

BIG-R BAA
Topic # 008 - Spotting Objects from Weak Labels (SPECTRE)

Document Change History

Date	Description
05/31/2022	Draft Topic 8 Posted
07/07/2022	Final Topic 8 Posted (Updated Sections 'Expected Awards', 'Budget', and Sections 4, 5, 7, 12)

Title

Spotting Objects from Weak Labels (SPECTRE)

Expected Awards

The government plans to make two awards, but may make more than two awards, one award, or none. Contract award under this topic is subject to the availability of funds.

Budget

The planned budget for each phase is approximately

Base Period: \$1.6M for 1 award (or \$800k each for 2 awards)
Option Period 1: \$1M for 1 award (or \$500k each for 2 awards)
Option Period 2: \$300k for 1 award (or \$150k each for 2 awards)

Funding may be increased or decreased at the sole discretion of the Government.

Topic Technology Readiness Level

3 (Base Period goal) to 5 (Option Period 2 goal).

This topic is limited to projects that meet technology readiness level (TRL) definitions in the TRL range 1-5. Upon completion of proposed developmental efforts, solutions should strive to meet a TRL of 5.

Dates

Draft Topic Posting	05/31/2022
Proposer's Day Registration Deadline	06/10/2022 @ 11:00am EDT
Proposer's Day	06/13/2022
Questions Due	06/22/2022 @ 5:00pm EDT
Final Topic and Q&A Posting	07/07/2022
Abstracts Due	07/21/2022 @ 5:00pm EDT
Abstract Feedback	08/11/2022
Proposals Due	09/06/2022 @ 5:00pm EDT

Point of Contact (POC)

Please send all questions and submissions to the Contracting Officer (CO): Daniela Garavito

Email Address: BigRBAA@nga.mil

→ Use the following email subject for all email correspondence: **‘TOPIC 8 - SPECTRE’**

Proposer’s Day Information

The government hosted a “SPECTRE Proposer’s Day” in conjunction with this topic to outline challenges, concerns, and expectations to potential proposers. In addition, this event intended to provide potential proposers with the opportunity to identify potential teaming partners.

Post-Event Slack Chat Room: To allow interested vendors to communicate with each other for teaming purposes, NGA is providing a Slack Chat Room.

Please note that NGA/Government personnel do not have access to the chat room and will consequently neither monitor nor get involved in any of the discussions.

Link to Chat Room: <https://join.slack.com/t/topic8teamingroom/signup> (alternate signup: https://join.slack.com/t/topic8teamingroom/shared_invite/zt-1b3zbfmn4-88Zh2u0tOAIMh9NetEi6uQ)

Event Recording: A complete recording of this event is accessible at <https://nga.webex.com/nga/ldr.php?RCID=0a428180b90ddbebac338547aef6314f> (passcode: BAATopic8)

1. Topic Research Opportunity Description

The National Geospatial-Intelligence Agency Research Directorate (NGA Research) is soliciting proposals for pioneering research in weakly supervised object detection [1] using text and remote sensing imagery.

This topic call seeks to find breakthrough solutions for lowering data acquisition risk and cost for computer vision, particularly in unstructured data. Proposals should investigate novel methods to advance the state-of-the-art in automatically extracting training data from a variety of structured and unstructured text captions.

1.1 Outside of the Scope of this Topic

Specifically excluded is research that results in manual-labeling of images, and any purely heuristics-based labeling functions [2] or rule-based knowledge bases [3] that do not substantially advance the state-of-the-art. Research resulting in only gradual improvements to the state-of-the-art are out of scope for this solicitation.

2. Statement of Need(s)

The goal of this research is to significantly advance the state-of-the-art in weakly supervised object detection techniques.

Proposals should address the following research elements in pursuit of joint text and image processing with weak supervision:

- Automated methods of curating data sets from structured and unstructured (narrative) text captions, vice existing manual processes
- Multi-class weakly supervised object detection in large-scale remote sensing imagery.

Generative and projective models appear to be robust to variances in data sources. They generally outperform heuristics-based models, which are often limited by information coarseness [4, 5] from the data set, and inability to generalize beyond supervised methods [6, 7, 8]. Other state-of-the-art solutions utilize a paradigm similar to teacher-student methods [9, 10, 11, 12] though those solutions also suffer from similar limitations as heuristics-based models [13, 14, 15].

3. Project Structure

The research will consist of three periods – a base period and two option periods, with key milestones that determine advancement to the next period. The key technical elements are described below:

3.1 Technical Elements

3.1.1 Technical Element 1

Foundations of Natural Language Processing (NLP) (Base Period, Option Period 1) This technical element seeks methods to extract computer vision training data from existing, narrative text captions. This narrative text will include information such as object locations and counts, but may not fully specify all object attributes. For example, object locations might be exact coordinates, rough or relative descriptions, or not provided. Text extraction algorithms must be capable of extracting clear context from English text. Bidirectional encoder representations from transformers (BERT) and generative pre-trained transformer 2 (GPT-2) are the current state-of-the-art for NLP. Generative or projective language models are within the scope of acceptable research, though stemming or heuristics-based language models should be avoided unless a strong justification with a novel approach is made.

