

Appendix

Appendix I: Pre-Interview Questionnaire: FAIRification and Data Landscape Interview (for Repository Teams and Owners)

Q1 Full name	
Q2 Your project or repository affiliation	
Q3 Your organizational affiliation	
Q4 Your email address	
Q5 Are you a...	I. Repository Owner II. *Repository Services & Development team member
<i>N.b.:</i> Questions marked with an asterisk (*) are only to be responded to by repository services & development team members.	
Q6 What is your most common role related to your repository work?	I. Repository manager II. Data curator/steward III. Research manager IV. Software Developer V. Web Developer VI. Data wrangler VII. Metadata Specialist VIII. Other (please specify): _____
Q7 Please rate your familiarity with the FAIR Principles for purposes of managing and sharing data.	I. Familiar and put them into practice II. Familiar but do not currently put them into practice III. Heard of them but not familiar with what they mean IV. Never heard of the FAIR principles
Q8* What percentage of your repository's data has a globally unique and persistent identifier?	
Q9* What is the smallest entity of your repository's data that has a globally unique and persistent identifier?	I. Data set/collection II. Data file III. Data record

	IV. Other (please specify): _____
Q10* Please name and describe how each of the identifier types used in your repository's data is created, e.g., IDs for institutions, funding agency, research products, etc.	
Q11 Does your repository use ORCIDs or other digital personal identifiers for researchers?	
Q12* What percentage of your repository's metadata has a globally unique and persistent identifier?	
Q13* What is the smallest entity of your repository's metadata that receives a globally unique and persistent identifier? Please provide an example for each entity, if possible and as applicable.	I. Data set/collection description: _____ II. Data file description: _____ III. Data record description: _____ IV. Other (please specify): _____
Q14 Do your repository's globally unique and persistent identifiers explicitly identify and record versions for data and metadata entities?	I. Yes II. No III. Not sure IV. Other – please specify: _____
Q15* How many metadata fields does your repository require to describe each data entry? (Use your best estimate, or -1 if unknown.)	
Q16* If available, please provide or point to a list of the required metadata fields.	
Q17* How many metadata fields does your repository allow to describe each data entry? (Use your best estimate, or -1 if unknown.)	
Q18* If available please provide or point to a list of the allowed metadata fields.	
Q19 Do you believe the metadata of your repository sufficiently describes the data to meet your end users' discovery needs, i.e., finding and choosing appropriate data for research needs? (If not "Yes," please explain.)	I. Yes II. Sometimes; _____ III. No; _____

Q20 Does the metadata describing your repository's data include the identifier of the data being described? (If not "Yes," please explain.)	I. Yes II. Sometimes; III. No;
Q21 Does your repository track publications that have been produced based on your data or other evidence of data reuse? If Yes, what is the process or method that you use to identify the relevant publications to track or the types of data reuse?	I. Yes; II. No III. Don't know
Q22* When using a data set identifier, what standardized retrieval protocol(s) are supported by your repository?	I. None II. Direct download via User Interface (UI) III. Direct download via API command IV. SFTP V. Other (please describe): VI. Don't know
Q23* Using a metadata record identifier, what standardized retrieval protocol(s) are supported by your repository?	I. None II. Direct download via UI III. Direct download via API command IV. SFTP V. Other (please describe): VI. Don't know
Q24* Have your repository users had any difficulty accessing any data sets? If yes, please provide an explanation for the difficulty, if known.	I. Yes; II. No III. Don't know
Q25 If necessary, how do you enforce authentication and authorization in your repository for data download? (Please describe or list any protocols or rules that you use.)	
Q26 Are metadata for your repository's data freely available for download or export? If not, please explain.	I. Yes II. Sometimes III. No;
Q27 Please describe your response to the following statement: My repository's preservation plan makes an explicit	I. Strongly disagree II. Somewhat disagree III. Neither agree nor disagree

statement and provides assurance that metadata remain available after the corresponding data are removed.	IV. Somewhat agree V. Strongly agree
Q28* What standard representation format(s) do you support for sharing your data (e.g., CSV, JSON, XML or other)?	I. Comma-separated Values (CSV) II. JSON III. XML IV. Other: _____
Q29* What standard representation format(s) do you support for sharing your metadata (e.g., JSON, YAML, JSON-LD, RDF/XML)?	

Q30* For the following entities (A through F), describe how your repository represents controlled terms. Select all that apply.

