

Protocol for the Spectral Digitization of Herbarium Specimens



Version 1.1

Table of contents

Document Scope, Versioning, and Terms of Use	3
Changes Since Last Version	4
Current Version: 1.1	4
Summary of changes since Version 1.01	4
Preface	5
Abstract	5
IHerbSpec and Protocol Development	5
Authors	6
Author Contributions	6
Acknowledgements	7
Suggested Citation	7
Navigate the protocol	7
Part 1 – Overview	9
1.1 Purpose and Structure of the IHerbSpec Protocol	9
1.2 Measurement and Metadata Overview	9
1.2.1 Filename Conventions	10
1.2.2 White Reference and Background Measurements	10
1.2.3 Strategy and Number of Tissue Measurements	11
1.2.4 Measurement Technique	12
1.2.5 Metadata Requirements and Flexibility	12
1.2.6 Scoring Metadata Pertaining to Tissue Condition and Contamination	12
1.2.7 Specimen and Tissue Annotation	13
Navigate the protocol	13
Part 2 – Measurement and Metadata Workflow	14
Overview	14
Step 1. Instrument Setup and Connection	14
Step 2. Prepare Bench and Specimens	14
Step 3. Start Session and Score Session Metadata	15
Step 4. Collect White and Black Session Measurements	15
Step 5. Tissue Measurement Sequence	15
Step 6. Quality Assessment, Quality Control	16
Step 7. Score Specimen and Tissue Metadata	16
Step 8. Specimen and Tissue Annotation	16
Step 9. Move to Next Specimen	16
Navigate the protocol	18
Part 3 – Filename Conventions and Formats	19
Overview	19

3.1 Recommended Filename Conventions	19
Table 3.1: Full and Simple Filename Conventions with Different Measurement Target Classes	19
3.2 Filename Components	20
Table 3.2: Filename Components and Corresponding Metadata Fields.	20
Navigate the protocol	21
Part 4 – Metadata and Databasing	22
Overview	22
4.1 Metadata Tables	22
Table 4.1: Session Metadata	22
Table 4.2: Specimen Metadata	24
Table 4.3: Tissue Metadata	25
4.2 Controlled Vocabularies	31
Table 4.4: Developmental Stage Class Codes	31
Table 4.5: Target Class Codes.	31
Table 4.6: Tissue Descriptor Codes	32
Table 4.7: Background and White Reference Class Codes	33
4.3 Guidelines for Data Archiving and Sharing	34
Navigate the protocol	35
Part 5 – Instrumentation and Materials Guidelines	36
5.1 Precautions	36
5.2 Spectra Data Files	36
5.3 Instrumentation and Materials Quality Control	37
5.4 Specimen and Tissue Annotation	39
Navigate the protocol	40
Part 6 – Selecting Tissues for Spectral Measurement	41
6.1 Sources of Variation in Herbarium Plant Tissues	41
General Guidelines for Tissue Selection	41
Decision tree for tissue selection	44
Navigate the protocol	44
References	46
Navigate the protocol	46
Appendix I – Considerations Regarding the Number of Measurements per Specimen	48
Considerations	48
Navigate the protocol	52
Appendix II – Representing Tissue Conditions in Metadata for Quality Control	53
Scoring metadata for tissue conditions	53
Table A1: Examples for Scoring of Tissue Conditions.	53
Navigate the protocol	56

Document Scope, Versioning, and Terms of Use

Version 1.1

This PDF is a compiled and citable version of the **Protocol for the Spectral Digitization of Herbarium Specimens** published by **IHerbSpec**, aka the “IHerbSpec Protocol”. It is a single document for offline use, printing, and long-term reference that represents the latest compiled and versioned release of the protocol.

This versioned release is archived on Zenodo doi.org/10.5281/zenodo.15849668.

The IHerbSpec Protocol is maintained as a living standard on the project website, iherb-spec.github.io, where the most current version is always active and may include minor updates or clarifications made since this archived release.

Reuse and adaptation of this material are permitted under the **CC BY 4.0 license**.

When redistributing or adapting content, please include the attribution: “**IHerbSpec Protocol (CC BY 4.0)**”

Thank you!

Changes Since Last Version

Current Version: 1.1

Summary of changes since Version 1.01

- Reorganized the protocol into a Quarto book structure to enable consistent rendering of a single, compiled PDF while preserving the web-based version of the protocol: – Beginning with this release, versioned snapshots of the IHerbSpec Protocol will be generated **directly from the website source repository** and archived on Zenodo. This change streamlines version control, reduces duplication, and ensures that archived releases correspond exactly to the content presented on the website at the time of release. – As part of this transition, the **legacy IHerbSpec-protocol GitHub repository will be deprecated**. It will remain available for reference and historical context, but future protocol updates, releases, and Zenodo archives will be managed exclusively through the main IHerbSpec website repository.
 - Introduced **full title** for the protocol as “Protocol for the Spectral Digitization of Herbarium Specimens”. The alternate title “IHerbSpec Protocol” is still used, but the full title is better for a citation that differentiates the author and title.
 - Updated **citation**: “IHerbSpec” as author along with full title above.
 - Added a Downloads section to the protocol, available in HTML, where users can easily access all tables. These tables are called from in `protocol/tables.qmd` and live in `protocol/tables/`.
 - Added this “**Changes Since Last Version**” section to document versioned updates transparently.
 - Added logo to homepage and icon to navbar.
 - Expanded **Part 5 - Instrumentation and Materials Guidelines** to include guidance on depositing spectral data and metadata in the **IHerbSpec Dataverse**.
 - **Section 5.3.2:** Added Fig. 5.1 and new text describing the spectral properties of background materials and their effects on measured leaf reflectance.
-

Preface

Abstract

Reflectance spectroscopy is a powerful, broadly integrative tool for capturing plant phenotypes. But variation in instrumentation and measurement procedures introduces a high risk of data incompatibility and signal contamination across datasets. These challenges are amplified in herbarium specimens because they are subject to complex variation from age, preservation techniques, and mounting practices.

To support global data harmonization and scientific reproducibility, this document presents a standardized, interoperable protocol for the spectral measurement of herbarium tissues. It provides data-driven justification and community-informed guidance for measurement procedures and metadata fields, with the goal of improving data quality and enabling confident data aggregation across institutions.

IHerbSpec and Protocol Development

The International Herbarium Spectral Digitization Working Group (IHerbSpec) was established in December 2024 as a global consortium focused on advancing the use of reflectance spectroscopy of herbarium specimens for ecological and evolutionary studies.