3.1.2 Technical Element 2:

Foundations of Computer Vision (Base Period, Option Period 1) This element focuses on the localization of individual objects of interest from multiple classes using weak supervision. For example, state-of-the-art approaches might use labels at the image-level to characterize the data distribution [16]. However, many of those methods still rely on the creation of image-level labels by hand, which is not relevant for this effort: detectors should primarily utilize information extracted from text, as stated. Proposed methods may utilize pre-trained models from model zoos, but should not solicit any new human labels during the pre-training process.

The scale of the imagery itself should also be considered. Methods must be compatible with imagery on the order of tens to hundreds of megabytes (MB) in size, with each target object occupying no more than 0.5% of the raw image's total pixel area. Many existing approaches mitigate this by dividing imagery into chips and using multiple-instance-learning [17], but any method capable of dealing with full-scale remote sensing imagery is relevant here.

3.1.3 Technical Element 3:

Extensions to Remote Sensing (Option Period 1) – Potential text sources for weak labels of remote sensing images present unique challenges due to the scale of the imagery and nature of descriptions for overhead imagery.

Both structured and unstructured descriptions of remote sensing imagery often describe objects in a hierarchical manner. For example, a fighter jet might be described as an “aircraft” if it is impossible to visually distinguish from a commercial aircraft. Methods under this technical

element should leverage known semantic information to classify objects within such a hierarchy: standard multi-class approaches are inadequate. Some existing methods use known semantic hierarchies to detect objects at varying levels of detail, but have not been applied to remote sensing imagery [18].

Another key concern is a lack of hard negative labels in unstructured text. Since text data is typically not created with machine learning in mind, the absence of information about an object does not necessarily correlate with the absence of that object. Thus, standard methods for negative sampling are not immediately applicable here [19].

3.2 Program Periods

The goals of the research should be focused on performance while maintaining accuracy and robustness relative to the diversity of data.

3.2.1 Base Period (12 months)

The Base Period will focus development of critical capabilities for a complete end-to-end weakly supervised object detection training scheme. At a minimum, Base Period plans should address Technical Elements 1 and 2. English will be the language used for validation and verification of text extraction algorithms. Finally, both text and image solutions must be combined at the end of this period to demonstrate an ability to perform weak supervision compliant with the performance expectations as identified in Section 5. Metrics.

3.2.2 Option Period 1 (12 months)

Option Period 1 will focus on extending the algorithms developed in the Base Period to the unique challenges of remote sensing data. Proposals should address Technical Element 3 within their Option Period 1 plan, clearly describing improvements to elements 1 and 2 resulting from this additional work, such as the extension of text extraction algorithms to domain-specific vocabulary and semantics, the scalability of computer vision algorithms to large-scale imagery, and hierarchical classification of individually detected objects.

3.2.3 Option Period 2 - Transition (6 months)

Option Period 2 will focus on maturing the technological readiness level of algorithms developed in Option Period 1 to technological readiness level 5. This includes adapting algorithms to interface specifications that will be provided by the government at the beginning of this period (see [Section 7](#) Government Furnished Information) and may include demonstrating performance on government-provided data sets. Access to controlled unclassified information (CUI) will be required beginning with this period of the project. Any contractor personnel working with CUI information must receive a favorable HSPD-12 and/or HSPD-12 Tier 1 adjudication prior to accessing CUI information. See section 9 Security Information for additional detail regarding this requirement.

4. Team Expertise

This topic encourages collaborative relationships and multidisciplinary teaming between academic and industry partners. The development and delivery of weakly supervised object detection techniques will likely require expertise in several relevant technical fields, including computer vision, machine learning, remote sensing, and natural language processing. This list is included only to provide guidance for the proposers; satisfying all the areas of technical expertise is not a requirement for selection, nor is partnering with another organization.

5. Metrics

Algorithms and/or solutions resulting from this research will be measured against object detection and classification performance. Specifically, object detection performance will be measured using the standard mean average precision (mAP) and recall metrics (where a successful detection is one with a center point within 3 meters of the center point of a ground truth label) in both the Base Period and Option Period 1. Proposers may utilize their preferred method of detection, which includes any form of localization that would be useful for a user to understand where objects are located (for example, image segmentation, bounding box prediction, or center point prediction). Proposers should clearly describe in their Verification and Validation plan (see Section 6, Table 2) how algorithm outputs can be mapped to point detections for the purposes of evaluation against the metrics below.

In the Base Period, object classification performance will be measured with the F_1 score, as in a standard multi-class classifier. In Option Period 1, when algorithms will be expected to respect a given taxonomy, object classification performance will be measured using the semantic distance of the predicted class label from the true class label, as defined using a provided taxonomy and the measure defined in [20] ("Semantic Similarity in a Taxonomy").