	Free text string (1)	Validated text string (2)	Local code or ID (3)	CURIE (4)	Semantic IRI (5)	Mixed (6)	Other (7)
(A) Data file variable names (e.g., in top row of CSV file) (1)							
(B) Data file values from a short list of options (e.g., checkbox or radio options) (2)							
(C) Data file values from a large managed vocabulary (e.g., diseases, symptoms, or procedures) (3)							
(D) Metadata field or attribute names (e.g., in data element definition) (4)							
(E) Metadata field values from a short list of options (e.g., checkbox or radio options) (5)							
(F) Metadata field values from a large managed vocabulary (e.g., diseases, symptoms, or procedures) (6)							

Q31* Does your repository specify how to reference other relevant data sets or resources in data or metadata? Please describe or point to the policy.	I. Yes; _____ II. No III. Not sure
Q32* Does your repository specify a particular set of relations used to describe the related data sets or resources? If Yes, please describe the relations.	I. Yes; _____ II. Somewhat III. No
Q33* Does your repository have a list of required and allowed data set topic descriptions that you could provide to us? If yes, please provide or tell us how to acquire it.	I. Yes; _____ II. No III. Don't know
Q34 Do you believe the metadata of your repository sufficiently describes the repository's data to meet its end users' re-use needs? (If not "Yes," please explain your answer.)	I. Yes II. Sometimes; _____ III. No; _____
Q35 What conditions for access apply to the data and metadata in your repository?	
Q36 Does your repository have a clear and clearly visible data usage license? (Typically, this would mean the license is directly accessible from a link on the front page, it is no longer than a page, and it does not require sophisticated judgments to determine if data can be used.)	I. Yes II. No III. Don't know
Q37 Does your repository have a clear and clearly visible metadata usage license? (Typically, this would mean the license is directly accessible from a link on the front page, it is not longer than a page, and it does not require sophisticated judgments to determine if data can be used.)	I. Yes II. No III. Don't know
Q38 Does your repository support describing where and by what processes the data were created, including processing applied along the way? If so, please describe the support offered by the repository and how the information is made available to users.	I. Yes; _____ II. No III. Don't know
Q39 Please list any community standard supported by your repository for describing provenance information.	I. PROV II. CWLProv III. None IV. Don't know V. Other (please explain): _____

Q40 Does your repository's data follow domain-specific (documented) community standards? If so, please list; if not, please describe relevant details.	I. Yes; _____ II. No; _____ III. Not sure
Q41 Does your repository's metadata follow domain-specific (documented) community standards? If so, please list; if not, please describe relevant details.	I. Yes; _____ II. No; _____ III. Not sure
Q42 Are interested in further discussion and collaboration with the project team?	I. Yes II. No III. Maybe
Q43 Please acknowledge that you have completed the survey to the best of your knowledge.	I. Yes II. No

Appendix 2: Generic Work Plan Template

1. Introduction

1.1. Document Status

Document Version: 2.1

Document Status: [Preliminary, Pending Internal Review, Available for External Review, or Ongoing Maintenance]

Last Updated: [YYYY.MM.DD]

1.2. Work Plan Summary

1. **Target Repository:** [Repo Name]
2. **GO FAIR US Coordinator(s):**
3. **Target Repository Program Officer:** [Program officer(s), primary first]
4. **Target Repository Work Plan Representative:** [Program officer or similar with responsibility for directing work tasks]
5. **Project Technical Lead:** [Target Repository or grantee]
6. **Division Name, if applicable:**

The Work Plan tasks outlined in this document are intended for discussion with the repository team. In the discussion process, the repository team will select the most appropriate tasks to undertake at this stage of the project, and the work plan will be updated to reflect the chosen tasks and points of contact for each task. To offer maximum flexibility and a broader general picture, we describe more tasks, in more sections, than will be targeted in the final work plan.

Details about each task, including technical approaches and responsible persons (from the repository and/or from GO FAIR US) will be expanded (in Table 2 and auxiliary documents as needed) during the discussion process. The final Work Plan will include those details.

The Work Plan describes many tasks ambitiously, but less aggressive approaches may be better. We may reference other approaches in the Work Plan Notes, in the Appendix, or during discussions.

1.3. Work Categories

For reference, here we outline the five major categories of work. The Work Plan section details specific tasks.

The major categories of work are:

- **Persistent Identifier (PID) Implementation Strategies**
These strategies focus on having global, persistent, resolvable, and citable identifiers for data and metadata artifacts, and their component entities.
- **Metadata (MD) Improvement Strategies**
These strategies focus on improving descriptions for the data, supporting its discovery, accession, and re-use.
- **Community and Assessment Strategies**
These strategies focus on improving scientific collaboration and sharing within and across domains.
- **Architectural Strategies**
These strategies focus on adopting implementation approaches that make the repositories more broadly and easily interoperable.

- **Additional Specific Strategies**
These strategies focus on repository-specific ways to simplify the implementation of effective and consistent data services.

