appendix B: Catalog Metadata Framework	14
Resources	14
Questions	15
Collections	16
Assessments	17
Lessons	18
Units	19
Courses	21
Rubrics	22
Metadata Details	22

Curated Catalog

One challenge with Navigated Learning is to find resources that could be used to construct learning pathways. Gooru's initial focus on developing a catalog of resources by human-driven methods, ranging from direct accretion of resources by Gooru staff to crowd-sourced acquisition of resources from hundreds of educators who were using Gooru. These resources were tagged directly with competencies and were in many cases placed into collections for classroom use by teachers. This step was very important because, as explained later in the paper, it provided data that could be used to train machine learning algorithms to automatically carry out this tagging step.

Tagging resources with associated competencies is commonly called alignment [REF] and is often considered the most critical aspect of a resource, but in our view many other types of metadata are needed to algorithmically determine how to use a resource in a personalized learning pathway. We therefore computed other metadata elements such as "relevance," "engagement," "efficacy," "rigor," "transcript," "summary," and "phrase-clouds" for each resource. For each of these, we developed algorithms that derived their values from information within a resource (intrinsic information) as well as event log data from students using these resources (extrinsic information). We also created processes, tools and incentives for users of the Gooru platform to organize resources into learning units and courses. The process of tagging resources with metadata and then selecting a small subset for each competence is what we call curation.

The above methods were used to curate millions of resources, and the results were applied in Learning Navigator to recommend learning pathways based on a learner's current knowledge, history, and learning goals. This paper discusses the metadata we used for curation, the algorithms used to generate this metadata, and how this metadata was updated based on activity stream data as students used the learning system. It is our contention that the curation and recommendation process described can deliver personalized learning at scale and provide substantial savings over developing new resources while enabling training to be developed and deployed more quickly in response to changing environments.



Metadata Requirements for Using Resources in NILE

Instructors play important roles in many real-world education and training environments, and in adult learning, the learner is often involved in setting their own direction. A globally scalable model of personalized learning should, therefore, enable humans to construct outcomes-focused learning experiences for students with the aid of technology that finds the most appropriate resources for every given learner. This requires tagging resources with standard metadata such as title, description, thumbnail, and copyright that are very valuable to humans and that are represented in educational metadata standards such as Dublin Core Education and IEEE Learning Object Metadata [REF]. These metadata elements, however, are largely

unusable by a machine that is making a learning recommendation. While standard search algorithms can find resources relevant to key terms and, if the alignment data is present, to a given competency, this does not provide enough support and involves too much manual labor. To enable machines to mediate between the learning needs of an individual and large numbers of potential learning resources, it is necessary to curate resources with metadata that can

algorithmically be used to determine the appropriateness of a resource in a given learner's circumstance.

Resources

The resources we indexed consist of openly accessible digital content such as websites, videos, interactive games, and text documents. These can be digital activities (e.g. videos) or offline tasks (e.g. a project described using Google Docs). We crawled, indexed and extracted metadata information from these resources. The metadata we extracted included descriptive metadata that is useful to humans and other metadata described below.

The resources indexed covered K-12 and higher education subject areas including language arts, mathematics, social sciences, science, personal finance, technology & engineering, and arts & humanities. Resources in the catalog are regularly updated and curated using the activity data from the community of users.

Competency Metadata

The primary characteristic of a resource is the competencies it helps students acquire, defined in the context of a subject, domain, and level, as described above. The algorithms we developed were able to determine the subject, level, domain, and competencies associated with a resource. These algorithms analyzed the text in the resource, available metadata (e.g. existing title, subject, description), links to other resources, and the placement of these resources in collections developed by teachers. The algorithms were trained on data gathered by human means during the early stages of Gooru development and were then permitted to proceed without human intervention and corrected as teachers and students continued to use Gooru.

In practice, teachers are less interested in how Learning Navigator classifies a resource as in how it aligns with state standards or common core standards, which is additional metadata about a resource. Since each competency in our framework is cross-walked to state standards, we were able to display this information for teachers. For the purposes of constructing pathways, Learning Navigator also needs to know the prerequisite competencies for a resource. These too are present in our framework and may also be viewed as additional metadata.

A fair amount of work has been done on automated alignment of resources to competency frameworks. Results from the 1990's using latent semantic analysis (LSA) and similar techniques were promising and were improved on to show that in an educational digital library, machine-learning-based text analysis techniques do a reasonable job of matching the topical classification of resources derived from examining where contributors placed the resources in a topical hierarchy. The algorithms used in Gooru are in the same general category and have the same drawback, which is that they operate on text. This makes them less applicable to videos and certain other types of resource (e.g. serious games) that are rich in media but poor in text, usually only containing a title and description that can be analyzed. Even web pages today are often visual and auditory in nature. Several projects are now working on using transcripts of videos and speech-to-text techniques to generate more linguistic data with the goal of improving the quality of classifications.

Relevance

Relevance is defined as the (perceived) usefulness of a resource to educators and trainers. This was determined from user activity stream logs. For example, these logs describe how instructors have arranged resources into collections, lessons, units and courses. As a result, relevance is based in part on how many instances a resource is used by educators in courses and assignments.

Let c be a competency with which a resource (or activity) a is aligned, let $COL(c)$ be the collections that use any resource aligned with c , and let $COL(a)$ be the collections in which a is used. Then the relevance $r(a)$ is defined as

$$r(a) = \frac{|COL(a)|}{|COL(c)| + \epsilon}$$

where $| \cdot |$ denotes the size of a set.

Engagement

The (perceived) engagement score of a resource is a measure of how much the resource is used, shared, and liked by students and learners. This includes time spent using that resource, whether a student shared the resource, the student's reaction to the resource (e.g., based on surveys and "likes"). Let $l(t)$ represent the set of all reactions (likes, recommendations, etc.) from users concerning an activity t , and let $u(t)$ represent the set of users who have successfully completed activity t . The engagement score of $e(a)$ is calculated as:

$$e(a) = r(a) * \left[\alpha \frac{|l(a)|}{|l(t)|} + (1 - \alpha) \frac{|u(a)|}{|u(t)|} \right] .$$

The parameter α determines how the measure of engagement is split between actions that communicate an affinity for a resource and actions that demonstrate an affinity for a resource.