Data sets will be provided in each period for proposers to track performance, as specified in [Section 7](#) Government Furnished Information. The government will utilize private splits of these data sets to measure performance against the metrics below, using an independent validation and verification (IV&V) protocol. Proposers are free to define additional data sets, but should not use the labels of the original xView 1 data set in any training, pretraining, or model development. Additional data sets may be used either to augment training or to demonstrate a facet of performance that the provided xView 1-derived dataset may not allow for. If being used to augment training, the dataset must be publicly available (e.g., ImageNet) or labeled without manual intervention (e.g., images and captions scraped from a public blog).

The Government will consider data sets from performers that demonstrate the breakthroughs beyond the current state-of-the-art. Any proposed datasets must be representative of the challenge of using text for weak supervision, and so must include captions as training data. Neither the validation nor the test dataset will be expected to

contain text captions. Nominated data sets will not be shared between performers unless approved by the nominating performer. Competency and understanding of the nominated data sets is expected of the performers, and performers should keep in mind that all independent validation and verification will be done using the Government provided data set described in [Section 7](#) Government Furnished Information.

Table 1 Metrics

Metrics	Base Period	Option Period 1
Object Detection	Mean Average Precision (mAP)	
	Threshold: 0.25 Objective: 0.45	Threshold: 0.35 Objective: 0.60
	Recall	
	Threshold: 0.70 Objective: 0.80	Threshold: 0.75 Objective: 0.90
Object Classification	F_1 -score	Semantic Distance [20]
	Threshold: 0.70 Objective: 0.80	TBA

6. Deliverables

The proposer shall support a project kickoff briefing providing a detailed schedule and experimental research plan as well as mid-point and close out briefings that provide a quantitative assessment of progress on key performance metrics.

Table 2 Schedule of Technical Deliverables

ITEM	DESCRIPTION	DUE DATE
Technical Exchange Briefing	Live briefing and associated materials that detail the experimental approach, internal schedule with expected milestones, and execution plan.	15 Days after beginning of Base, Option Period 1, and Option Period 2
Monthly Status Report (MSR)	Provide a brief summary of work accomplished, any challenges or issues that may impact cost/schedule/performance or needs input, and brief summary of intended actions for the next reporting period. (Not anticipated to exceed one to two pages)	Monthly (Base, Option Period 1, and Option Period 2)
V&V Plan	Describes the methodology for evaluating the performance of the proposed solution, and any required data and code that will need to be provided to the independent, third-party, validation and verification team at Month 9 (see <i>Algorithm(s)</i> below).	Month 6 (Base & Option Period 1)

Midpoint Status Report	<p>Live briefing and associated materials to inform the Government of the current technical progress, communicate any impediments or limitations, anticipated difficulties, and present any preliminary discoveries from the current research.</p> <p>In Option Period 2, this should include a live demonstration of the system to the Government.</p>	<p>Month 6 (Base & Option Period 1)</p> <p>Month 3 (Option Period 2)</p>
Algorithm(s)	<p>Algorithms for the extraction of computer vision labels from text (Technical Element 1) and the use of those labels for object detection (Technical Element 2). In Option Period 1, this includes updated algorithms addressing Technical Element 3. To be delivered to the government as source code and in a dockerized container for implementation in a commodity cloud environment.</p> <p>The performer will be expected to have all necessary code and data available to the Government for an initial evaluation of performance at Month 9, and again for a final IV&V evaluation at Month 11.</p>	<p>Prototype algorithms: Month 6 (Base & Option Period 1)</p> <p>IV&V ready algorithms: Month 9 and 11 (Base & Option Period 1)</p>
Algorithm Description Document (ADD)	To aid in validation and capability integration, provide a detailed ADD and signal model to describe how the detections are created through processing. To be delivered with each algorithm.	<p>Month 9 (Base & Option Period 1)</p> <p>Month 5 (Option Period 2)</p>
Final report & Debrief	Final technical report and live briefing summarizing the work performed and the final results of the investigation.	Up to two weeks before the end of each contract period (Base, Option Period 1, and Option Period 2)
Finalized Algorithms	Finalized, documented algorithms for all three technical elements in a unified system, at technical readiness level 5 and conformant to interface specifications, which will be provided by the government at the start of Option Period 1.	Month 6 (Option Period 2)

7. Government Furnished Information

The government-provided Base Period data set will be a modified version of the xView 1 data set [21], with a training split of imagery and corresponding structured and unstructured text, and a validation split with ground-truth object center points and classes. Proposers can expect text captions on the order of 10 sentences, containing descriptions of object presence, object counts, and spatial relationships, among other information. Not all types of information will be present in each caption. The data will contain noise and complexity, particularly in terms of how objects are described and what objects the annotator chose to describe, that does not exist in a dataset weakly annotated by hand for the express purpose of computer vision. All text data will be provided in a plain text format, and no annotations will be present in the imagery itself: there will be no requirement for optical character recognition. This dataset will be delivered at the beginning of the Base Period.