The group actively collaborates to identify common challenges, facilitate collaborations, and develop best practices.

This protocol is one of the primary products of IHerbSpec's initial collaborative phase. From December 2024 to June 2025, members convened virtually to share project-specific workflows and align measurement strategies. The protocol was developed through an open, collaborative process, with a majority of IHerbSpec members contributing to discussions and decisions and serving as authors. The effort culminated in an in-person workshop at the Harvard University Herbaria (May 1–3, 2025), where members reached consensus on key elements of the protocol, including:

1. The minimum and recommended number of measurements for leaf tissues (see [Strategy and number of tissue measurements](#)).
2. Tissue condition and contamination metadata (see [Scoring metadata pertaining to tissue condition and contamination](#)).
3. Tissue development stage metadata (see [Table 4.4](#)).

IHerbSpec will continue to refine and expand this protocol and welcomes inquiries from interested researchers or institutions wishing to ask questions, provide feedback, or explore opportunities for participation.

Authors

- **Dawson M. White** — Harvard University Herbaria and Department of Organismic and Evolutionary Biology, Harvard University, USA
- **Natalie I. Ahlstrand** — Natural History Museum of Denmark, University of Copenhagen, Denmark
- **Matthew W. Austin** — Missouri Botanical Garden, USA
- **Denis Bastianelli** — CIRAD / University of Montpellier, France
- **Samantha Bazan** — CIRAD / University of Montpellier, France
- **Khalil Boughalmi** — University of Montpellier, France
- **Warren Cardinal-McTeague** — University of British Columbia, Canada
- **Thomas L.P. Couvreur** — University of Montpellier, CIRAD, IRD, France
- **Flavia Durgante** — Karlsruhe Institute of Technology / National Institute of Amazonian Research (INPA), Brazil
- **Olwen M. Grace** — Royal Botanic Garden Edinburgh, UK
- **J. Antonio Guzmán Q.** — Harvard University, USA
- **Kimberly Hansen** — Field Museum of Natural History, USA
- **Michael Hopkins** — National Institute of Amazonian Research (INPA), Brazil
- **Rykkar Jackson** — University of British Columbia, Canada
- **Jonathan A. Kennedy** — Harvard University, USA
- **Shan Kothari** — University of Alberta, Canada
- **Aaron K. Lee** — University of Minnesota, USA
- **Étienne Léveillé-Bourret** — Université de Montréal, Canada
- **José Eduardo Meireles** — University of Maine, USA
- **Barbara M. Neto-Bradley** — University of Cambridge, UK
- **Cornelius Onyedikachi Nichodemus** — University of Maine, USA
- **Natalia L. Quinteros Casaverde** — NASA Goddard Space Flight Center, USA
- **Carlos Rodrigues-Vaz** — Muséum National d'Histoire Naturelle, France
- **Michaela Schmull** — Harvard University Herbaria, USA
- **Pamela S. Soltis** — University of Florida, USA
- **Daniel Spalink** — Texas A&M University, USA
- **Caroline C. Vasconcelos** — National Institute of Amazonian Research (INPA), Brazil
- **Jéssica Lira Viana** — Aarhus University, Denmark
- **Hanna Tuomisto** — University of Turku, Finland
- **Tom Wells** — Royal Botanic Gardens, Kew, UK
- **Jeannine Cavender-Bares** — Harvard University, USA

Author Contributions

All authors contributed intellectually to the development of the protocol, including reviewing, editing, and approving the final version. DMW organized and led protocol development, writing, and website management.

- Intellectual leadership: Led by JCB with contributions from MWA, FD, JAK, JEM, JAGQ, MS, and DMW.
- Draft writing: NIA, MWA, JCB, ELB, TLPC, FD, OMG, JAGQ, JAK, SK, DS, NLQC, DMW

- Figures and tables: SB, FD, JAK, DMW
- Appendix II examples: KB, SB, FD, AKL, CCV, CON, DMW
- Metadata standardization: Led by JAK with contributions from MWA, JCB, ELB, FD, SK, JAGQ, PSS, DMW
- Validation: MWA, CON, JLV, HT

Acknowledgements

We gratefully acknowledge the staff, managers, and curators who steward biodiversity collections and make them accessible for scientific discovery. We thank Jorge M. Robles for artistic contributions to Figs. 2.1 and 6.1.

IHerbSpec has been supported by:

- US NSF Biology Integration Institute ASCEND (DBI-2021898)
- Harvard University Herbaria
- iDigBio (NSF DBI-2027654)
- NSF grant DBI-1930007
- Canadian NSERC Discovery Grant
- The Missouri Botanical Garden's Revolutionizing Species Identification project.

Individuals have received additional support from:

- ERC Horizon 2020 (grant 865787; TLPC and KB)
- Royal Botanic Garden Edinburgh – Botanics Foundation and American Friends of the Botanic Garden
- NYBG Visiting Scientist Travel Award (CON)
- FAPEAM 'Women in Science Program' (SPECTRA POP project, 2024; FD, MH, CCV)
- BMBF grant 01LK2101D (FD, MH, CCV)
- Danish National Research Foundation grant DNRF179 (JLV, HT)
- SURA award 80NSSC22M0001 (NLQC)

Suggested Citation

IHerbSpec. 2025. *Protocol for the Spectral Digitization of Herbarium Specimens, v.1.1.* <https://doi.org/10.5281/zenodo.1584966>

Navigate the protocol

- Proceed to [Part 1](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview

- **Part 2** – Measurement and Metadata Workflow
- **Part 3** – Filename Conventions and Formats
- **Part 4** – Metadata and Databasing
- **Part 5** – Instrumentation and Materials Guidelines
- **Part 6** – Selecting Tissues for Spectral Measurement
- **References**
- **Appendix I** – Number of Measurements per Tissue
- **Appendix II** – Tissue Metadata for Quality Control

Part 1 – Overview

1.1 Purpose and Structure of the IHerbSpec Protocol

This document outlines a “base” protocol for the spectroscopic measurement of herbarium tissues. It defines clear minimum requirements and recommended practices that promote data quality, methodological transparency, and interoperability across projects and institutions. In addition to these foundations, the protocol has been expanded to incorporate broader discussions surrounding instrumentation, materials, and herbarium tissues to further guide best practices.

The protocol specifies two categories of procedural elements:

- **Minimum requirements** – essential components to enable data harmonization that must be met for a project to be considered aligned with the IHerbSpec Protocol.
- **Recommended practices** – additions that enhance data quality and support broader utility, but are not mandatory.