Efficacy

The efficacy of a resource is a measure of its impact on a student's learning. This can be most accurately determined by evaluating a student's learning outcomes after using a given resource. Every assessment and resource is linked to one or more learning standards. After a student takes an assessment, we can evaluate the efficacy of the resources the student has used to learn those standards, based on the student's resulting assessment performance. To do this, we use a standard approach known as Bayesian Knowledge Tracing (BKT). Because individual learners have used varying combinations of resources to learn the standards connected to a given assessment, we can deduce exactly which resources contributed to the student's resulting performance on an assessment. As a result, the efficacy of a single resource can be isolated from the efficacy of other resources.

In more detail, the efficacy score for a learning activity measures how effective it was in making an *observable difference* in outcomes displayed by the learner after performing this activity. This is computed using the model of Bayesian Knowledge Tracing (BKT), which is quite popularly used in automated tutoring. BKT is based on computing the probability of acquiring a competency, based on four priors:

- P_{init} : the probability that a learner already has the said competency
- $P_{transit}$: the probability that a learner will transit from not having the competency to having the competency after participating in the learning activities of this node
- P_{slip} : the probability of a learner having the said competency failing to show it in the signature assessment
- P_{guess} : the probability that a learner not possessing the said competency, clears the signature assessment by luck.

Given the above, the probability of a given learner u obtaining the said competency based on positive and negative outcomes, is given as follows:

$$P_u(c|outcome = 1) = \frac{P_{init}(1 - P_{slip})}{P_{init}(1 - P_{slip}) + (1 - P_{init})P_{guess}} \quad (6)$$

$$P_u(c|outcome = 0) = \frac{P_{init}P_{slip}}{P_{init}P_{slip} + (1 - P_{init})(1 - P_{guess})} \quad (7)$$

The overall probability of a learner obtaining the competency is given by:

$$P_u(c) = P_u(c|outcome) + P_{transit}(1 - P_u(c|outcome)) \quad (8)$$

The efficacy score for a competency node c is computed by summing up all the competency scores for all users who have completed the signature assessment:

$$efficacy(c) = \sum_u P_u(c) \quad (9)$$

The efficacy score for every learning activity a that is mapped to c is computed by multiplying the cosine score of its relevance and engagement scores with the efficacy score of the competency:

$$efficacy(a) = \frac{relevance(a)engagement(a)}{\sqrt{relevance^2(a) + engagement^2(a)}} efficacy(c(a)) \quad (10)$$

Rigor

Rigor is scored against Bloom's taxonomy. Questions are classified to establish not only the competency they measure, but also the rigor with which the competency will be measured. We use a standard classification approaches that use the words in the question to establish the rigor for the question.

Transcript, Summary, and Phrase Cloud

The transcript of a resource is the text contained in a resource, including a video or audio file. The transcript is an important tool because we can use it to summarize the content of the resource and pull out key words as a "phrase cloud." This, in turn, can be used to analyze the resource. In our system, we associate a transcript with every resource, and then generate key words which are also associated with the resource.

Plot & Exposition Similarity

Plot similarity means that multiple resources cover the same subject matter (including domains and competencies). Essentially, this means that, within a student's learning pathway or sequence, either of these resources could fulfill the "next" activity that the student encounters. Two resources with plot similarity can potentially serve as alternatives for different students with different needs and levels of understanding.

Generating metadata related to plot similarity is essential for suggesting an alternate resource if a given resource is not efficacious and for engaging a student, based on their performance, needs, or interests. Effective resource libraries contain multiple resources with plot similarity so that learning can be truly personalized to an individual, unique student.

Exposition similarity refers to the design, context, and style of a learning resource. As is the case with plot similarity, it is an important metadata tag for identifying alternate resources for a student. If the Learning Navigator were to select an alternate resource with an exposition that differed radically from that of the overall learning sequence or the learning resources immediately preceding and following the selected resource, then the student would have a jarring and confusing learning experience. For example, if a student is engaged in a series of learning activities that use comic book scenes to describe planetary motion, the student will be surprised to see a resource composed of technical text. By including metadata on exposition coherence, the machine can select alternate resources of similar look and feel, so that the student has a seamless and coherent experience.

The proposed problem is presented in a setting, comprised of a large corpus of open learning resources, created by several authors independently. A learning resource is any kind of digital or offline resource, that provides learning content or activity for a given topic. A learning resource could be in the form of a text document, a video, a set of slides, a podcast, etc. The only assumption about a learning resource is that the resource as a whole addresses one competency. One of the first steps towards computing semantic coherence is to first convert disparate kinds of learning resources into a canonical, text-based model. For this, we use several tools like transcript generators, OCR tools, etc. to convert every learning resource in the corpus, into a collection of sentences.

The semantic coherence problem is now defined as a function that measures consistency in exposition between any two learning resources. Exposition refers to the style in which a particular narrative is presented. To model this, we think of an exposition as an unfolding sequence of topics, that is used to present the competency addressed by the learning resource. The unfolding sequence of topics is in turn, computationally modeled as a random walk. We explore two approaches for implementing topical unfolding over a random walk. In the first approach based on counting co-occurrences, learning resources are represented as co-occurrence graphs and a kernel function is defined over the graphs to represent semantic coherence, based on the random walk. In the second method based on co-occurrence prediction, a neural skip gram model is used to predict the neighborhood of possible co-occurring terms, given a base term. This skip gram model is used to define a random walk.

In the Graph Kernel Method, a term co-occurrence graph is built for each learning resource in the corpus. Given a pair of learning resources, semantic coherence between them is defined as a kernel function, that is in turn computed on a combined co-occurrence graph. For each pair of learning resources that are compared, first the key-phrases are extracted from the textual description of the resource in the key-phrase extraction phase, then a semantic context graph is generated for each learning resource which includes all the words that the key-phrases in learning resources are semantically associated with. Then, in the merge-graph generation phase, the semantic context graphs of the two learning resources are merged in a way that emphasizes or highlights the intersected nodes and edges of the semantic context graphs. On this merged graph, several hypothetical sequences of random walks are then run to compute the semantic coherence score between the learning resources (Figure 7).