Data format will be identical to the xView 1 dataset, with identical preparation: ortho-rectification, georeferencing, pan-sharpening, and 0.3-meter ground sample distance. Each image in the training dataset will be labeled with an unstructured text caption reflective of the challenges in Technical Elements 1 and 2. Bounding boxes will not be given, but object center points will be provided for the validation dataset as well as a subset of the training dataset (the distribution of this subset will not be uniform: some images and classes will be labeled with text only, and some will contain a combination of text and center points). The breakdown of this data is detailed below:

Table 3 Provided Dataset

Data Split	Number of Images	Number of Objects	Number of Objects with Center Points
Training	847	600,345	25,000
Validation	282	200,291	200,291
Testing	284	-	-

The Option Period 1 data set will also be a modified version of the xView 1 data set, but the text data will be updated to reflect the challenges described in Technical Element 3. This dataset will be delivered at the beginning of Option Period 1.

The objects of interest will be clearly defined within the dataset, and will not include objects outside of the 60 original xView 1 classes. A semantic taxonomy of the objects within the data set will also be provided for both the Base Period and Option Period 1, so that proposers may prototype (and, in Option Period 1, fully implement) hierarchical classification algorithms described under Technical Element 3.

In Option Period 2, the government will provide interface specifications at the beginning of the performance period that the final system will be expected to conform to at the conclusion of Option Period 2.

8. Intellectual Property

NGA anticipates that achieving the goals of the program will necessitate a minimum of government purpose rights in all deliverables. Furthermore, NGA anticipates that the proposer will secure to the government a paid license of suitable scope for any third-party software incorporated into contract deliverables. Notwithstanding these assumptions, the government acknowledges the possibility that other approaches proposed by proposers may also achieve program goals.

9. Security Information

9.1 Unclassified Work Performance Security Requirements

- 9.1.1 Uncleared contractor personnel are authorized to work on this contract up to the unclassified level, with access to Department of Defense (DoD) controlled unclassified information (CUI) at the contractor site without the requirement of a security clearance.
- 9.1.2 Any contractor personnel working with CUI information must receive a favorable HSPD-12 and/or HSPD-12 Tier 1 adjudication prior to accessing CUI information. Contractor personnel who require access to CUI for 60 days or less must receive a favorable HSPD-12 adjudication. Contractor personnel who require CUI access for greater than 60 days must receive a favorable HSPD-12 Tier 1 adjudication.
- 9.1.3 NGA will sponsor the HSPD-12 and HSPD-12 Tier 1 background investigation for required program personnel. NOTE: Contractor personnel submitted for SCI access cannot be submitted for a HSPD-12 or HSPD-12 Tier 1 adjudication while waiting for their SCI approval.
- 9.1.4 Foreign nationals are not permitted to perform unclassified work under the terms of this contract.
- 9.1.5 Contractor personnel shall not release any unclassified information, regardless of medium (e.g. film, tape, document), pertaining to any part of this contract or any program related to this contract, unless the contracting office representative (COR) has given prior written

approval or in performance of a project that has been scoped and negotiated by NGA.

9.1.6 Contractor personnel visiting NGA facilities and/or sites will receive the appropriate visitor badge and be escorted, as appropriate. The visitor badge will be returned at the end of each visit day. NOTE: NGA reserves the right to refuse access to any personnel.

9.1.7 Contractor personnel are forbidden from bringing in prohibited, unauthorized, and/or portable electronic devices (PEDs) items into any NGA installation or any secure office/working location covered under this agreement. A list of PEDs includes but is not limited to cell phones, cameras, two-way pagers, laptops, recorders (e.g. digital, tape, etc.), flash drives, or any other kind of removable media, without prior approval and approval paperwork from NGA. See NGA instructions/regulations/policy for a full list of prohibited and unauthorized items. Security violation repercussions will be determined on the severity of the violation.

9.2 Information Handling

9.2.1 Contractor personnel will comply with the NGA, DoD, and intelligence community (IC) policies and regulations (to include, but not limited to, the Consolidated NGA (CoNGA) Security Classification Guide) to properly mark (to include portion marking) classified and unclassified documentation, media, etc.

9.2.2 Document markings will be in accordance with the lowest security classification possible to ensure the confidentiality and integrity for the greatest release to partners in accordance with NGA and mission partner marking guides for classified information.

9.2.3 All government-furnished information released to the contractor or created in the performance of this contract will be destroyed or returned by the contractor to NGA upon contract termination or when no longer required for contract performance. The determination to destroy or return will be at the direction of the NGA contracting officer (CO) or COR.

10. Proposal Requirements

In addition to the proposal submission requirements of the BIG-R BAA General Solicitation, Rev5, proposers shall include the following information within

Proposal Volume 2:

Proposals shall clearly address all technical elements (see Section 3.1) for each period (see Section 3.2).

Further, each proposal submitted in response to this topic shall contain a preliminary validation and verification (V&V) plan that describes the methodology for evaluating the performance metrics of the proposed solution. Proposers will have the opportunity to finalize this V&V plan post-award.

The preliminary V&V plan will not count towards the page limitation for Volume 2.

Proposal Volume 4:

Proposers are reminded that the submission of a detailed List of Data Right Restrictions/Assertions (Commercial and Non-Commercial) – per General Solicitation, Rev 5, p. 36/37 (Part IV, Section 6.3.4, Paragraph iii) is required as part of Proposal Volume 4.

11. Place of Performance

Research activities outlined in this proposal will be conducted at the performers' location(s).