The IHerbSpec Protocol is organized into modular parts:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

While consistency is central to aggregating spectral datasets, **the protocol is not intended to prescribe a rigid workflow**. The core requirements are designed to enable data harmonization through methodological consistency and transparency, while the overall structure preserves the flexibility needed for its broad adoption, continued innovation, and refinement.

1.2 Measurement and Metadata Overview

This section provides a quick conceptual guide to the required and recommended components of the protocol with relevant notes on adaptability.

1.2.1 Filename Conventions

To ensure key identifiability components for every unprocessed spectra file (white target, black target, calibrated reflectance standards, tissues), the IHerbSpec Protocol **strongly recommends** using the filename formats described in [Part 3](#).

Adaptability: Projects may use simplified filenames (see [Table 3.1](#)) during metadata scoring, then convert to the full format for permanent storage and distribution.

1.2.2 White Reference and Background Measurements

- White-reference measurements are **required** at the start of each session and at regular intervals (e.g., every 20 minutes; check manufacturer guidance).
 - This measurement is not recorded as a unique file (see [Section 5.3](#)).
 - A *measurement session* is defined as the period between switching the instrument on and off.
- White, black, and/or gray calibrated reflectance standard target measurements are **recommended** to be taken at the beginning of each measurement session or at regular intervals (e.g. once per week) and be locally archived (see [Section 5.3](#)).
- A single white target measurement is **required** at the beginning of each session to confirm instrument performance (see Fig. 1.1).
 - It is **strongly recommended** to use the `sessionId` field in filenames in order to link the white-reference spectrum to every tissue measurement in the session (see [Table 4.1](#)).
- The protocol **requires projects to use a <4% reflective black surface whenever possible** as a background behind tissues during measurement ([Section 5.3](#)).
 - A black-background target measurement is **required** at the start of each session (see Fig. 1.1).
 - It is **strongly recommended** to use the `sessionId` field in filenames in order to link the black-background spectrum to every tissue measurement in the session (see [Table 4.1](#)).
- If tissues are glued and must be measured against herbarium sheet paper as the background, then a **paper background target measurement** must be taken and linked to the corresponding tissue spectra (Fig. 1.1; [Part 3](#)).

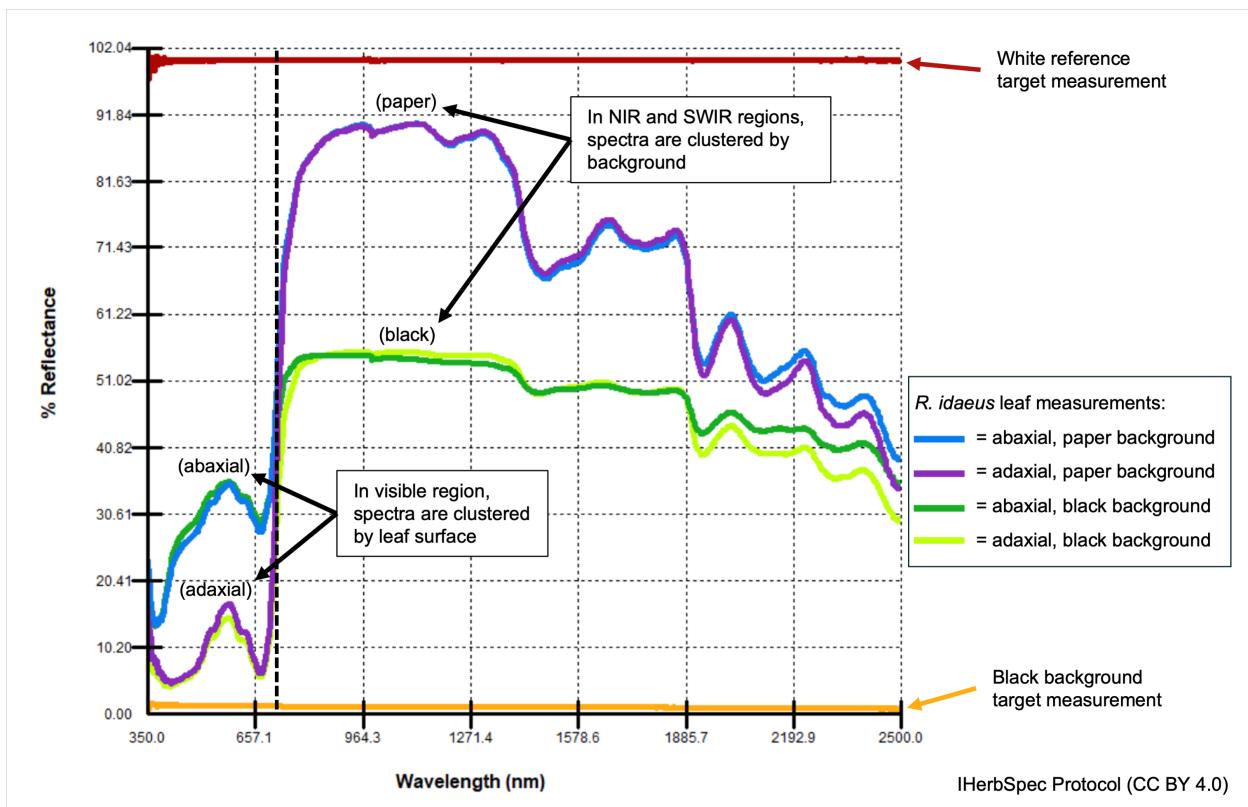


Fig. 1.1. Influence of background on leaf spectral measurements. For a single *Rubus idaeus* loose leaf (see NEBC 02618198 in [Appendix II](#)), the green adaxial surface and strongly glaucous abaxial leaf surface are separated spectrally in the visible region (400–700 nm) regardless of paper or black backgrounds. However, the spectra in the near-infrared (700–1,100 nm) and shortwave infrared (1,100–2,500 nm) regions are different as they cluster by the background material, revealing both biological and background contamination effects on reflectance spectra. The plot shows target measurements of the white reference, black background, and adaxial and abaxial leaf surfaces over black and paper (herbarium sheet) backgrounds. All measurements were made with an SVC HR-1024i spectroradiometer with the LC-RP Pro leaf clip.

1.2.3 Strategy and Number of Tissue Measurements

- **Tissue selection strategy**
 - Considerations and guidelines for selecting specimens and tissues for measurement are provided in [Part 6](#), including a decision tree diagram (Fig. 6.1).
- **The number of leaf tissue measurements per specimen**
 - If both adaxial and abaxial leaf surfaces are accessible and suitable, **a minimum of three representative measurements per surface (six total) is required per specimen**. If only one surface is available or if the tissue class is **leaf** with undifferentiated abaxial/adaxial surfaces (see [Table 4.5](#)), then a minimum of three representative measurements are required.