1.4 Table Key

- **Column Definitions:** The meaning for each column in a specific row is described below.
 - **Item #:** Each category gets its own category number ('1', '2'), while each strategy gets its own sub-number ('1.1', '1.2'). If there are several exclusive strategies to address the same strategic goal, each is appended with a letter ('1.1A', '1.1B').
 - **Priority:** Described tasks (next column) are prioritized as High, Medium, or Low according to their value/dependencies.
 - **FAIRification Task Description:** The strategy name, followed by specific tasks to achieve that strategy (each task is identified with a parenthetical letter ('(A)').
 - **Expected Outcome:** The Expected Outcome summarizes the goal of the described work, usually as a functional result.
 - **FAIR Principle:** The FAIR Principle(s) that apply to this strategy. FAIR Principle identifier codes and their descriptions may be viewed in [Appendix A](#).
 - **Metrics:** Proposed metrics for evaluating the usefulness of the strategy. (In some cases metrics may also assess the level of completion of the strategy.)
 - **Resources Required, Dependencies, and Notes:** Additional information about the strategy and corresponding tasks. The Resources Required are summarized as 'Effort'; see Effort and Resource Indicators bullet below for more information.
- **Ordering, priorities, and dependencies**
 - The strategic categories—single-digit sections (1, 2, 3, etc.)—are ordered by notional priority (in part due to dependencies), but community priorities may differ.
 - Within each category, the strategies—numbered items (1.1, 1.2, etc.)—are in our suggested priority order, and typically depend on previous strategies.
 - Within each strategy, the tasks—lettered items ((A), (B), etc)—are chronological, and typically depend on previous tasks.
- **Effort and Resource Indicators**
 - Pink text background on the task letters like '(A)' indicates relatively large effort (and associated costs), with lower priority.
 - Effort indicators are conceived as calendar time, yet also roughly correspond to work time. However, the repository goals, architectures, priorities, and chosen approaches may have big impacts on scoping any particular numbered item.
 -  = modest effort, 1 week or less for most tasks.
 -  = moderate effort, a few weeks to a month for most tasks.
 -  = major effort, at least some months will be required overall.
 -  = use of highlight color indicates that the highlighted effort is for the highlighted task bullets

2. Work Plan

2.1. Detailed Strategies

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1	Persistent Identifier (PID) Implementation Strategies These strategies focus on identifying the data and metadata artifacts, and their component entities, using identifiers that are globally unique, persistent, resolvable in a browser for humans or computers, and citable in documentation and publications. All of these features increase the value of the underlying data and metadata to the scientific researchers, and to anyone else who wants to discover or work with the resulting resources.					
1.1	High: A, B, C, D Medium: E	PID Strategy: Annotate Data Sets with Appropriate Terms A. Determine appropriate vocabulary(ies) to use B. Apply to existing studies and review with submitters C. Recommend use of the corresponding metadata field and vocabulary to future submitters. D. Display corresponding tags for each dataset, and support in sorting algorithm. E. Implement searching for datasets based on corresponding tags	Dataset search, sort, and/or browse by disease	R1, I2	1. # of time tag is used (e.g., users sort by or search by tag, or click on the tag to learn more) 2. % of submissions and datasets identified with one or more of the added tags	Effort: ⏳ ⏳ Dependencies: • Submitter input for existing datasets Notes: • Searching service could be external, in addition to or instead of internal service • GO FAIR US team can help choose appropriate vocabularies; this task would be called out in Final Work Plan

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1.2	High: A, B, C	PID Strategy: Implement ORCIDs as People Identifiers / RORs as Organization Identifiers A. Use and show ORCIDs of data submitters / RORs of data originators B. Present data-related ORCID / ROR data (if any) C. Recommend use of metadata fields (based on ORCIDs/RORs) to submitters	Establishment of the core practice to simplify later analysis and improve display of metadata (as the common ORCIDs and RORs can be used to connect researchers and organizations across data sets throughout the ecosystem).	R1, R1.2	# of attributes in system that support ORCIDs/RORs # of unique ORCIDs/RORs referenced in datasets # of submitters defining 'people metadata' with ORCIDs/RORs	Effort: ⌚ to ⌚⌚ Dependencies: - Researchers must have ORCIDs and be happy to provide them - Maximization of ORCID and ROR adoption and value requires completion of some metadata tasks Notes: - ORCIDs effectively 'market' researchers by connecting their diverse products and activities - RORs effectively 'market' organizations by connecting their diverse products and activities - It may be necessary to allow for alternatives or opting out