Word Embedding Method is based on co-occurrence prediction, a neural skip gram model is used to predict the neighbourhood of possible co-occurring terms, given a base term. This skip gram model is used to define a random walk. We use the seminal work of word embedding or word2vec for this method. The word2vec embedding model generated an m-dimensional vector for terms from a corpus, where semantically close terms are closer to one another in the vector space. The parameter m is a hyperparameter, to be specified during the training phase. The embedding is performed by training a shallow neural network with one input layer, one output

layer and one hidden layer. The hidden layer comprises of m neurons. The input and output layers comprise of $|V|$ neurons, where V is the vocabulary of all terms in the corpus.

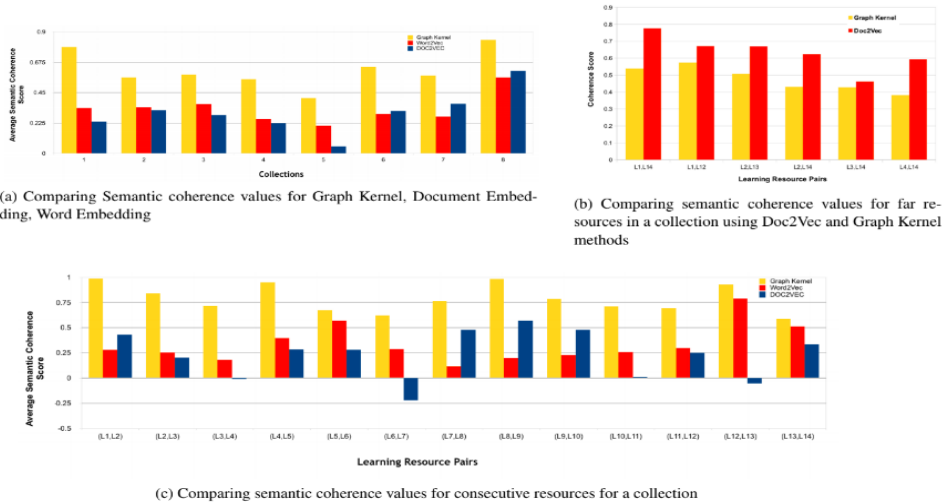


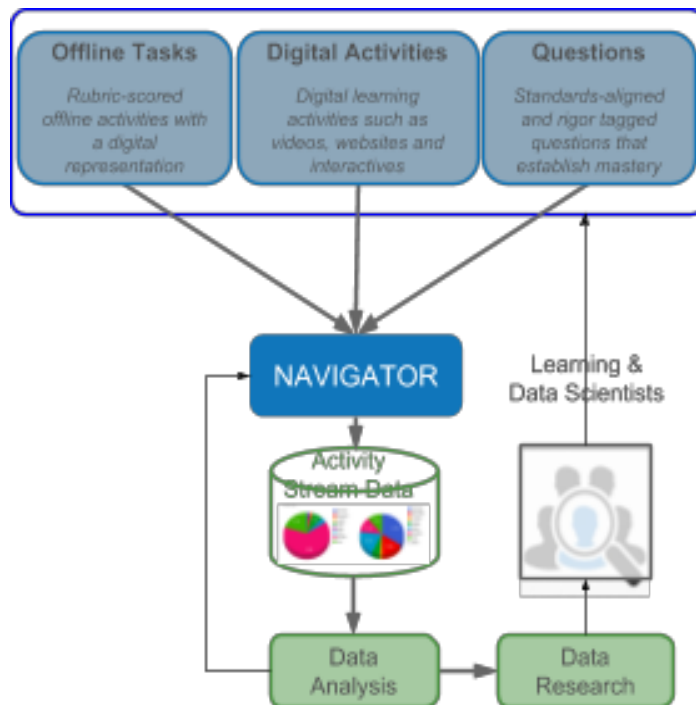
Figure 7: Comparing Different Semantic Coherence Approaches

Educator Signalling by Organization Resources

While the generation of the metadata described above is vitally important to successfully selecting resources to meet a student's needs, signals from experts organizing these resources is critical to creating an effective learning experience for the student. It is important that resources are organized into coherent collections, units, and courses that include assessments and informal "checks for understanding" to track each learner's progression.

Machines can use a variety of techniques to curate resources using the content and usage data. However, one of the most important ways to improve the quality of machine classification is to pay attention to expert signaling and to train algorithms with data derived from expert activities. We achieved this by providing educators with the tools to organize resources into courses (with assessments) in which we can assume that students experience a structured and orderly learning environment. Using the resources in the Gooru catalog, and leveraging the metadata that describes those resources, educators can design and share collections, units, courses, and assessments. The quality of machine curation is significantly aided by the training data set created through expert activity.

The quality of curation can also be improved through an iterative process in which we start with reasonable initial values for the different metadata elements associated with resources, use these metadata values to trigger



the logic that results in the resources being presented to students, and then update the metadata based on their engagement and outcomes from the activity stream data. This is shown in the figure above..

To evaluate our algorithms, we used the dataset of open learning resources aggregated by the educational platform www.gooru.org. The dataset comprised of about 4.2 million learning resources in total that were independently created by several authors. A team of educators previously created several learning pathways or “collections” over subsets of these resources. Each learning pathway represented a coherent sequence of learning resources that a student could use, to achieve a learning goal. We used these manually curated collections of learning resources as our ground truth, for testing.

Signature Assessments and Collections

Signature assessments and signature collections are developed at the competency level; they are tied to the competency and not the content. Each signature assessment and collection is created to address only one competency. Signature assessments and collections can be introduced individually or merged together to address one or more competencies. Signature assessments are presented to the student post regular tests based on users’ performance at similar competency aligned tests. Signature collections are presented to the student as backfills and infills.

At the completion of any assessment or collection within a lesson, Navigator will suggest 0-3 performance-based suggestions (which we also refer to as reroute suggestions) as part of the summary report.

While studying along a learning path in the classroom course, a signature assessment or collection can be introduced at any point in their journey to support their individual learning.