12. Optional Performance

If the government determines that end results and overall performance within a given period warrant continued research, then the government may exercise subsequent option periods.

Optional periods will be exercised based on quantitative evaluations, such as achievement of performance metrics, and quantitative metrics, such as overall assessment of base product, computational tractability, potential benefits of exercising the option, and availability of funding. Note that not exercising an option may not be solely based on performance.

There will be no explicit weighting between qualitative and quantitative evaluations.

The government may, at its discretion, transition the program to a follow-on effort. The results of the program at its completion will determine whether a follow-on program is warranted.

315

13. Questions & Answers (Q&A)

316

#	Topic 8 Reference	Question	Government Response
1	Topic Cover Page	Is there a not to exceed contract value associated with BIG-R Topic 8?	Yes, see updated language on Topic cover page: “Budget”.
2	HM0476-20-BAA-0001: Part III, 1.0 Eligible Applicants	Will you consider submissions from the robotics community?	All responsible sources capable of satisfying the Government's needs may submit a proposal for NGA's consideration.
3	Section 9.1.4 Unclassified Work	Are US citizens only allowed to work and respond to the BAA? Does that exclude US permanent residents?	The Topic 8 prohibits participation of non-US citizens.
4	HM0476-20-BAA-0001, Rev5: Part III, 1.0 Eligible Applicants, and 1.1.1 FFRDCs	Could you please confirm any specifics about the FFRDCs and whether they are allowed to respond to the BAA?	All responsible sources capable of satisfying the Government's needs may submit a proposal for NGA's consideration. FFRDCs are subject to applicable direct competition limitations and cannot propose to this topic in any capacity unless they meet the conditions listed in BIG-R BAA General Solicitation, Part III, Section 1.1.1.
5	Section 7 GFI	Does the government have practical examples of sources for test captions that can be discussed with proposers?	38North (https://www.38north.org/topics/satellite-analysis/) is an unclassified, open source imagery analysis organization whose articles are similar to the data that will be provided for this work. Their satellite imagery analysis articles are longer and more news-like than the captions will be for this effort, but are generally representative of the syntax and grammar performers can expect. Shorter articles from 38North are particularly representative.
6	Section 5 Metrics Section 7 GFI	Does the government expect performers to propose potential sources for test captions, for example for Technical Element 3?	As stated in Section 7, validation/test data will not contain captions. “The Government will consider data sets from performers that demonstrate the breakthroughs beyond the current state-of-the-art. Any proposed datasets must be representative of the challenge of using text for weak supervision, and so must include captions as training data. Neither the validation nor the test dataset will be expected to contain text captions.” (See updated language in Section 5.)
7		What is the status of the SAFFIRE program and does its development impact SPECTRE?	Due to classified efforts being outside the scope of this topic, the Government will not discuss SAFFIRE.
8	Section 1.1 Out of the Scope	Section 1.1 of the draft RFP reads “Specifically excluded is research that results in manual-labeling of images...”. If a state-	Yes, it is disqualified. (See Section 1.1

		of-the-art solution requires an intermediary manual-labeling step, is that solution disqualified?	
9	Section 3.2.1 Base Period	For the NLP portion, are we focusing on English, or will there be other languages represented?	“English will be the language used for validation and verification of text extraction algorithms.” (See Section 3.2.1)
10	Section 5 Metrics & 7 GFI	There is a dataset to evaluate technical elements 2 and 3, but how will technical element 1 be evaluated?	Technical Element 1 involves the extraction and/or representation of information from the text component of the dataset to enable the computer vision components in Technical Element 2. Thus, the Government expects that any methods for Technical Element 2 will be heavily dependent on the quality of methods developed for Technical Element 1. The metrics and dataset defined in Sections 5 and 7 deal explicitly with Technical Element 2, but the interdependence of the elements means that any evaluation of Technical Element 2 will also reflect performance under Technical Element 1. (See updated language in Section 5 and 7.)
11	Section 5	You explicitly mention object detection and object classification. Can you describe what you envision as the object classification experimental setup?	“In the Base Period, object classification performance will be measured with the F1 score, as in a standard multi-class classifier. In Option Period 1, when algorithms will be expected to respect a given taxonomy, object classification performance will be measured using the semantic distance of the predicted class label from the true class label, as defined using a provided taxonomy and the measure defined in [20] (“Semantic Similarity in a Taxonomy”).” (See Section 5)
12	Section 7 GFI	Can you give more examples of how this corpus of existing data differs from manually labeled data for weak supervision?	Since GEOINT is made for policymakers and warfighters, not computer vision, this corpus of existing data is unlike manually labeled data for similar weak supervision tasks. The existing data will contain noise and complexity, particularly in terms of how objects are described and what objects the annotator chose to describe, that does not exist in a dataset weakly annotated by hand for the express purpose of computer vision.
13	Section 7 GFI	Is the distinction that manually labeling has exact counts while what you are looking at will have less count information and free-flowing natural language with class noise?	Technical Element 3 describes many of the unique distinctions. A lack of count information, free-flowing (narrative) language, and class noise are all important, but not comprehensive in terms of the challenges that may be encountered. As noted in the response to question 12, the distinctions do not simply come from manual vs. non-manual labeling, but rather from the fact that existing data was labeled for purposes that may not align with an ideal computer vision dataset.
14	Section 7 GFI	Are we assuming documents have been parsed? Do we need to pull labels out of large analysts word documents, PDFs, and powerpoints?	“All text data will be provided in a plain text format, and no annotations will be present in the imagery itself: there will be no requirement for optical character recognition.” (See updated language in Section 7)
15	Section 1.1 Outside of the Scope.	Are you interested in weakly supervised methods that work with manually created high-level labels?	Any manual labeling is out of scope for this effort. (See Section 1.1)
16	Section 7 GFI	Are the objects or categories of objects of interest known apriori (noise or completeness issues notwithstanding)?	“The objects of interest will be clearly defined within the dataset, and will not include objects outside of the 60 original xView 1 classes.”