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1.3 A	High: A, B, C, D Medium: E, F Low: G	PID Strategy: Implement Study and Data IDs A. Evaluate cost/benefit of updating any existing study and data identifier schemes to adopt best practices for persistent identifiers (e.g., minimize semantic content, standardize identifier creation, make identifiers resolvable). Agree on final scheme. B. Allocate unique identifier for each study, associated dataset, and contained data file(s) C. Resolve study, dataset, and data file IDs to views for each entity type D. Make identifiers visible to external systems (Vivli, Google Data Search, ...) E. Allow download of entities via API using their identifier. F. Extend to all files within study (track part-of relations) G. Extend to versions of datasets (track relations) H. [Alternate or supplement to E,F,G] Support download of all metadata via JSON-LD or other documented computable format.	Persistent identifiers for content that promote the repository and bring more visitors to it.	F1, F3, A1, R1	# of data entity types that have unique identifiers # of visits/resolutions of each identifier type # of external repositories that cite or reference repo's identifier # of citations in papers that use repo's identifiers	Effort: ⌚⌚⌚ Dependencies: - Must decide if data file identifiers are feasible. Notes: - Task 1.3A targets file-based repositories, 1.3B targets record-based repositories. Choose one or both as appropriate. - Datasets are collections of one or more data files; unique IRIs are key for both entity types. - DOIs are available through doi.datacite.org

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1.3 B	High: A, B, C, D Medium: E, F Low: G	PID Strategy: Implement IDs for Targeted Entity Type(s) A. Identify any existing identifier scheme for the target entity(ies). Evaluate cost/benefit of updating existing identifier schemes to adopt best practices for persistent identifiers (e.g., minimize semantic content, standardize identifier creation, make identifiers resolvable). Agree on final scheme. (existing example if any: ...) B. Allocate unique identifiers for entities of each targeted entity type C. Resolve identifiers to views for each targeted entity type. (Forward any old identifiers to their replacement IRIs.) D. Make identifiers visible to external systems (Vivli, Google Data Search, ...) E. Allow download of targeted entities via API using their identifier F. Extend to all subsidiary targeted entities (track part-of relations) G. Extend to versions of targeted entities (track relations and timestamps) H. [Alternate or supplement to E,F,G] Support download of all metadata via JSON-LD or other documented computable format.	Persistent identifiers for content that promote the repository and bring more visitors to it.	F1, F3, A1, R1	# of targeted entity types that have unique identifiers # of visits/resolutions of each identifier by targeted entity type # of external repositories that cite or reference repo's identifiers # of citations in papers that use repo's identifiers	Effort: ⌚⌚ to ⌚⌚⌚ Dependencies: - Ability to decide which entity types should be targeted - Ability to define an identifier for the targeted entity type in a rigorous, reproducible, intuitive way Notes: - ADVANCED - Entity Types are classes of digital objects that are primary representations in record-based repositories. Detailed documentation regarding Enhanced Interoperability for Entity-Based Repositories will be provided. - Task 1.3A targets file-based repositories, 1.3B targets record-based repositories. Choose one or both as appropriate.

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1.4	High: A, B Medium: C, D, E	PID Strategy: Dataset Citation A. Present to submitters the format and use of citation strings for datasets and data files (using corresponding identifiers) in addition to publications. B. Publish the citation strings on the site for each dataset. C. Identify data sets in the best way possible, and report results to data providers and repo managers D. Add citation information to presented information in the repository. E. Evaluate the prospects of adding dataset identifier/citation fields to relevant data portals, e.g. clinicaltrials.gov.	Increased familiarity with, adoption of, and respect for data citations; increased visibility into dataset reuse.	F1, F2, F4	# of citations of repo's citation identifiers # of other references to repo's dataset citations - Altmetrics value for repo's datasets	Effort: ⌚ ⌚ Dependencies: Notes: - Data citation and identification strategies can work hand-in-hand. - GO FAIR US can suggest best options for citation strings for the repo. - GO FAIR US could offer tracking strategies to detect dataset citations of all of the repo's datasets. This is a large scope item but would have benefits across many repos.
1.5	Medium: A, B	PID Strategy: Auxiliary IDs A. Recommend to submitters the use of shared external identifiers for specifying Projects (tbd), Locations (tbd), and other common identifier types in metadata and data B. Present that information in metadata for the related dataset(s) in user interfaces and via APIs.	Increased access to contextual information about data; increased analytic understanding of the repo's contents; more shared interest in maintaining shared external identifier sources.	F2, F4, R1.2	# of unique Projects, Locations, and other shared entities represented by repo datasets % of submissions with at least one auxiliary ID (by entity type)	Effort: ⌚ Dependencies: - Need agreement on best Project and Location identifiers to use in across different repositories Notes: - Multiple identifier sources (e.g., Location identifier sources) may be appropriate - Recommend a central tracking system across repositories