1. Signature assessments/collections are exclusively created by Gooru.
2. Each signature assessment and collection is aligned to one competency.
3. Signature assessments and collections can be introduced individually or merged together to address one or more competencies.
4. Signature assessments:
 - a. Typically contains 5 questions to ensure a variety of Depth of Knowledge and question types
 - b. Contain only machine-gradable questions
 - c. Can be suggested a maximum of 3 times total
 - d. Can be used to earn mastery for a given competency.
 - i. If a student earns a score of 80% or greater for a given signature assessment, they earn mastery for that competency.
 - ii. A student earns mastery for successful completion ($\geq 80\%$) of a signature assessment, regardless of the context in which the signature assessment is presented to the student.

Future Work

The greatest need for curating learning resources using ML and AI is the availability of high fidelity usage data. This requires that we have a full-fledged system that is commercially useful and engaging. Developing a product, integrating within the student’s learning environment, training educators and students and supporting them involves significant costs, range of talent different from those required for curating learning resources. Once the environment has been setup, given that the data spans many learning activities, the computation of efficacy is challenging in such circumstances. There is no objective way to determine if the curated value is accurate - this is rather subjective. The proof will be in the learning outcomes when the resource

is used. Algorithms for the computation of many useful metadata tags show promise, but need to be validated only against learning outcomes, which take time to establish

Learning is a very broad space and addressing this for every domain is a very huge task. We will initially focus on learning mathematics and get the curation right as evidenced by positive learning outcomes in users who learn using these curated resources.

We proposed a model for curating metadata for millions of resources by using the big-data from user activity streams and ML and AI techniques such as classification, doc2vec, BKT, etc. We illustrated the results with classification for typical metadata such as subject, grade and standard as well as for the more unique metadata such as exposition similarity between a pair of learning resources. We argued how a closed-loop approach would iteratively improve the quality of the metadata using new usage data. This work is pivotal to enable the recommendation of personalized learning pathways.

Appendix A: Navigator Glossary

Terminology	Definition
Air Traffic Control View	Quick view of the entire class plotted on a graph, depicting progress vs performance of each student.
Assessment	A set of scored questions that teachers and students can use to monitor understanding and performance.
Baseline	A black line on the Proficiency Chart that indicates student's starting location, the base-level concepts and skills they have mastered in each domain.
Classroom	"Virtual classroom" within Gooru where teachers add students or students join the classroom and move through content that the teacher has "assigned" or made visible
Classroom Code	A code unique to each classroom that students can use to join the classroom
Classroom Roster	The students (Gooru users) who joined a classroom via classroom code
Collection	A playlist of multimedia resources (videos, website, images, etc) and questions
Competency	This term is synonymous with learning standard. K-12 educators generally understand competencies and content developed at the competency level. Much of Gooru's current content is developed around learning at the competency level.
Competency Report	Competency summary of all students in the class that includes the domain names, how many competencies are in each domain, and each student's mastery status.
Completion	Measured progression through a course by a student
Content Editor	Where a user can build and edit their content on Gooru. Also where they develop the course hierarchy (unit, lesson, and collection folders) and add and organize content
Content Information	A designated section in a collection, unit, or course to provide additional thoughts about the content
Content Settings	The status of content in terms of visibility - can be private, shared individually, assigned to a student or students, or publically visible.
Course	A "folder" that allows you to organize your learning content into units and lessons.

Course Map	View of all the content available in a classroom
Course Objectives	Goals for academic mastery within the course designated and written by the author
Current Location	Where a student is in a Gooru course in terms of mastery of a standard with regard to the learning targets. Demonstrates understanding and progression through the collections, lesson, unit, and course.
Daily Class Activities	Collections and assessments that teachers want students to complete that day while monitoring their results in real-time. It allows teachers to provide real-time help to students, who may or may not complete the whole assessment or face difficulties. Teachers can view student performance data by launching the Go Live dashboard to see how students responded.
Data	Data of student assessments, attempts, scores, etc for a classroom
Drag and Drop Order	Question type in which student clicks and drags multiple answer choices so that they are in the order as requested in the question.
ECAP	Event-Condition-Action-Principle (ECAP) refers to the event such as poor performance in an assessment, condition is the learner profile, action is the suggested remedy - study a collection or retake an assessment, based on a principle of learning
Fill-in-the-blank	Question type in which student enters text to complete a statement in the question that is missing words or phrases.
Gooru Catalog	The set of all content available in Gooru, searchable by collection, resource, and standard.
Highlight Text	Question type in which student highlights a text or a sentence to select the correct answer as requested in the question.
Lesson	A lesson is a sub-folder in the course folder hierarchy.
Library	The content on a user's profile--their courses, collections, assessments, questions, and uploaded resources
Metadata	Information associated/tagged on resources, assessments, collections, courses; includes areas such as standard, grade level, type of media, etc.
Multiple Answer	Question type in which student selects "Yes" or "No" out of multiple choices.
Multiple Choice	Question type in which student selects one answer out of multiple.
Multiple Select - Image	Question type in which student selects one or more images available as answer choices.
Narration	Text that pops up before a resource to describe it and/or give instructions to

	the student studying the collection
Narrative Arc	A coherent pathway made of different learning resources, that are potentially created independently, but which offer a semantically coherent and cognitively engaging learning experience.
Go Live	When the teacher selects “Go Live” for an assessment, the results for each question for the whole class is visible on the teacher’s dashboard in real-time
Grade Line	Line connecting the highest competencies in each domain within a subject that is expected to be covered at a specified Grade level.
Learner Profile	Biography, Progress, Performance, Proficiency, Portfolio, Preference and Content
Performance Data	View of progress within a unit, lesson, or collection
Proficiency	Student skill level.
Profiles	A collection of information on the user and the user's content (the user could be an individual or an organization) that the user can update. Profiles are publicly visible.
Reaction	Student response to a resource via a likert-type scale
Real-Time Performance	Visual representation of completion and mastery within a collection, lesson, unit and course
Remix	Copy and save content to your account
Rescope	A course is tailored to include learning activities for which the learner does not have mastery.
Resource	A resource is multimedia content in a variety of formats such as videos, interactives, websites, images, Google docs, and more.
Route-0	This is a list of lessons to help the learner gain knowledge in competencies not covered by the course, but, required to learn the competencies that the course covers
Search (Collection)	User enters keyword in search box to find and filter Gooru's catalog to find relevant collections
Search (Resource)	User enters keyword in search box to find and filter Gooru's catalog to find relevant resources
Signature Assessment	Assessment offered to students when they receive $\geq 80\%$ on their Course Assessment, allowing them the opportunity to gain mastery of a competency.

