# Step-by-Step Guide to Using OAI Data

*All the OAI CORE Knowledgebase services described in this guide are offered free of charge.*

*The knowledgebase offers other services described on the website for a fee-for-service (e.g., image assessments).*

*Please cite the NIH grant any time you use OAI CORE Knowledgebase resources or services: R24 AR085006.*

## Step 1. Develop an OAI-feasible research question

- Define the scientific question, target population, exposure(s), outcome(s), unit of analysis, and time horizon.
- Determine early the unit of analysis (e.g., participant- or knee-level).
- Confirm that OAI contains the required variables and follow-up structure.
    - OAI Data is distributed across three repositories:
        - NIMH Data Archive (NDA): clinical data, image assessments, images, and nested studies data
        - dbGaP: whole-genome genotyping data
        - Biospecimen repository
    - Review study design, subcohorts, visit schedules, imaging protocols, and availability of desired data.

⚠ **Common mistake:** Investigators define a question before confirming that the exposure and outcome were collected at compatible visits.

**How the OAI CORE Knowledgebase can help**

- Consulting: Help refine feasible OAI-based questions.
- Consulting: Identify variables, visits, subcohorts, and analytic limitations.
- Consulting: Identify key documents to review
- Consulting/Resource: Connect investigators with experienced collaborators.
- Resource: Offer webinars and archived presentations describing OAI resources.

## Step 2. Obtain access to OAI resources

- Create an NDA account and complete the Data Use Certification (DUC).
    - Go to https://nda.nih.gov/oai
    - Login via Login.Gov
    - Go to "My Account" (upper right) --> Click "Data Permissions"
    - Scroll down to "Request Access to a Permission Group" and find the "Osteoarthritis Initiative" and click "Request Access"
    - Click" Start Request" and follow any other prompts
- Apply separately for dbGaP genetic data or biospecimen access if needed.

⚠ **Common mistake:** Discovering late in a project that genetics or biospecimens require additional approvals.

**How the OAI CORE Knowledgebase can help**

- Consulting: Help investigators navigate repository access requirements.
- Consulting: Assist with the preparation of biospecimen request applications.

## Step 3. Review documentation

- Read protocols, data dictionaries, and manuals, including codebooks and release comments.
- Confirm variable definitions, coding conventions, and visit availability.
    - Check OAI Examination Measures
    - Check OAI Questionnaire Measures
- Verify relevant missing data codes (e.g., .A, .D, .M, .P)
- Review the image assessment project documentation.

⚠ **Common mistake:** Beginning analyses without understanding variable definitions or missing-data codes.

**How the OAI CORE Knowledgebase can help**

- Consulting: Guide investigators to the most relevant OAI documentation.
- Consulting: Support onboarding through tutorials and training resources.

## Step 4. Confirm data availability

- Map variables to visits before downloading data.
- Verify participant-level versus knee-level structure.
- Check whether variables exist only within nested substudies.

⚠ **Common mistake:** Assuming a variable exists longitudinally when it only exists at selected visits.

**How the OAI CORE Knowledgebase can help**

- Consulting: Assist with mapping variables to specific visits and flag data only available in nested substudies.
- Resource: The planned OAI Explorer will support variable-by-visit searches and basic descriptive analyses.

## Step 5. Download and organize the data

- Download the files in a zipped folder or create parsimonious datasets using NDA tools.
- Download images, as needed.
- Track dataset versions and file names.

⚠ **Common mistake:** Mixing multiple dataset versions in the same analysis pipeline.

**How the OAI CORE Knowledgebase can help**

- Consulting: Assist with dataset selection, organization, and reproducible workflows.
- Consulting: Create parsimonious image sets (e.g., knee radiographs at selected visits)

## Step 6. Understand the dataset structure before merging

- Determine whether datasets contain one row per participant, knee, or repeated observation.
- Identify key merge variable(s) such as ID, SIDE, VISIT, and READPRJ.
- Review imaging project structures carefully.

⚠ **Common mistake:** Incorrectly merging different dataset structures (e.g., participant-level with knee-level data).

**How the OAI CORE Knowledgebase can help**

- Consulting: Help investigators understand row structures and merge logic.

## Step 7. Merge datasets carefully

- Verify row counts and duplicates after every merge.
- Document all merge logic.
- Avoid combining incompatible imaging projects without harmonization.

⚠ **Common mistake:** Combining multiple image assessment projects without accounting for sampling design or repeated readings.

**How the OAI CORE Knowledgebase can help**

- Consulting: Troubleshoot merge strategies and harmonization decisions.
- Consulting: Create merged or parsimonious datasets for investigators

## Step 8. Handle missing data appropriately

- Differentiate structural missingness from true missingness.
- Review contingent variables and skip patterns carefully.
- Evaluate attrition and informative missingness.

⚠ **Common mistake:** Treating all missing values identically.

**How the OAI CORE Knowledgebase can help**

- Consulting: Provide guidance on OAI missing-data codes and coding strategies.

## Step 9. Build the analytic cohort carefully

- Clearly define inclusion and exclusion criteria.
- Specify OA definitions, visits, and subcohort handling.
- Document participant- versus knee-level eligibility.

⚠ **Common mistake:** Using subcohort assignment as a surrogate for disease definition.

**How the OAI CORE Knowledgebase can help**

- Consulting: Review cohort definitions and discuss generalizability.

## Step 10. Select appropriate analytic methods

- Account for repeated measures and correlated knees.
- Consider mixed-effects models, GEE, weighting, or imputation where appropriate.
- Account for nested substudy sampling frameworks.

⚠ **Common mistake:** Treating knees as independent observations.

**How the OAI CORE Knowledgebase can help**

- Consulting: Provide consultation on analytic strategies and interpretation.

## Step 11. Special considerations for AI and machine learning

- Partition datasets at the participant level to avoid data leakage.
- Clearly define prediction windows and temporal alignment.

- Document preprocessing, feature engineering, and validation strategies.

⚠ **Common mistake:** Allowing the same participant to appear in training and testing datasets.

**How the OAI CORE Knowledgebase can help**

- Help identify leakage risks and longitudinal design challenges.

## Step 12. Report transparently and reproducibly

- Report repositories, file names, dataset versions, visit definitions, and merge strategies.
- Describe missing-data handling, OA definitions, and imaging projects.
- Include required OAI acknowledgments and funding statements.

⚠ **Common mistake:** Omitting dataset versions or insufficiently describing cohort construction.

**How the OAI CORE Knowledgebase can help**

- Review reporting completeness and reproducibility.