- * Representative measurements should be free of technical errors and generally reproducible (see Section 6.2; [Appendix I](#)).
- Five or more representative measurements per available and suitable leaf tissue class (adaxial, abaxial, or leaf) are recommended (see [Appendix I](#) for discussion).

Adaptability: Projects may choose to collect measurements from a single leaf or from multiple leaves to capture tissue- versus specimen-level variability. The different tissue units are recommended to be distinguished using the `targetTissueId` field and be identifiable via annotation labels and/or their location on the specimen (see [Table 4.3](#)).

- **Other tissue classes**
 - **No required minimum** is prescribed.
 - **Recommended:** At least three representative measurements per specimen per tissue class, if feasible.

1.2.4 Measurement Technique

- Multiple measurements per tissue unit may be taken by targeting multiple suitable areas on the tissue or—taking care not to heat-damage the tissue—by targeting a small area and slightly rotating or shifting the optical probe between measurements.
- Avoid overlapping tissue layers, debris, or partial coverage of the probe aperture (Fig. A1).
- If a measurement appears affected by technical error, delete and repeat it.

1.2.5 Metadata Requirements and Flexibility

- All required and recommended protocol metadata are available for use in the IHerbSpec Metadata Spreadsheet, which can be downloaded on the [protocol home page](#).
- The protocol recommends recording tissue metadata for every individual measurement because every measurement will generate a spectral data file. This means that **every tissue measurement will be recorded in a unique row in the metadata spreadsheet** (e.g., 10 measurements have 10 rows).
- All fields in Session metadata ([Table 4.1](#)), Specimen Metadata ([Table 4.2](#)), and Tissue Metadata ([Table 4.3](#)) are marked as required or recommended in the field descriptions.

1.2.6 Scoring Metadata Pertaining to Tissue Condition and Contamination

- To record key elements of tissue condition and contamination, these fields are required metadata (see Section 6.1; field descriptions in [Table 4.3](#)):

- `tissueDevelopmentalStage`
 - `hasGlue`
 - `hasNonGlueContamination`
- To further describe tissue conditions, the `measurementFlags` and `tissueNotes` fields are **recommended**:
 - `measurementFlags` allows use of standardized tissue condition descriptors (see [Table 4.6](#)) for preservation status, damage, or contamination (e.g., `GoodPreservation|GluePresent`).
 - `tissueNotes` enables free-text documentation of additional tissue conditions or context inside or outside the measurement area.

Adaptability:

- `tissueDevelopmentalStage` may be coded as `notScored` if assessment is not possible (e.g., due to technician training or specimen condition; see [Table 4.4](#)).
- `hasGlue` and `hasNonGlueContamination` may be marked `uncertain` if presence is not confidently determined.

1.2.7 Specimen and Tissue Annotation

- The IHerbSpec Protocol recommends annotating the location of measured tissues on the sheet using the `targetTissueId` field. Additional annotations to the herbarium sheet or packet are encouraged in accordance with herbarium policies (see [Section 5.4](#)).
-

Navigate the protocol

- Proceed to [Part 2](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Part 2 – Measurement and Metadata Workflow

Overview

The recommended step-by-step workflow for spectral data collection and metadata entry. Steps include references to protocol sections providing further details.

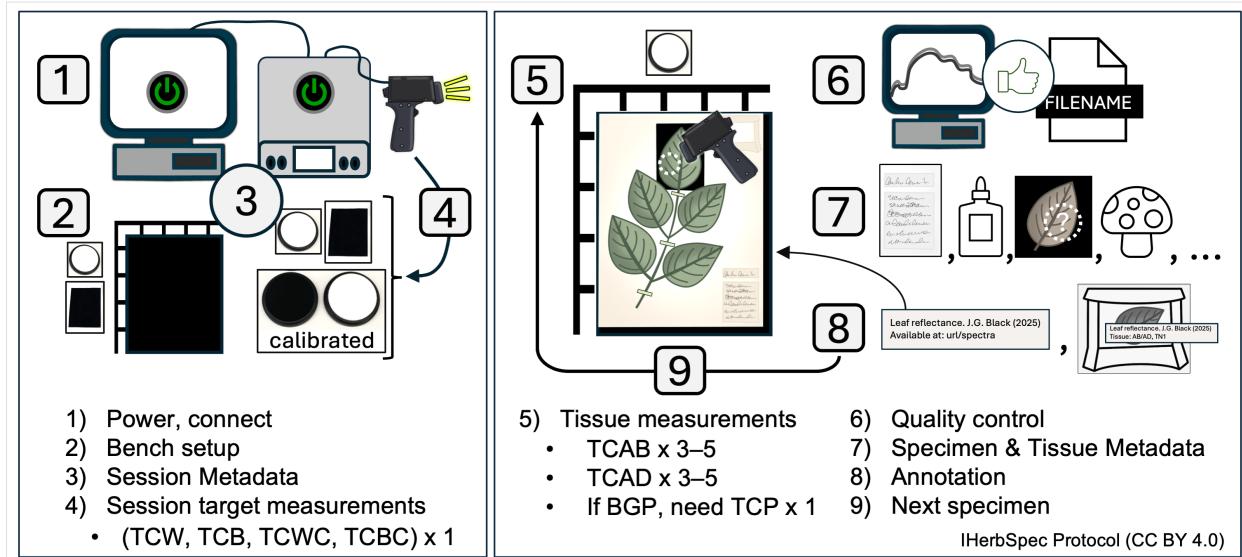


Fig. 2.1. Diagram of IHerbSpec Protocol measurement and metadata workflow. Steps 1–4 describe the setup procedure and steps 5–9 describe the measurement sequence. For decisions regarding specimen tissue selection and using black backgrounds, see [Part 6](#).

Step 1. Instrument Setup and Connection

- 1.1. Plug in and turn on the instrument and light source.
- Allow >15 minutes or follow manufacturer SOPs for lamp warm-up and sensor cool-down.
- 1.2. Set up the computer and software, and connect the instrument following project SOPs.

Step 2. Prepare Bench and Specimens

- 2.1. Select specimens and tissues for measurement (see [Part 6](#); Fig. 6.1).
- 2.2. Optionally, prepare filenames (see [Part 3](#)) in a separate document for copy–paste entry.
- 2.3. Prepare bench with the following materials (see Fig. 2.2):
 - **Required:** white reference, black background
 - **Recommended:** benchtop black background, rulers for tissue coordinates, tweezers, laboratory gloves (see [Part 5](#)), annotation labels, archival envelopes (see [Section 5.4](#)).