Signature Collection	Additional resources offered to students to enhance their understanding of the concept if they receive < 80% on the Course Assessment.
Skyline	A white line on the Proficiency Chart that represents the highest-level concepts and skills they have mastered in each domain.
Skyline Algebra	An algebra over the set of all polylines that is used to compute pathways, cover or union of skylines, etc.
Standards	District, state or national frameworks that define content areas and skills. Well-known standards include Common Core, Next Gen Science Standards
Teacher Suggestion	Additional collections or assessments offered to students to reinforce concepts where students may be struggling.
Time Spent	Amount of time spent on a question, resource, assessment, or collection
Unit	A unit is a sub-folder in the course folder hierarchy.

Appendix B: Catalog Metadata Framework

This document describes the different content types (also called Learning Activities) in the Navigator’s content model, their structure, associated metadata and the hierarchy. We will describe the various content types and how these are organized into collections, assessments, lessons, units and courses. Navigator identifies and captures several metadata for the learning activities. The metadata captured can vary depending on the type of learning activity. The list of all metadata captured with description about them can be found in the Appendix. The metadata can be grouped under various categories such as Descriptive, Educational, Media, Framework etc. and also based on whether user tagged or curated and machine extracted or computed. The following tables list the related metadata for each activity. Refer to the Appendix for description about each metadata, which of these are mandatory etc.

Resources

Resources are individual learning units aligned to a specific competency and can be one of the following types:

- Web pages: Web pages are hosted externally and can be accessed via a URL.
- Video: Videos are publicly accessible Youtube or Vimeo videos
- Text: Any uploaded PDFs or documents on Google Drive.
- Image: Image of any format
- Audio: Audio resources with embedded players in external web pages or in Google Drive
- Online textbooks: Offline digital activities. Link and instructions provided within a document on Gooru and task completed outside of Gooru
- Interactive (including games and manipulatives): Resources on external web pages
- Offline activities: Activities that are completed offline and data about these reported online

Group	Metadata
Descriptive	Title Description
Educational	Educational Use Audience Time Required Language
Framework	Subject Taxonomy Course Domain Standards (Competency) Concepts
Instructional	Habits of Success (21st Century Skills) Instructional Model Scaffold

Vectors	Relevance Engagement Efficacy
Media	Transcripts Alternate Language URLs (Uniform Resource Locator) Video translation Format Url Advertisement Level Access Hazard Media Features Mobile Friendly Data Type Keywords Access Mode Has Adaptation* Is Adaptation* Control Flexibility Visibility Framebreaker Broken Thumbnail
Creation	Publisher Aggregator Host Author License Copyright Country Code Created Date Modified Modified By Creator Id Date Created At Source
Other	Tags Teacher Resource

Questions

Questions are used to evaluate the student's understanding of concepts and can be one of the following types:

- Multiple choice

1. These questions present multiple options as answer choices of which only one option is the right answer. Student can select only one of the options to mark as the right answer.
- Multiple Answers
 1. These questions present multiple options as answer choices of which more than one is the right answer. Students can select multiple options to mark as their answer
 - Fill in the Blanks
 1. A Fill in the Blank question consists of a phrase, sentence, or paragraph with one or more blank spaces where students provide the missing word or words.
 - True or False
 1. Students choose True or False in response to a given statement in the question
 - Highlight Text
 1. Students select specific words and highlight them as a response in the given statement in the question
 - Highlight Image
 1. Students are presented with multiple images for a question statement and select one or more images as a response
 - Drag and Drop Order
 1. Students are provided with a list of items that need to be sequenced in a particular order. Students can use drag and drop to sequence them in the right order
 - Multiple Select Image
 1. Students are presented with multiple images as answer options and are required to select one or more of these as the right answer
 - Multiple Select Text
 1. Students are required to select multiple words from a sentence or sentences as answer to the question
 - Free Response
 1. Students provide answers in free form text. Free Response questions have a rubric associated with them for the teacher (or student) to grade the answers. Free response questions have the flexibility to be activities, projects, labs, concept maps, and any offline, experiential question. Each FRQ can have a rubric associated with it.

Collections

A collection is a playlist of resources on a topic curated by a teacher. Each collection is comprised of resources, which may include images, videos, webpages, games, textbooks, questions, and more.

Students interact with the content at the collection level. Content author is expected to include learning objectives and consider using a variety of resource types to expose students to the concepts in multiple ways.

The resources are expected to be sequenced in the right order to build on the concepts. Progression through a collection should flow in a logical manner and take the intended audience from a general to more complex level of understanding if appropriate, or allow adequately for student exploration. Questions are used in collections to check for understanding along the way.

Group	Metadata
Descriptive	Title
Educational	Audience Learning Objective Time Required
Framework	Subject Taxonomy Course Domain Standards
Media	Keywords Thumbnail Visibility
Creation	Publisher collaborator instance creator original creator Aggregator License Created Date Modified Modified By Creator Id
Other	Tags Teacher Resource

Assessments

An assessment is a set of scored questions, which are used to monitor understanding and performance. Assessments typically have a variety of question types (including several based on the SBAC) so that students can demonstrate understanding in different ways. Each question should be tagged to standards, micro-standards, and Webb's Depth of Knowledge.

Group	Metadata
Descriptive	Title
Educational	Assessment Goal Audience Time Required
Framework	Subject Taxonomy Course Domain Standards
Vectors	Relevance Engagement Efficacy
Media	Keywords Visibility Thumbnail
Creation	Publisher collaborator instance creator original creator Aggregator Author License Created Date Modified Modified By Creator Id
Other	Tags Teacher Resource

Lessons

Lessons have two types of activities: learning collections and assessments. Lessons are built with collections and assessments that are created by the author or remixed from collections and assessments that are in the Navigator catalog. Lessons are tagged to standards and learning objectives along with other metadata.