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
1.6	High: A Medium: B, C, D	PID Strategy: Provide Shared ID Service A. Identify any collection of identifiers for external resources that is, or could be, used as a shared resource service by other repositories B. Establish that necessary conditions exist for sharing the resource: - GUPRIs exist and are resolved for each identifier in the collection - Entity type being shared is well covered - API mechanism exists to submit, curate, ingest, publish, and search new instances - Public documentation exists on how to access all services C. Agree with NIH or other funders on policies, practices, and funding arrangements that will be followed to continue supporting the shared resource and assure its persistence. D. Share and announce the service(s)	A single source of truth across multiple organizations and repositories; more complete, reliable, and maintainable source of knowledge; easier adoption of standard identifiers by external parties; allows building of more complex systems by relying on external services.	F2, F4, R1.2	- Metrics of information contained in the service - Metrics of requests served by the service. - Metrics of organizational adoptions of the service as a primary provider of information.	Effort: ⌚⌚⌚ Dependencies: - Some scale must be achieved before it becomes obvious that the single source of truth is more useful and efficient than multiple sources. - Commitment from senior management to support service in an ongoing way is needed before others will have the confidence to adopt it. Notes: - ADVANCED - Information in the repository must be shared with and open to the relevant community so as to assure transition of information if the system goes down.
2	Metadata (MD) Improvement Strategies The metadata strategies improve the descriptions of the data, including their findability, availability and usability. This makes it easier to discover the data, get access to it (if that is appropriate), and re-use it for validation of scientific research or pursuing new scientific discoveries. The strategies also simplify the effort required for scientific communities to work together on shared problems.					

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
2.1	High: A, B, C, D Medium: E, F	MD Strategy: Standardize Metadata Representation A. Determine metadata schema* for dataset and data file to include at attributes important for domain-specific or general search portals B. Provide effective guidance to data providers on following metadata schema (including controlled vocabularies and other identifiers) C. Complete minimum required metadata for existing projects D. Expose repository's metadata for machines (download one or all entries by metadata ID using a common API architecture) E. Expose repository's metadata for people (browse and view by metadata ID) F. Automatically validate metadata and enforce appropriate standards	Each task further prepares the repository for a major upgrade in FAIR capabilities. Completing at least through task (D) enables external repositories and services to offer advanced discovery and re-use of data and metadata, increasing interest and visits.	F2, F4, A2, I1, I2, I3, R1, R1.2	- Datasets # of required attributes supported by all submitted datasets # of optional attributes supported (on average) by all submitted datasets - Repo APIs # of required attributes supported by repo APIs # of optional attributes supported (on average) by repo APIs # of repo-specified metadata attributes (required or optional) that come from, or are mapped to, attributes in community specification	Effort: ⌚⌚⌚ Dependencies: - The 'minimum required metadata' attributes should be determined in (A) and described in (B) in order to be filled out in (C). Notes: * A metadata schema is a document listing what metadata the repository will support for each entity (e.g., dataset, data file). Metadata schemas should specify 'required' and 'optional' metadata attributes. - A tool like CEDAR can make the first three tasks relatively simple - A phased approach could adopt basic schema with limited semantics, but with less value - A significantly less desirable alternative to (D) might be to support download of all metadata via JSON-LD or other documented computable format. See Appendix B: Architectural Considerations below for repo-specific observations.

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
2.2	High: A, B, C Medium: D, E, F	MD Strategy: Standardize Data Dictionaries A. Adopt a computable Data Dictionary format with semantic capabilities B. Encourage use of new Data Dictionary format (include guidance) C. Expose Data Dictionaries via API D. Provide guidance on selecting vocabularies and terms E. Provide tool for building Data Dictionaries F. Provide evaluation/validation of submitted Data Dictionaries	Uniform machine-parsable data dictionary formats used to describe all data files (provides major discovery and re-use value).	F2, I2, R1, R1.2	1. # and % of submissions with computable semantic Data Dictionaries 2. # and % of submissions with repo-validated Data Dictionaries	Effort: ⌚⌚⌚ Dependencies: - Needs a Data Dictionary format that is acceptable to most adopters Notes: - A 'computable semantic DD format' exists and can be adopted to desired level of rigor. - Each attribute in a Data Dictionary is analogous to a Common Data Element (CDE), which can be shared to build other Data Dictionaries

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
2.3	Medium: A, B, C	MD Strategy: Extend Metadata Representation A. Recommend particular properties when describing contextually related artifacts and encourage adoption of those terms and metadata attributes. B. Recommend particular terms to submitters for describing data provenance, including addressing data integrity concerns. C. Appropriately provide added metadata with corresponding data, including via API	More mature and complete representation of the context of each data set, for better understanding of the data set users.	F2, R1.2	1. # of property types used to define contextually related artifacts 2. # of contextually related artifacts referenced (by property type)	Effort: ⌚⌚⌚ Dependencies: • Previous standardization strategies should be in place. • Tooling to simplify metadata entry is recommended. • Data provenance and data integrity are advanced metadata topics and often depend on specifics of the scientific activity and domain. Notes: • ADVANCED • For (C) See Appendix B: Architectural Considerations below