Step 3. Start Session and Score Session Metadata

3.1. Score Session Metadata (see [Table 4.1](#)):

- Record `sessionId` datetime in the format YYYYMMDDHHMM (e.g., 202507011351).
- Score **required fields**: `projectID`, `sessionId`, `instrumentModel`, `opticalSetupDescription`, `measurementSettings`, `whiteReferenceDescription`.
- Score **recommended fields**: `operator`, `lightSourceType`, `distanceTargetToSensor`, `lensFieldOfView`, `angleLightToSensor`, `measurementAreaDiameter`.

3.2. Create a storage folder for spectral data named with the `sessionId` (e.g., 202507011351) and set it as the destination folder for saving measurement files.

Step 4. Collect White and Black Session Measurements

4.1. Take a white reference.

4.2. Set simple filename for white target (see [Table 3.1](#)):

- `PI<projectId>_SN<sessionId>_TCW`
- Example: PIHUhCoca_SN202507011351_TCW

4.3. Take one white target measurement.

4.4. Set simple filename for black target:

- `PI<projectId>_SN<sessionId>_TCB`
- Example: PIHUhCoca_SN202507011351_TCB

4.5. Take one black target measurement.

4.6. Optionally, take target measurements of each calibrated reflectance standard (see [Section 5.3](#)) using filename conventions ([Table 3.1](#)):

- `PI<projectId>_SN<sessionId>_TC<targetClass><serial number>` - Example: PIHUhCoca_SN202507011351_TC

Step 5. Tissue Measurement Sequence

5.1. Take white reference at regular intervals (e.g., every 20 min.; see [Section 1.2.2](#)).

5.2. Determine if tissue is suitable for measurement (see [Part 6](#)). If not, skip to the next specimen.

5.3. Place specimen on benchtop black background with a ruler grid (recommended; see Fig. 4.2).

5.4. Place black background behind target tissue (required whenever possible; see Fig. 2.2).

5.5. *Adaxial (AD) surface available:*

- Filename: `SI<specimenId>_TC<targetClass>_TN<targetTissueId>`
- Example: SI02022418_TCAD_TN1
- Take 3–5 adaxial measurements.

5.6. *Abaxial (AB) surface available:*

- Filename: `SI<specimenId>_TC<targetClass>_TN<targetTissueId>`
- Example: SI02022418_TCAB_TN1
- Take 3–5 abaxial measurements.

5.7. *If tissue measured on herbarium paper:*

- Filename: `SI<specimenId>_TC<targetClass>_TN<targetTissueId>`
- Example: SI00746092_TCP_TN1
- Take 1 paper measurement.

5.8. *Additional tissues (optional):*

- Select other tissue units with new `targetTissueIds` (recommended) and/or additional `targetClass` values (see [Table 4.5](#)).
- Repeat tissue measurement steps.

Step 6. Quality Assessment, Quality Control

- 6.1. Visually review each spectrum.
- 6.2. Delete and repeat any measurement that appears anomalous.
- 6.3. If unsure, take additional measurements.
- 6.4. Check for and fix any filename errors before proceeding.

Step 7. Score Specimen and Tissue Metadata

- 7.1. Score **Specimen Metadata** for every measurement (see [Table 4.2](#)):

- Required: `herbariumCode`, `specimenId`
- Recommended: `scientificName`, `identificationQualifier`, `identifiedBy`, `dateIdentified`, `isTempControlled`, `annualTempMin`, `annualTempMax`, `isHumidityControlled`, `annualHumidityMin`, `annualHumidityMax`

- 7.2. Score **Tissue Metadata** for every measurement (see [Table 4.3](#)):

- Required: `backgroundClass`, `hasLowReflectanceBackground`, `targetClass`, `tissueDevelopmentalStage`, `hasBackgroundInMeasurement`, `hasGlue`, `hasNonGlueContamination`, `measurementIndex`
- Recommended: `backgroundDescription`, `targetTissueId`, `percentBackgroundInMeasurement`, `measurementFlags`, `tissueNotes`, `tissueLocation`, `comment`

Step 8. Specimen and Tissue Annotation

- 8.1. Add project annotation label to sheet (**recommended**; see [Section 5.4](#)).
- 8.2. Annotate loose tissues in packets/envelopes with `targetTissueClass` and `targetTissueId` labels; store in an envelope on the sheet (see Fig. 4.2).
- 8.3. *If applicable*, annotate attached target tissues on herbarium sheet (unless already recorded in `targetLocation`).