Group	Metadata
Descriptive	Title
Educational	Pacing guide Time Required Grade Level Learning objective
Framework	Subject Taxonomy Course Domain Standards
Instructional	Lesson Plan
Vectors	Relevance Engagement Efficacy
Media	Keywords Visibility
Creation	Publisher Aggregator Instance Creator License Created Date Modified Modified By Creator Id
Other	Tags Teacher Resource

Units

Units is another container layer for organizing lessons together to create a diverse experience for the learner. Units are identified with Big Ideas and Essential Question. Units are sequenced chronologically, by topic, or by standard.

Group	Metadata
Descriptive	Title
Educational	Big Ideas Essential Questions Pacing guide Time Required
Framework	Subject Taxonomy Course Domain Standards
Instructional	Unit plan
Media	Keywords Visibility
Creation	Publisher Aggregator License Created Date Modified Modified By Creator Id
Other	Tags Teacher Resource

Courses

A course is a container folder that allows the author to organize learning content into units and lessons, and easily share the entire course with the



learners.

Group	Metadata
Descriptive	Title
Framework	Subject Taxonomy Course Domain Standards
Media	Keywords Visibility
Creation	Publisher Aggregator License Created Date Modified Modified By Creator Id

Other	Tags Teacher Resource
-------	--------------------------

Rubrics

A rubric is a document that articulates the expectations for a free response question or an offline activity by listing the criteria, or what counts (called categories in the Navigator), and describing levels of quality. Each level can have a score associated with it. The score for the question or activity is the one on the level assigned for each category.

Metadata Details

Resources

Metadata Type	Name	Description	Group	Mandatory
primary	Publisher	The organization credited with publishing the resource (i.e. who the content belongs to).	Creation	Y
secondary	Aggregator	The organization that catalogues content from external publishers (i.e. who catalogues publisher content).	Creation	Y
secondary	Host	The institution that hosts user-generated content (i.e. where the content is hosted physically).	Creation	N
secondary	Author	The individual(s) credited with the creation of the content at the original site.	Creation	
secondary	License	License code	Creation	Y
secondary	Copyright	Copyright information of the page.	Creation	N
secondary	Country Code	Country origin of resource.	Creation	N
internal	Created	The date and time when the resource was created in Gooru.	Creation	Y
internal	Date Modified	The date and time when any of the resource metadata was modified in Gooru.	Creation	N
internal	Modified By	User id of user who last modified the resource.	Creation	N

internal	Creator Id	The ID of the Gooru user who added the resource.	Creation	N
internal	Date Created At Source	The date that the website was last updated.	Creation	N
primary	Title	The title of the resource.	Descriptive	Y
primary	Description	A paragraph to describe what the resource is about. Should include keywords that would aid in finding this resource via search	Descriptive	Y
secondary	Educational Use	The purpose of the resource in the context of education.	Educational	Y
secondary	Audience	The audience for which the resource is intended.	Educational	Y
secondary	Time Required	Approximate time (min) required to use the resource.	Educational	N
secondary	Language	Language of resource. For videos/audios, it is the language of the dialogue.	Educational	Y
secondary	Transcripts	Transcripts of video/audio. This is an array of (language, transcript)	Media	N
secondary	Alternate Language URLs	If the resource is available for different languages at different urls, it is captured here. This is an array of (language, url)	Media	N
secondary	Video translation	State the language(s) that the video is translated into. Selectable in the video	Media	N
primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
primary	Instructional Model	The competencies identified in 5Es that the activity is aligned to	Instructional	N
primary	Habits Of Success	The competencies in 21st century skills that the activity is aligned to.	Instructional	N
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	

internal	Flagged	Resource can be flagged for multiple reasons - needs a review for content or broken link or any issues with the resource	Internal	
internal	Crawled Subject	The crawled subject of the resource.	Internal	N
internal	Import Mode	If a resource has been recrawled, "Update" replaces the old resource with new metadata fields.	Internal	N
internal	File Code	The GAT identifier for a group of ingested resources.	Internal	N
internal	Crawl Origin	Gooru if crawled by the Gooru crawl team. NextWealth if crawled by the NextWealth crawl team.	Internal	N
internal	Crawl Method	Manual if crawled by hand. Mozenda if crawled using Mozenda.	Internal	N
primary	Format	The resource type, such as a video or a webpage. A "resource," as used in Gooru, is any open source web material that can be used for the purpose of instructing or studying an academic discipline covered in K-12 curriculum.	Media	Y
secondary	Url	The URL of the resource.	Media	N
secondary	Advertisement Level	Identifies the presence of ads on a particular domain, and how severe.	Media	N
secondary	Access Hazard	A characteristic of the described resource that is physiologically dangerous to some users. This is related to WCAG 2.0 guideline 2.3.	Media	N
secondary	Media Features	Content features of the resource, such as accessible media and alternatives. These derive primarily from AfA AdaptationTypeRequired.Type/adaptationRequest.	Media	N
secondary	Mobile Friendly	Determines whether or not the content is playable on a mobile device (currently refers to content which contains flash).	Media	N
secondary	Data Type	The data type of the resource.	Media	N
secondary	Keywords	Vocabulary and key terms associated with the resource	Media	Y
secondary	Access Mode	The access modality refers to whether the user requires vision, the ability to distinguish between colors, hearing, touch and/or text	media	N

		literacy to access the resource.		
secondary	Has Adaptation*	Identifier of a resource that is an adaptation, for accessibility purposes, for this resource.	Media	N
secondary	Is Adaptation*	Identifier of a resource for which this is an adaptation, for accessibility purposes.	Media	N
secondary	Control Flexibility	Identifies one or more input methods can be used that allow access to all of the application functionality.	Media	
internal	Visibility	The visibility setting of the resource. Determines whether the resource will be displayed in search.	Media	Y
internal	Framebreaker	Whether the original source of the resource allows the resource to be displayed in an i-frame.	Media	Y
internal	Broken	Whether the resource has been marked as broken.	Media	Y
primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other	Y
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	N
secondary	Relevance	Computed value.	Vectors	Y
secondary	Engagement	Computed value	Vectors	Y
secondary	Efficacy	Computed value	Vectors	Y
secondary	Lexile Level			N
internal	Editorial Tags	Identifies Content Quality	Internal	Y
primary	Scaffold	Optional help text to students about resource that pops-up for users in study player	Instructional	N
primary	Thumbnail	Thumbnail image for the resource	Media	N
secondary	Cultural relevance	Country or region where the content would be relevant		N
secondary	Exclusionary content	Is content void of diversity (gender, race etc.)		Y

secondary	Rubric	Reference to the Rubric that is associated with the resource.		N
primary	Concepts	The micro competencies that the resource is aligned to	Framework	N