			“A semantic taxonomy of the objects within the dataset will also be provided for both the Base Period and Option Period 1.” (See updated language in Section 7)
17	Section 5 Metrics	Is image segmentation a valid approach? What would be the metrics in that case?	“Proposers may utilize their preferred method of detection, which includes any form of localization that would be useful for a user to understand where objects are located (for example, image segmentation, bounding box prediction, or center point prediction). Proposers should clearly describe in their Verification and Validation plan (see Section 6, Table 2) how algorithm outputs can be mapped to point detections for the purposes of evaluation against the metrics in Section 5.” (See updated language in Section 5)
18	Section 5 Metrics	You mentioned that high-level labels and aliases are a big problem. How is the dataset designed to assess that? For example, if one label has four different alias and one high-level category, what do you expect the "correct" output to be?	The intention is for the natural language processing to fuse or map the extracted labels to the provided taxonomy (that is, to map the four aliases to one standard label), and for the final output to be as correct as possible within that hierarchy. A metric has been defined within the topic (see Section 5) that provides a numerical error value for classification within a hierarchy, but it is up to proposers to define how they wish to push their algorithm towards a more correct output.
19	Section 12 Optional Performance	The BAA references specific quantitative metrics to evaluate solutions, but you mentioned some program goals that may not be well covered by these metrics. Will there be any qualitative evaluation of solutions? If so: What might that look like? Will it involve analysts in the loop? How will qualitative evaluations be weighted compared to quantitative metrics?	Yes, “optional periods will be exercised based on quantitative evaluations, such as achievement of performance metrics, and quantitative metrics, such as overall assessment of base product, computational tractability, potential benefits of exercising the option, and availability of funding. Note that not exercising an option may not be solely based on performance.” There may be analyst/user evaluation during the Transition Period. “There will be no explicit weighting between qualitative and quantitative evaluations.” (See updated language in Section 12.)
20		Will the dataset be publicly available down the road like xView is?	No. There is no intent for the Government to make this dataset publicly available.
21	Section 3.1.2 Tech Element 2	What role do you see pre-trained models playing here? Do prefer approaches that have little to no pre-training using classical weak/strong labels?	The use of pretrained models for remote sensing and other computer vision tasks is well-studied and relevant in the context of this effort. There is no preference between strong or weak labels, or between using or not using pre-training, so long as models are not pre-trained using the original xView 1 labels. (See Section 3.1.2.)
22	Topic Cover Page: Budget	Is there a set budget for this project?	See updated language on Topic cover page: “Budget”.
23	Section 7 GFI	Is there a text component at validation/inference time? Or just an image?	No, as mentioned in Section 7, the validation data includes only “ground-truth object center points and classes” (See Section 7)

24	Section 7 GFI	Given the data will be a subset of xView, roughly what fraction will include the unstructured text (or structured data)?	See Section 7, Table 3 (“...objects with center points” refers to structured data.)
25	Section 4 Team Expertise	Do you require academic partners on a team?	No. “This topic encourages collaborative relationships and multidisciplinary teaming between academic and industry partners,” but no form of partnership is required. (See updated language in Section 4.)
26	Section 7 GFI	It was mentioned that the text descriptions would be a few paragraphs. Is that correct? Or did you mean a few sentences? I imagine the average length of a paragraph to be 4-6 sentences, so it seems like a lot.	“Proposers can expect text descriptions on the order of 10 sentences.” (See updated language in Section 7)
27	Section 1 1. Topic Research Opportunity Description	Is state-of-the-art considered any approach that is published or currently available?	Yes. Because the concept of using GEOINT as weak supervision is novel, the Government is unaware of existing approaches that comprehensively address all aspects of this effort. Thus, any existing approaches represent the state-of-the-art that proposers should improve upon.
28	Section 7, Lines 195-196.	Will there be object classes outside of the 60 categories of the original xView dataset?	“The objects of interest will be clearly defined within the dataset, and will not include objects outside of the 60 original xView 1 classes.” (See updated language in Section 7.)
29	Section 3.1.3 Tech Element 3	‘Negative label’ is a term that has multiple disparate definitions. Can the government define ‘negative label’ in this context and provide examples of a negative label and the desired utilization of this labeling type?	‘Negative label’ refers to any indication that an object of interest is not present. For example, “no aircraft are observed in the image.” The challenge for this effort is not in utilizing this label type, but in creating models that can respect the fact that it will often be unavailable (the absence of information about an object does not necessarily correlate with the absence of that object).
30	Section 7 GFI	Can we get example images and captions?	See question 5.
31	Section 7 GFI	Do the captions indicate spatial relationships?	Yes. “Proposers can expect text captions ... containing descriptions of object presence, object counts, and spatial relationships, among other information. Not all types of information will be present in each caption.” (See updated language in Section 7)
32	Section 7 GFI	Does the validation set have captions?	No, see question 23.
33	Section 7 GFI	Will images be annotated like an imagery analysis product (e.g. have callout boxes, directional arrows, etc.)?	No, see question 14.
34	Section 7 GFI	Should vehicle detection be limited to just military or does the government wish to include civilian as well?	The data set for this effort will be a modified version of the xView 1 data set, which consists primarily of civilian vehicles and buildings.