Step 9. Move to Next Specimen

- 9.1 Repeat the sequence beginning at [Step 5](#).
-

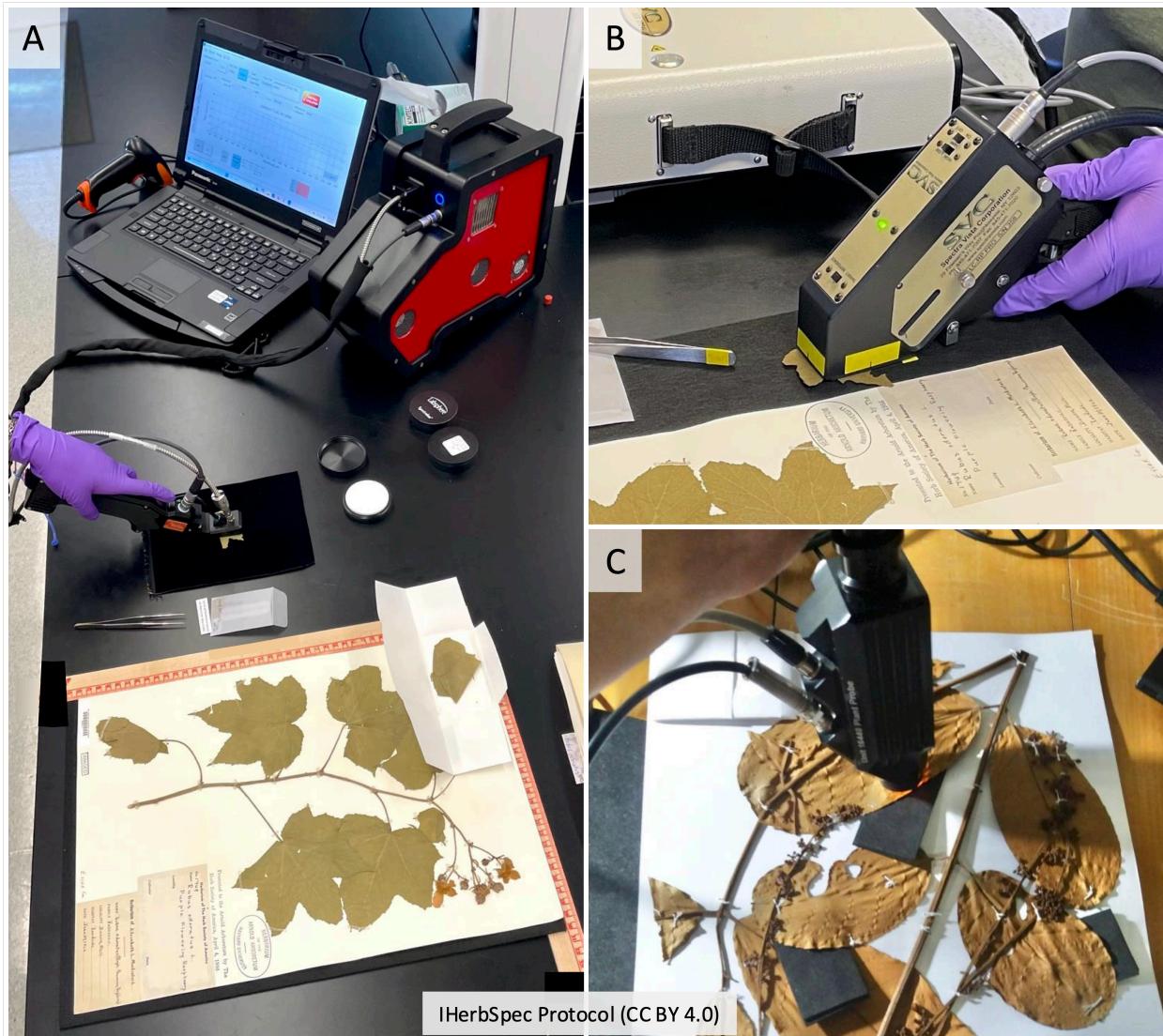


Fig. 2.2. Reflectance spectroscopy bench setup and optical probe measurements. (A) Benchtop setup with Spectral Evolution NaturaSpec spectroradiometer with herbarium probe measuring a loose leaf fragment on black background. The setup also includes a barcode scanner, a laptop with data acquisition software and display for quality control and recording metadata, 2-inch diameter white and black calibrated reflectance standards, a 2-inch diameter Spectralon® white reference (top removed), a benchtop black background underneath the herbarium sheet, and a tissue annotation label stored in a glassine envelope inside the herbarium sheet packet. (B) Spectral Vista Corporation HR-1024i spectroradiometer and LC-RP Pro leaf clip (with clip removed) measuring a loose leaf fragment over a black background. (C) ASD optical probe measuring mounted tissue with black background slid underneath. Photos A and B by D.M. White; specimen from A: Herbarium of the Arnold Arboretum of Harvard University. Photo C by F. Durgante.

Navigate the protocol

- Proceed to [Part 3](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Part 3 – Filename Conventions and Formats

Overview

The full and simple filename conventions specified in [Table 3.1](#) provide a consistent format for encoding key metadata into filenames, supporting traceability, metadata parsing, and long-term interoperability.

Each filename follows a defined format, built from metadata-coded segments with standardized prefixes (see [Table 3.2](#)). These are **recommended filename standards** designed for reliable parsing regardless of segment order, **corresponding to required metadata fields** described in [Part 4](#). The spectrometer software will automatically append a measurement index (**IDX**) to the end of each filename—no additional text should be added after the **IDX**.

Filename formats differ slightly depending on the type of target material (i.e., `targetClass`; see [Table 4.5](#)). For white and black-background targets, the `projectId` (`PI`) and `sessionId` (`SI`) are key segments for traceability and the `SI` is also critical for linking to all associated tissue measurements via the tissue full filename. For tissues, the full filename convention also includes the minimum session-level, specimen-level, and tissue-level metadata needed for confident data aggregation.

A simplified filename convention may be used during measurement sessions to streamline data collection. However, projects should convert all filenames to the full format before archiving or sharing. When using the simplified format for local files, projects should maintain consistent file organization and take precautions to prevent ambiguity.

Tables in Part 3 can be downloaded from iherbspec.github.io/protocol.

3.1 Recommended Filename Conventions

Table 3.1: Full and Simple Filename Conventions with Different Measurement Target Classes

Note: The filenames should have **no blank spaces**.

Target (<code>targetClass</code>)	Convention Type	Filename format and example
White target (TCW)	Full	PI<projectId>_SN<sessionId>_TC<targetClass>_<IDX> Example: PIERYspec1_SN202406180932_TCW_0001
Black background target (TCB)	Full	PI<projectId>_SN<sessionId>_TC<targetClass>_<IDX> Example: PIERYspec1_SN202406180932_TCB_0001
White calibrated reflectance standard (TCWC)	Full	PI<projectId>_SN<sessionId> _TC<targetClass><serialNumber>_<IDX> Example: PIFagaceae_SN202506171532_TCWC7254_0000

Target (targetClass)	Convention Type	Filename format and example
Black calibrated reflectance standard (TCBC)	Full	PI<projectId>_SN<sessionId> _TC<targetClass><serialNumber>_<IDX> Example: PIFagaceae_SN202506171532_TCBC7210_0000
Tissue target on black background (BGB + TCAD/TCAB)	Full	PI<projectId>_SN<sessionId>_BG<backgroundClass> _HC<herbariumCode>_SI<specimenId>_TC<targetClass> _TN<targetTissueId>_<IDX> Example: PIERYspec1_SN202406180932_BGB_HCGH _SI02022418_TCAD_TN1_001
Tissue target on black background (BGB + TCAD/TCAB)	Simple	SI<specimenId>_TC<targetClass>_TN<targetTissueId>_<IDX> Example: SI02022418_TCAD_TN1_0001
Tissue target on paper (BGP + TCAD/TCAB)	Full	PI<projectId>_SN<sessionId>_BG<backgroundClass> _HC<herbariumCode>_SI<specimenId>_TC<targetClass> _TN<targetTissueId>_<IDX> Example: PIERYspec1_SN202406180932_BGP_HCNEBC _SI00746092_TCAD_TN1_001
Tissue target on paper (BGP + TCAD/TCAB)	Simple	SI<specimenId>_TC<targetClass> _TN<targetTissueId>_<IDX> Example: SI00746092_TCAD_TN1_0001
Paper target (BGB + TCP)	Full	PI<projectId>_SN<sessionId>_BG<backgroundClass> _HC<herbariumCode>_SI<specimenId>_TC<targetClass> _TN<targetTissueId>_<IDX> Example: PIERYspec1_SN202406180932_BGB_HCNEBC _SI00746092_TCP_TN1_001
Paper target (BGB + TCP)	Simple	SI<specimenId>_TC<targetClass>_TN<targetTissueId>_<IDX> Example: SI00746092_TCP_TN1_0001

3.2 Filename Components

[Table 3.2](#), below, defines the components of filenames and their direct links to metadata fields. Each segment of the filename follows a specific format, with a standardized prefix that links directly to a **required** metadata field (see [Part 4](#)). These components form the structured **conventions** shown in [Table 3.1](#) and enable automatic parsing and alignment between spectral files and metadata records.