Questions

Metadata Type	Name	Description	Group	Mandatory
primary	Publisher	The organization credited with publishing the resource (i.e. who the content belongs to).	Creation	Y
secondary	Aggregator	The organization that catalogues content from external publishers (i.e. who catalogues publisher content).	Creation	Y
secondary	Host	The institution that hosts user-generated content (i.e. where the content is hosted physically).	Creation	N
secondary	Author	The individual(s) credited with the creation of the content at the original site.	Creation	Y
secondary	License		Creation	Y
secondary	Copyright	Copyright information of the page.	Creation	N
secondary	Country Code	Country origin of resource.	Creation	N
internal	Created	The date and time when the resource was created in Gooru.	Creation	Y
internal	Date Modified	The date and time when any of the resource metadata was modified in Gooru.	Creation	N
internal	Modified By	User id of user who last modified the resource.	Creation	N
internal	Creator Id	The ID of the Gooru user who added the resource.	Creation	Y
internal	Date Created At Source	The date that the website was last updated.	Creation	N
primary	Title	The title of the question	Descriptive	Y
secondary	Time Required	Approximate time required to answer the question.	Educational	N
secondary	Language	Language of question.	Educational	Y
secondary	Grade Level	The intended grade level of the question.	Educational	N

primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
primary	Habits Of Success	The competencies in 21st century skills that the activity is aligned to.	Instructional	N
primary	Depth of Knowledge	The categorization of content based on the Depth of Knowledge graph. This only applies to question resources.	Instructional	N
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	N
internal	Flagged		Internal	N
internal	Crawled Subject	The crawled subject of the resource.	Internal	N
internal	Import Mode	If a resource has been recrawled, "Update" replaces the old resource with new metadata fields.	Internal	N
internal	File Code	The GAT identifier for a group of ingested resources.	Internal	N
internal	Crawl Origin	Gooru if crawled by the Gooru crawl team. NextWealth if crawled by the NextWealth crawl team.	Internal	N
internal	Crawl Method	Manual if crawled by hand. Mozenda if crawled using Mozenda.	Internal	N
primary	Format	Question Type	Media	N
secondary	Url	The URL of the resource.	Media	N
secondary	Keywords	Vocabulary and key terms associated with the resource	Media	N
secondary	Access Mode	The access modality refers to whether the user requires vision, the ability to distinguish between colors, hearing, touch and/or text literacy to access the resource.	media	N
secondary	Has	Identifier of a resource that is an adaptation, for accessibility purposes, for	Media	

	Adaptation*	this resource.		
secondary	Is Adaptation*	Identifier of a resource for which this is an adaptation, for accessibility purposes.	Media	Y
secondary	Control Flexibility	Identifies one or more input methods can be used that allow access to all of the application functionality.	Media	Y
internal	Visibility	The visibility setting of the question. Determines whether the question will be displayed in search.	Media	Y
internal	Broken	Whether the resource has been marked as broken.	Media	N
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	Y
internal	Editorial Tags	Identifies Content Quality	Internal	Y
primary	Common wrong answers	Common wrong answers are tagged at the concept level		N
primary	Has solution	Solution or answer present for the item		Y
primary	Hint/Scaffold present	Scaffold or hint option available if the student is stuck on the item		N
secondary	Rubric	Reference to the Rubric that is associated with the question.		N
primary	Concepts	The micro competencies that the question is aligned to	Framework	N
primary	Mis-concepts	The mis-concepts that this question addresses in its answer options	Framework	N

Collections

Metadata Type	Name	Description	Group	Mandatory
primary	Publisher	The organization credited with publishing the activity (i.e. who the content belongs to).	Creation	Y
secondary	collaborator	The people who collaborated in the creation of the activity	Creation	N
secondary	instance creator	Username of the gooru user who created this instance of the activity.	Creation	Y

secondary	original creator	For collections, assessments, units, lessons, courses which were copied/remixed, the username of the gooru user who created the original collection.	Creation	N
secondary	Aggregator	The organization that catalogues content from external publishers (i.e. who catalogues publisher content).	Creation	N
secondary	Author	The individual(s) credited with the creation of the content at the original site.		Y
secondary	License		Creation	Y
internal	Created	The date and time when the resource was created in Gooru.	Creation	Y
internal	Date Modified	The date and time when any of the metadata was modified in Gooru.	Creation	N
internal	Modified By	User id of user who last modified the collection.	Creation	N
internal	Creator Id	The ID of the Gooru user who added the collection.	Creation	Y
primary	Title	The title of the Collection.	Descriptive	Y
primary	learning objective	The expected goal of interms of demonstrable skills or knowledge that will be acquired by a student. Student will be able to...	Educational	Y
secondary	Audience	The audience for which the resource is intended.	Educational	Y
secondary	Time Required	Approximate time required to study the collection	Educational	N
primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	Y

internal	Flagged		Internal	Y
primary	thumbnail	Thumbnails are relative paths to CDN URLs	Media	N
secondary	Keywords	Vocabulary and key terms associated with the resource	Media	N
internal	Visibility	The visibility setting of the collection. Determines whether the collection will be displayed in search.	Media	Y
primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other	Y
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	N

Assessments

Metadata Type	Name	Description	Group	Mandatory
primary		Publisher	Creation	Y
secondary		collaborator	Creation	N
secondary		instance creator	Creation	Y
secondary		original creator	Creation	N
secondary		Aggregator	Creation	N
secondary		Author	Creation	N
secondary		License	Creation	Y
internal		Created	Creation	Y
internal		Date Modified	Creation	N
internal		Modified By	Creation	N
internal		Creator Id	Creation	Y
primary	Title	The title of the assessment.	Descriptive	Y
primary	Assessment Goal	The expected goal of assessment in terms of demonstrable skills or knowledge that will be acquired by a student. Student will be able to...	Educational	Y
secondary	Audience	The audience for which the assessment is intended.	Educational	Y
secondary	Time	Approximate time required to complete the	Educational	N