35	Section 7 GFI	Will the government provide a desired format for the labels and image captions?	The dataset, including all relevant data format information, will be delivered at the beginning of the Base Period. (See Section 7.)
36	Section 7 GFI	Will the government provide bounding boxes for the objects?	“Bounding boxes will not be given, but object center points will be provided for the validation dataset as well as a subset of the training dataset.” (See Section 7)
37	Section 7 GFI	What is the anticipated format of images included in the government-provided datasets? Are the images georeferenced?	“Data format will be identical to the xView 1 dataset, with identical preparation: ortho-rectification, georeferencing, pan-sharpening, and 0.3-meter ground sample distance.” (See updated language in Section 7)
38	Section 3.2.1 Base Period	Is it required for submissions to include innovations in both NLP and computer vision, or is a focus on one area sufficient?	At a minimum, proposals should address Technical Elements 1 and 2. This includes both computer vision and natural language processing. (See Section 3.2.1.)
39	Section 7 GFI	It's mentioned in the Draft BAA that 25k of the 600k objects in the training data set are labeled with center point indicators. Are these 25k points distributed uniformly across classes and separate images? Are some classes/images in the training set completely unlabeled?	“... object center points will be provided for the validation dataset as well as a subset of the training dataset (the distribution of this subset will not be uniform: some images and classes will be labeled with text only, and some will contain a combination of text and center points).” (See updated language in Section 7)
40	Section 5 Metrics	Also mentioned in the Draft BAA is the fact that the validation dataset is labeled with center points and that recall metrics are based on a 3-meter center point threshold. Is the desired output of a proposed solution only a center point label, or are other types of localization desired so long as they can be reduced to a center point for evaluation?	See question 17.
41	Section 7 GFI	The original xView dataset contains images that span large areas and/or are densely populated with targets. Will the 5-10 sentences of unstructured text associated with each image usually refer to random groups of objects anywhere within the image? Or will the text usually refer to objects clustered in a specific region of interest within the larger image?	The dataset will include examples of both cases.
42	Section 12 Optional Performance	Are there any memory, speed, or hardware requirements? (Does the model have to analyze a certain number of images per minute?)	No. There are no hardware requirements for this effort, however, computational tractability is an element of the qualitative evaluation that will determine advancement to the Option Period. (See updated language in Section 12.)
43	Section 5 Metrics	The BAA states "Proposers are free to define additional data sets, but should not use the labels of the original xView data set in any training or pretraining." Are we allowed to use the original xView data for other purposes, such as testing and validation? For	No. “Proposers are free to define additional data sets, but should not use the labels of the original xView 1 data set in any training, pretraining, or model development.” (See updated language in Section 5)

		example, if our model were to generate object bounding boxes as an intermediate step, could we use existing xView bounding boxes to test that processes?	
44	Section 7 GFI	Are the distribution of object types similar to that in the original xView dataset?	Yes, but the data set may not necessarily include all object types from the original xView 1 data set. (See updated language in Section 7.)
45	Section 5 Metrics	Lines 140-142 of the published Draft_BIGR_BAA_Topic+8_05312022 document state: “Proposers are free to define additional data sets, but should not use the labels of the original xView data set in any training or pretraining.” What is meant by “define additional data sets?” Will NGA allow non-xView datasets to be introduced into the overall solution and, if so, what are the limitations, if any, to those non-xView datasets?	“Define additional data sets” refers to proposers being permitted to introduce non-xView 1 datasets into their solution, either to “augment training or to demonstrate a facet of performance that the xView 1-derived dataset may not allow for. If being used to augment training, the dataset must be publicly available (e.g., ImageNet) or labeled without manual intervention (e.g., images and captions scraped from a public blog).” (See updated language in Section 5)
46	Section 9.1.4 Unclassified Work	Section 9.1.4 (Line 215-216) of the DRAFT states: “Foreign nationals are not permitted to perform unclassified work under the terms of this contract.” Can you please state whether U.S. persons are allowed or if U.S. citizens are required to perform work on this project?	See question 3.
47	HM0476-20-BAA-0001: Part III, 1.0 Eligible Applicants	Can hardware companies submit alone if we have some unique value to add to these specific use-cases?	See question 2.