Table 3.2: Filename Components and Corresponding Metadata Fields.

Code	Metadata field	Description	Example
PI	projectId	Unique identifier for the spectral measurement project (Table 4.1).	PIHUhcoa, PIFagales1

Code	Metadata field	Description	Example
SN	sessionId	A unique identifier generated from date/time when the session begins (SNYYYYMMDDHHMM; (Table 4.1)).	SN202406180932
BG	backgroundClass	Enumerated code from Background Class Codes (Table 4.3).	BGB, BGP, BGO
HC	herbariumCode	Herbarium acronym or collection identifier (Table 4.2).	HCGH, HCINPA
SI	specimenId	Specimen ID (GUID, barcode, accession no., collector name + number; Table 4.2).	SI03774853, SIThorne24070
TC	targetClass	Enumerated code from Target Class Codes (Table 4.3).	TCAD, TCAB, TCW, TCP, TCB
TN	targetTissueId	Index tracking measured tissue units (e.g., 1, 2, ...; Table 4.3).	TN1, TN2
IDX	measurementIndex	Auto-appended by spectrometer software (Table 4.3).	0001, 0002

Navigate the protocol

- Proceed to [Part 4](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Part 4 – Metadata and Databasing

Overview

This section defines the metadata fields used to describe spectral reflectance measurements of herbarium specimens. This standardization supports interoperability, cross-project data aggregation, and future integration into biodiversity informatics platforms.

Projects can download the **IHerbSpec Metadata Spreadsheet** (available at doi.org/10.5281/zenodo.1584966) as a foundation for organizing their metadata. This spreadsheet includes all **required** and optional but **recommended** fields defined in the tables in [Section 4.1](#). The controlled vocabularies necessary for enumerating certain fields are provided in [Section 4.2](#). Finally, [Section 4.3](#) provides general guidelines for data and metadata storage and dissemination.

While metadata are expected to be disseminated in a flat file format (as in the IHerbSpec Metadata Spreadsheet), fields are presented here in logical groups—project, specimen, and tissue—to support conceptual understanding and integration with local databases.

Tables in Part 4 can be downloaded from iherbspec.github.io/protocol.

4.1 Metadata Tables

Table 4.1: Session Metadata

Session metadata are those metadata usually associated globally with a continuous digitization project. These metadata can usually be captured once per project and automatically populated for each measurement instance.

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
projectId	PI	Required	Unique identifier for the spectral measurement project. Example: HUHERYspec1.	TEXT
sessionId	SN	Required	Unique identifier for a measurement session, generated as YYYYMMDDHHMM. Example: 20240617132251.	TEXT
instrumentModel	-	Required	Spectroradiometer model name. Example: SVC HR-1024i.	TEXT

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
<code>opticalSetupDescription</code>		Required	Description of optical probe setup. Example: <code>LC-RP contact probe with leaf clip removed.</code>	TEXT
<code>measurementSettings</code>		Required	Instrument settings for measurements. Example: <code>2 seconds, high light setting.</code>	TEXT
<code>whiteReferenceDeserption</code>		Required	Material of the white reference. Example: <code>Spectralon SL Standard 99%.</code>	TEXT
<code>operator</code>	-	Optional	Name(s) of person(s) conducting measurements.	TEXT
<code>lightSourceType</code>	-	Optional	Light source for optical setup. Example: <code>tungsten halogen.</code>	TEXT
<code>distanceTargetToSensor</code>		Optional	Distance (mm) between target tissue and sensor face. Example: 12.	NUMERIC
<code>lensFieldOfView</code>	-	Optional	Angle (degrees) of sensor field of view. Example: 22.5.	NUMERIC
<code>angleLightToSensor</code>		Optional	Angle (degrees) of light source to sensor. Example: 10.	NUMERIC
<code>measurementAreaDiameter</code>		Optional	Diameter (mm) of illuminated tissue area. Example: 6.	NUMERIC

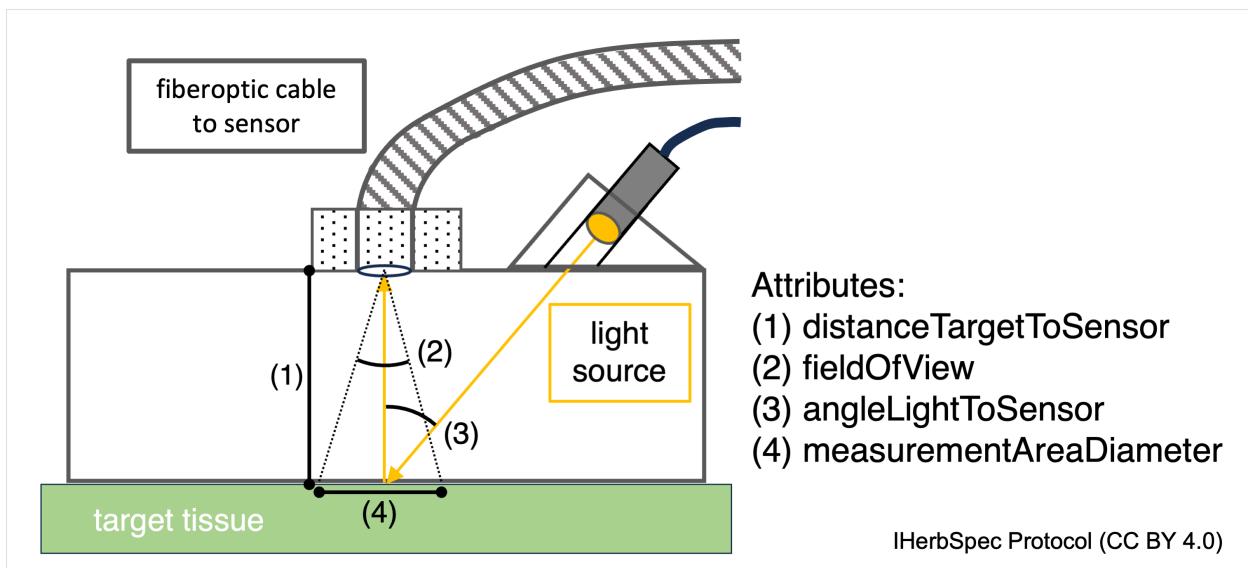


Fig. 4.1. Schematic of Session Metadata fields related to optical setup (see [Table 4.1](#)). `distanceTargetToSensor`, `fieldOfView`, `angleLightToSensor`, `measurementAreaDiameter`