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
2.4	High: A, B Medium: C Low: D	MD Strategy: Maximize Metadata Adoption A. Offer tooling to fill out metadata B. Provide tracking services that detect the metadata visibility and selection of all repo datasets and report results to data providers and repo managers C. Analyze the degree to which different metadata attributes are provided in submissions, are provided to data analysts by the repo, and contribute to discovery and use of the datasets. D. Extend IDs to versions of metadata	Increased use of tooling that enables more and better metadata to be provided with less effort.	A1, A2, I1, I2, I3, R1	- Percent of metadata completed over time - For tooling-assisted metadata, - Number and % of attributes completed - Completeness of completed attributes - Correctness of completed attributes - Time required to complete metadata	Effort: ⌚⌚⌚ Dependencies: - Tracking services to detect when repo datasets are selected usually have to be embedded in the repo's software framework. Notes: - ADVANCED (the tooling and some tracking is straightforward; many metrics are subtle) - Tooling described in (A) exists and can be easily integrated with other systems - Typically versioning is challenging, especially if it isn't built in from the start

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
2.5	High: A, D Medium: B, C Low: E	MD Strategy: Assess Domain-Centric Strategy A. Using a controlled terminology, characterize domains covered by existing submissions. (Include domains in dataset metadata at convenient time.) B. For each identified domain, determine what specifications (metadata, vocabularies, data formats) exist for that domain. Consider any other target repositories dealing with data in similar domains. C. Hold organization-wide FIPs* workshop to assess domain-specific activities across target repositories. D. Define recommendations to encourage future submitters in those domains to follow domain strategies. E. Recommend path to support adoption of domain guidance for any likely domain.	Identification and adoption of practical domain-specific approaches that each repository can support, and each data contributor can adopt. Knowledge of domain-level adoption of metadata and vocabulary schema and tools across all the participating repositories. Increased community awareness and engagement.	F2, R1.2, R1.3	1. # of domains covered by existing submissions 2. # of specifications existing for identified domains 3. # of specifications adopted (as in declared favorable) by this repo for existing domains 4. # of specifications for the relevant domain used by submission (total, and as a % of ideal)	Effort: ⌚⌚⌚ Dependencies: • Agreement on controlled terminology(ies) to use to characterize domains Notes: • * A FIP is a FAIR Implementation Profile, which describes some of the implementations a repository has chosen to meet their data and metadata management requirements. A FIPs workshop is an opportunity for repositories to describe, and see others describe, their (potentially shared) implementation choices. • Domain is most often a scientific domain, but can also be oriented to the purpose of the repository (e.g., ‘all data not stored in other repositories’)
3	Community and Assessment Strategies These strategies explicitly address the ability of research communities to work together, to understand and define their own discipline’s needs and any transdisciplinary needs, and to ensure their community’s objectives are well met. When these strategies are implemented, scientific cooperation within specific communities or domains and across domains can be enhanced while making the sharing process more efficient and effective.					

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
3.1	High: A, B, C Medium: D, E	Assess Strategies for Expanding Domain Researcher Data Sharing and Reuse A. Assess existing Data Sharing and Data Use Agreements to determine whether recommendations can be included that will allow the creation of citations for datasets. B. Solicit feedback from research data submitters or reusers on their experiences with the repository for purposes of assessing and improving submission and download workflows. C. Assess whether there are other places on the repository's public web pages where examples of recommended formats for citing published outputs from use and reuse of the datasets can be displayed prominently. D. Consider adding information about the repository to repository registries such as fairsharing, SciCrunch, etc. to expand awareness of the repository and its data. E. Assess whether more data usage metrics can be added to the download functions or to associated search methods.	Increased awareness of the repository's existence, with improvement in both download and upload (sharing) of data sets.	F4, I1	Measure impacts on data usage statistics of expanded sharing of information about the data collections within the repository (e.g., increased number of citations, altmetrics) For a given modification or set of modifications: 1. % change in citations 2. % change in altmetrics 3. % change in visits to relevant page type 4. % change in user access requests	Effort: ⌚ Dependencies: • GO FAIR US staff time Notes: • These strategies more broadly evaluate project-specific opportunities for improving data sharing and re-use, with GO FAIR US team members providing significant analytical support. • Each strategy pursues targeted improvements, at relatively low cost.