318 **14. References**
319

- [1] P. Varma, F. Sala, A. He, A. Ratner and C. Re, "Learning Dependency Structures for Weak Supervision Models," 2019. [Online]. Available: <https://proceedings.mlr.press/v97/varma19a/varma19a.pdf>. [Accessed Dec 2021].
- [2] A. Awasthi, S. Ghosh, R. Goyal and S. Sarawagi, "Learning from rules generalizing labeled exemplars," in *International Conference on Learning Representations*, 2020.
- [3] W. Xu, R. Hoffmann, L. Zhao and R. Grishman, "Filling knowledge base gaps for distant supervision of relation extraction," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013.
- [4] L. Oakden-Rayner, J. Dunnmon, G. Carneiro and C. Ré, "Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging," in *ACM Conference on Health, Inference, and Learning (ACM CHIL '20), April 2-4, 2020, Toronto, ON, Canada, 2020*.
- [5] C. P. Phoo and B. Hariharan, "Coarsely-Labeled Data for Better Few-Shot Transfer," in *Proceeding of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Virtual/Online, 2021.
- [6] Y. Wang, Q. Yao, J. T. Kwok and L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-Shot Learning," 29 March 2020. [Online]. Available: <https://arxiv.org/pdf/1904.05046.pdf>.
- [7] R. Roelofs, "Measuring Generalization and Overfitting in Machine Learning," 19 June 2019. [Online]. Available: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-102.pdf>.
- [8] J. Li, D. Li, C. Xiong and S. Ho, "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," 15 February 2022. [Online]. Available: <https://arxiv.org/abs/2201.12086>.
- [9] X. Yao, X. Feng, J. Han, G. Cheng and L. Guo, "Automatic Weakly Supervised Object Detection From High Spatial Resolution Remote Sensing Images via Dynamic Curriculum Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 675-685, 2021.
- [10] J. Zhao, X. Xie, X. Xu and S. Sun, "Multi-view Learning Overview: Recent Progress and New Challenges," 2017. [Online]. Available: <https://shiliangsun.github.io/pubs/MVLOverviewIF17.pdf>.

- [11] C. Liu, X. Zeng, K. Wang, Q. Guo and M. Xu, "Multi-task Learning for Macromolecule Classification, Segmentation and Coarse Structural Recovery in Cryo-Tomography," 16 May 2018. [Online]. Available: <https://arxiv.org/pdf/1805.06332.pdf>.
- [12] J. Chen, H. Yu, C. Qian, D. Z. Chen and J. Wu, "A Receptor Skeleton for Capsule Neural Networks," in *Proceedings of the 38th International Conference on Machine Learning*, Virtual Only, 2021.
- [13] H. M. D. Trong, N. N. Trung, L. V. Ngo and T. H. Nguyen, "Selecting Optimal Context Sentences for Event-Event Relation Extraction," in *Association for the Advancement of Artificial*, 2022.
- [14] M. Vierlboeck, C. Lipizzi and R. Nilchiani, "Natural Language in Requirements Engineering for Structure Inference -- An Integrative Review," arxiv, 2022.
- [15] S. Smith, M. Patwary, B. Norick, P. LeGresley, S. Rajbhandari, J. Casper, Z. Liu, S. Prabhumoye, G. Zerveas, V. Korthikanti, E. Zhang, R. Child, R. Y. Aminabadi, J. Bernauer and X. Song, "Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model," arxiv, 2022.
- [16] J. Han, D. Zhang, G. Cheng, L. Gu and J. Ren, "Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-level Feature Learning," 2015. [Online]. Available: https://strathprints.strath.ac.uk/53413/1/Han_etal_IEEE_TGRS_2015_Object_detection_in_optical_remote_sensing_images_based_weakly.pdf.
- [17] G. Campanella, M. Hanna, L. Geneslaw, A. Mirafior, V. Werneck Krauss Silva, K. Busam, V. Reuter, D. Klimstra, T. Fuchs and E. Brogi, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, vol. 25, pp. 1301-1309, 2019.
- [18] J. Deng, N. Ding, Y. Jia, A. Frome, K. Murphy, S. Bengio, Y. Li, H. Neven and H. Adam, "Large-scale object classification using label relation graphs," in *European conference on computer vision*, 2014.
- [19] P. Zhou, D. Zhang, G. Cheng and J. Han, "Negative Bootstrapping for Weakly Supervised Target Detection in Remote Sensing Images," in *2015 IEEE International Conference on Multimedia Big Data*, 2015.
- [20] D. Lin, "An Information-Theoretic Definition of Similarity," *Proceedings of the 15th International Conference on Machine Learning*, pp. 296-304, 1998.

- [21] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric, Y. Bulatov and B. McCord, "xView: Objects in Context in Overhead Imagery," *arXiv preprint*, 2018.
- [22] J. Cho, A. Zala and M. Bansal, "DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generative Transformers," *arXiv preprint*, vol. 2202.04053, 2022.