	Required	assessment		
primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	Y
internal	Flagged		Internal	Y
primary	thumbnail	Thumbnails are relative paths to CDN URLs	Media	N
secondary	Keywords	Vocabulary and key terms associated with the assessment	Media	N
internal	Visibility	The visibility setting of the assessment. Determines whether the assessment will be displayed in search.	Media	N
primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other	Y
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	N
secondary	Relevance		Vectors	Y
secondary	Engagement		Vectors	Y
secondary	Efficacy		Vectors	Y
primary	Show answer	Correct answers indicated at end of assessment		Y

Lessons

Metadata Type	Name	Description	Group	Mandatory
primary	Title	The title of the lesson.	Descriptive	Y
primary	Publisher	The organization credited with publishing the resource (i.e. who the content belongs to).	Creation	Y

primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
primary	Pacing guide	The recommended duration in which to complete the activity	Educational	Y
secondary	Audience	The audience for which the lesson is intended.	Student	Y
primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other	Y
secondary	sequence	The sequence of the lesson in the context of a unit or the unit in the context of a course		N
secondary	Relevance		Vectors	Y
secondary	Engagement		Vectors	Y
secondary	Efficacy		Vectors	Y
secondary	Instance Creator	Username of the gooru user who created the instance	Creation	Y
secondary	License		Creation	Y
secondary	Time Required	Approximate time required to study the lesson.	Educational	N
secondary	Keywords	Vocabulary and key terms associated with the resource	Media	N
secondary	Grade Level	The intended grade level of the material.	Educational	N
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	N
internal	Visibility	The visibility setting of the resource. Determines whether the resource will be displayed in search.	Media	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	Y

internal	Flagged		Internal	Y
internal	Created	The date and time when the resource was created in Gooru.	Creation	Y
internal	Date Modified	The date and time when any of the resource metadata was modified in Gooru.	Creation	N
internal	Modified By	User id of user who last modified the resource.	Creation	N
internal	Creator Id	The ID of the Gooru user who added the resource.	Creation	Y
primary	Learning objectives	The learning objectives to be covered in the lesson	Educational	Y
secondary	Lesson Plan	Defines how the lesson has to be taught. Usually over multiple sessions with each session covering specific collections and assessments from the lesson. Here the reference to the plan is stored	Instructional	N

Units

Metadata Type	Name	Description	Group	Mandatory
primary	Title	The title of the resource.	Descriptive	Y
primary	Publisher	The organization credited with publishing the resource (i.e. who the content belongs to).	Creation	Y
primary	Standards	The specific standard code(s) aligned to the content.	Framework	Y
primary	Subject	Subject(s) to which the activity is aligned	Framework	Y
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework	Y
primary	Domain	The taxonomy domain to which the activity is aligned	Framework	Y
primary	Big Ideas	2-3 Sentences stating the key concepts and practices learned in the Unit	Educational	Y
primary	Essential Questions	3-5 questions that align with the goals/big ideas of the unit	Educational	Y
primary	Pacing guide	The recommended duration in which to complete the activity	Educational	Y
secondary	Audience	The audience for which the unit is intended.	Student	Y

primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other	Y
secondary	sequence	The sequence of the unit in the context of a course		N
secondary	Aggregator	The organization that catalogues content from external publishers (i.e. who catalogues publisher content).	Creation	N
secondary	License		Creation	Y
secondary	Time Required	Approximate time required to study the unit.	Educational	N
secondary	Keywords	Vocabulary and key terms associated with the resource	Media	N
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal	Y
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other	N
internal	Visibility	The visibility setting of the resource. Determines whether the resource will be displayed in search.	Media	Y
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal	Y
internal	Flagged		Internal	Y
internal	Created	The date and time when the resource was created in Gooru.	Creation	Y
internal	Date Modified	The date and time when any of the resource metadata was modified in Gooru.	Creation	N
internal	Modified By	User id of user who last modified the resource.	Creation	N
internal	Creator Id	The ID of the Gooru user who added the resource.	Creation	Y
	Unit plan	The summary of the plan using which the unit needs to be taught. It is a free text field or an uploaded doc along with the aggregated lesson plans for all the lessons in the unit	Instructional	N

Courses

Metadata Type	Name	Description	Group
---------------	------	-------------	-------

primary	Publisher	The organization credited with publishing the resource (i.e. who the content belongs to).	Creation
secondary	Aggregator	The organization that catalogues content from external publishers (i.e. who catalogues publisher content).	Creation
secondary	License		Creation
internal	Created	The date and time when the resource was created in Gooru.	Creation
internal	Date Modified	The date and time when any of the resource metadata was modified in Gooru.	Creation
internal	Modified By	User id of user who last modified the resource.	Creation
internal	Creator Id	The ID of the Gooru user who added the resource.	Creation
primary	Title	The title of the course.	Descriptive
primary	Course goals		
secondary	Audience	The audience for which the course is intended.	Y
primary	Standards	The specific standard code(s) aligned to the content.	Framework
primary	Subject	Subject(s) to which the activity is aligned	Framework
primary	Taxonomy Course	The taxonomy course to which the activity is aligned	Framework
primary	Domain	The taxonomy domain to which the activity is aligned	Framework
internal	Id	Unique identifier for a resource; used internally to identify a resource	Internal
internal	Deleted	Whether the resource has been deleted in the product (soft delete).	Internal
internal	Flagged		Internal
secondary	Keywords	Vocabulary and key terms associated with the resource	Media
internal	Visibility	The visibility setting of the resource. Determines whether the resource will be displayed in search.	Media
primary	Teacher Resource	Identify if the activity is only for the teacher and does not show up for the student	Other
internal	Tags	Miscellaneous metadata that doesn't fit into the other data fields.	Other
secondary	Use case		

primary	Course plan	A summary of the plan for teaching the course. It is a free text entry or an uploaded doc	
---------	-------------	---	--