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
3.2	High: A Medium: B, C, D, E	Develop Key Use Cases A. Identify key interoperability use cases for your repository (things you want to be able to do, or do better). B. If desired, prioritize the use cases to provide input to the Work Plan. C. Assess whether implementing any of the use cases convincingly demonstrates the scientific value of the work (particularly by using quantifiable metrics in a controlled evaluation). If so, define necessary steps to achieve that demonstration. D. Implement necessary steps to achieve demonstration. E. Publish results in appropriate literature.	Improved understanding of repository goals, better Work Plan prioritization, and identification of rigorous criteria to evaluate the value of making your repo more FAIR.	All	1. # of use cases identified, prioritized, and addressed. 2. # of use cases identified as demonstrator use cases. 3. # of evaluated implementations published.	Effort: ⌚⌚⌚ Dependencies: Notes: - GO FAIR US team may help with heavy lifting in steps (C), (D), and (E) - Doing steps (A) and (B) is a valuable contribution to planning and prioritizing of work

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
3.3	High: A, B, C Medium: D	Community Training & Engagement A. Host Metadata for Machine (M4M) workshop* to bring FAIR-motivated participants to greater and shared awareness and interest. B. Embed mini-FIPs workshop** to identify currently adopted domain-specific practices for that community. C. Incorporate training on any of the other in-progress or completed aspects of this Work Plan. D. Support distinct groups in selection and/or creation of initial metadata profiles and needed controlled vocabularies.	Awareness and skills across the repository's user and management communities that increase support for and adoption of FAIR practices.	All, esp. R1.3	1. # of participants completing M4M workshop 2. # of participants completing FIPs or mini-FIPS workshop 3. # of unique participants and participant-hours attending FAIR training activities 4. # of projects receiving consultative support in particular product development (by product type) 5. # of consultation hours provided in particular product development (by product type) 6. # of products generated with consultative support from GO FAIR US	Effort: ⌚ ⌚ Dependencies: - Each task calls for interest from the targeted participants (both managers and attendees). Notes: * An M4M Workshop trains its participants to value, create, and model rich, principled metadata that describes data in their research domains. ** See Strategy 2.5 for FIPs workshop definitions - Though many hours are expended to host and attend a workshop, the benefit per hour for participants and providers is high. - If participants are skilled software/data experts, shorter but more targeted workshops and training will be optimal.
4	Architectural Strategies Architectural strategies concern the actual implementation approaches that are already in place, and that could be adopted by repositories. These are not one-size-fits-all prescriptions, but examples of approaches that could be effective in particular situations. Choosing good architectural approaches for a given repository can make that repository's data and metadata more easily discovered and used, and those repository products more often cited. Note that Custom observations may be offered in Appendix B: Architectural Considerations .					

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
4.1 A	Medium: A, B, C	Structure Metadata on Web Pages A. Establish sitemap.xml B. Generate JSON-LD metadata for defined metadata attributes following schema.org specifications C. Publish JSON-LD into script tags in the Study, Dataset, or other entity's landing page head tags	All defined metadata can be indexed into Google Dataset Search and other data portals as desired.	F4, A1, A2	1. Metadata can be accessed publicly and by Google Dataset Search 2. #/trend of page visits from sources Google Dataset search and plain Google search	Effort: ⌚⌚ Dependencies: - Requires ongoing access to web pages for each study, dataset, or entity - Must identify which entity(ies) to provide structured metadata for Notes:
4.1 B	Medium: A, B, C	Improve Existing Endpoint A. Improve the current (REST, RPC, GraphQL, other) API endpoint to include metadata described in earlier sections B. Make API publicly accessible C. Publicly document the endpoint	Allows metadata and identifier improvements identified in other tasks, and offers flexible public access to metadata.	F4, A1, A2	1. Queries can be made by anyone (or at least, anyone with an account) 2. # Projects using/analyzing metadata via the API	Effort: ⌚⌚ Dependencies: - Developer with appropriate experience Notes: - Existing API:

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
4.1C	Low: A, B, C, D	Provide New Metadata API A. Establish sitemap.xml B. Generate JSON-LD metadata for defined metadata attributes following schema.org specifications C. Publish JSON-LD in public endpoint (API) following common web standards (e.g., using HTTPS REST or RPC protocols). D. Publicly document the public endpoint	Provides REST-based querying interface to access system metadata.	F4, A1, A2	1. HTTPS queries can be made by anyone (or anyone with an account, at least) 2. # Projects using/analyzing metadata via HTTPS API	Effort: ⌚⌚⌚ Dependencies: - Allowance of the metadata to be publicly accessible. Notes: - Priority higher if existing endpoint is non-standard or poorly integrated - Is similar to 4.1, but avoids bespoke patterns - Could be combined with 4.1 - Example REST-based query to get metadata: https://data.ex.org/metadata/<metadata_ID>
4.2	High: A, B, C Medium: D	Architectural Integration Consult A. Hold a small series of design presentation meetings to understand existing repo system design and goals B. Assess and discuss best approaches to maximize integration with most important external system while minimizing upkeep C. Present and review proposed approach(es) and tradeoffs to enable repository's choice of implementation D. Implement selected approach	Determines best approach and practices for integrating existing or chosen architecture with external systems.	F4, A1, A2	1. Maintainable interoperable architectural approach(es) defined 2. # external systems supported by design and implementation	Effort: ⌚⌚⌚ Dependencies: Notes: - External systems include Google Data Sets and search, and external community services

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
5	Additional Specific Strategies A number of other specific strategies can be addressed for certain repositories. The common approaches identified in this section can simplify the implementation of effective data services, and help the repositories in the community provide more consistency to their users.					
5.1	High: A, B, C Medium: D	Assess Data Usage License Strategy A. Determine whether an appropriate data license exists to cover the repository's needs B. If necessary, determine what collection of well-defined data licenses are sufficient to cover (individually) existing data submissions C. Strategize best plan to publish data license for all datasets D. Present appropriate data license in the metadata for each dataset	A defined data license that offers reassurance that the data is available and likely reusable.	R1.1	1. % of datasets with data usage license 2. % of datasets with published data usage license 3. % of downloads with data usage license embedded with download	Effort: ⌚ to ⌚⌚ Dependencies: - A discussion may be necessary about preferred licensing options (n.b., generic government data license) Notes: - An ideal license is generic, but could refer to specific licenses that stand as exceptions - A goal is to maximize sharing, while maintaining flexibility needed by submitters

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
5.2	High: A, B Medium: C	Address Persistence Strategy A. Define repo's preservation approach (at minimum for metadata) B. Develop and publish Preservation Statement based on the repo's definition of its preservation approach C. Get overseeing organization's commitment to Preservation Plan	Publicly vouched metadata preservation strategy that increases user confidence in value of data submission.	A2	1. Data a. Existence of Preservation Statement for data b. Existence of published Preservation Statement for data c. Existence of overseeing organization-endorsed published Preservation Statement for data 2. Metadata a. Existence of Preservation Statement for metadata b. Existence of published Preservation Statement for metadata c. Existence of overseeing organization-endorsed published Preservation Statement for metadata	Effort: ⌚ ⌚ Dependencies: - Program Officers (POs) must be willing to contribute to, publish, and commit to a preservation statement that can capture the repository's preservation approach - Overseeing organization must be able to provide necessary resources to support the preservation statement. Notes: - Core Trust Seal is the high end of preservation statements. - The commitment in (C) is not a requirement for publishing a statement, if the statement is not too ambitious (for example, rehosting metadata in an external repository might not require additional commitments) - Even a minimal statement can be very useful.

Item #	Priority	FAIRification Task Description	Expected Outcome	FAIR Principle	Metrics	Dependencies, resources required, and notes
5.3	Medium: A, B, C	Address AI and LLM Integration A. Define community questions and concerns about intersection of existing data and developments in AI and Large Language Models (LLMs) B. Agree on training, reference, or discussion materials to be presented/hosted by GO FAIR US. C. Present the agreed materials and meetings for the repository to consolidate their knowledge and move forward.	Understanding (and potentially demonstration) of applicability of AI and LLM technologies in data repository	All	1. # of potential applications of AI and LLM to creating and using the repository's resources. 2. Assessed happiness of the participants with the information that was provided.	Effort: ⌚ ⌚ Dependencies: - GO FAIR US to provide the training and materials. Notes: - ADVANCED - Curriculum can be standard, or customized to address specific community concerns.

2.2. Task Management Table

This table will be filled out with the selected tasks, their owner and other task management comments.

Item #	Owner	Date of Completion	Notes

Appendix A

The 15 FAIR Principles are first described in the paper The FAIR Guiding Principles for scientific data management and stewardship, and are explained/justified in more detail at the GO FAIR FAIR Principles pages.

F1: (Meta)data are assigned globally unique and persistent identifiers

F2: Data are described with rich metadata

F3: Metadata clearly and explicitly include the identifier of the data they describe

F4: (Meta)data are registered or indexed in a searchable resource

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol

A1.1: The protocol is open, free and universally implementable

A1.2: The protocol allows for an authentication and authorisation procedure where necessary

A2: Metadata should be accessible even when the data is no longer available

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

R1: (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1: (Meta)data are released with a clear and accessible data usage license

R1.2: (Meta)data are associated with detailed provenance

R1.3: (Meta)data meet domain-relevant community standards

Appendix B: Architectural Considerations