Table 4.2: Specimen Metadata

Specimen metadata include identifiers and information about the physical specimen, with priority given to **required** fields needed to link spectral measurements to existing digital records (e.g., `specimenId`). Optional fields related to taxonomic determination and specimen storage environment are included to support integrative research, quality control, and downstream analysis.

Users should **avoid duplicating metadata that are already digitized, maintained, and available in herbarium or institutional platforms**, as these sources are better suited for future updates. Instead, users are encouraged to reference those records and supplement only missing required or recommended fields. Due to variation in the metadata recorded on institutional platforms, users should apply caution in the interpretation of presence or absence of determination information or any other recommended but optional fields.

Metadata Field	Code (Table 3.2)	Status	Field Description	Data Type
herbariumCode	HC	Required	Acronym for herbarium or collection (Index Herbariorum code). Examples: GH, P, US-Botany.	TEXT
specimenId	SI	Required	Identifier for specimen or record (catalog no., barcode, GUID, or collector+number). Examples: 00238762, Thorne24070a.	TEXT

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
<code>scientificName</code>	-	Optional	Full scientific name at lowest confident rank. Examples: <i>Quercus bicolor</i> , <i>Erythroxylum coca ipadu</i> .	TEXT
<code>identificationQualifier</code>		Optional	Uncertainty in taxonomic ID (Darwin Core). Examples: cf., aff.	TEXT
<code>identifiedBy</code>	-	Optional	Person/group who made the identification. Example: T. Plowman.	TEXT
<code>dateIdentified</code>	-	Optional	Date of identification. Examples: 1999, 2004-12-30.	TEXT
<code>isTempControlled</code>	-	Optional	Whether storage has active temperature control.	BOOLEAN
<code>annualTempMin</code>	-	Optional	Minimum annual storage temperature (°C). Example: 18.	TEXT
<code>annualTempMax</code>	-	Optional	Maximum annual storage temperature (°C). Example: 26.	TEXT
<code>isHumidityControlled</code>		Optional	Whether storage has active humidity control.	BOOLEAN
<code>annualHumidityMin</code>		Optional	Minimum annual storage relative humidity (%). Example: 20.	TEXT
<code>annualHumidityMax</code>		Optional	Maximum annual storage relative humidity (%). Example: 60.	TEXT

Table 4.3: Tissue Metadata

Fields describing the type, condition, and position of the tissue measured. Includes **required** and **recommended** metadata for linking spectral measurements to individual tissue units. Note that timestamps for tissue measurement files are often captured within the file.

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
<code>backgroundClass</code>	BG	Required	Enumerated abbreviated code from Background Class Codes (Table 4.7) describing the type of background used behind target tissue. Both abbreviated codes and descriptive codes are accepted. Examples: BGW, BGB, BGP.	ENUM (Table 4.7)

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
hasLowReflectanceBackground -		Required	True or False statement that the background (black or paper) has low reflectance as defined as less than 4% reflectance across the spectral range of the instrument. For a paper background, this would be scored false .	BOOLEAN
backgroundDescription	-	Cond. Req.	Description of the black or other background material, including manufacturer and product information when available. Not required for paper backgrounds. Required field when tissue has black or other type (not paper) of background. Example: Musou IR Flock	TEXT
targetClass	TC	Required	Free text or enumerated code from Target Class Codes (Table 4.5) describing type of tissue or background being measured. Both abbreviated codes and full codes can be used. Examples: AD, perigynium .	TEXT or ENUM (Table 4.5)
targetTissueId	TN	Optional	Character index tracking the measured tissue units when multiple tissues are measured from a single specimen (e.g., loose1, 1, 2). For compound or more complex structures, projects are encouraged to develop their own consistent naming convention. Examples: loose1, leaflet1, petal1 .	TEXT
tissueDevelopmentalStage	-	Required	Tissue developmental stage as coded in Developmental Stage Class Codes Table 4.4 . Examples: Mature, Uncertain .	ENUM (Table 4.4)

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
hasBackgroundInMeasurement	-	Required	True or false values indicating that the target tissue does not cover the full measurement area and the background is part of the measurement.	BOOLEAN
percentBackgroundInMeasurement		Optional	Numeric estimate for the percentage of the measurement area that is not covered by the target tissue and is background material (black background or herbarium paper). It is recommended to describe the estimation method in the comment field. Example: 25.	INTEGER
hasGlue	-	Required	true/false/uncertain: glue present in measurement area.	ENUM (true/false/ uncertain)
hasNonGlueContamination	-	Required	True, false, or uncertain values indicating a contaminant other than glue is present in the measurement area. This includes foreign biotic or abiotic agents on the target tissue, such as fungus or preservatives.	ENUM (true/false/ uncertain)
measurementFlags	-	Optional	Standardized categorical descriptors of the condition of the tissue within the measurement area. Values are selected from the predefined Tissue Descriptor Codes Table 4.6 . Multiple descriptors should be separated with a pipe character (). Example: GoodPreservation PathogenPresent.	ENUM (Table 4.6)

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
<code>tissueNotes</code>	-	Optional	<p>Free-text field used to record additional observations on the condition of the specimen that may aid interpretation of spectral data. It can be used to clarify or elaborate on descriptors already included in measurementFlags, such as the conditions evidencing the quality of preservation (e.g., a MediumPreservation flag could be explained with the note, ‘measurement area discolored and wrinkled’). Notes should specify whether the information applies to the measurement area, the tissue unit, or the specimen as a whole. Examples: <code>mold in measurement area, formaldehyde preserved.</code></p>	TEXT
<code>tissueLocation</code>	-	Optional	<p>The location of the target tissue on the herbarium sheet. For mounted tissues, record as an X,Y coordinate in centimeters, with 0,0 at the top-left corner of the sheet (e.g., 17,29; see Fig. 4.2). If the sheet has non-square angles, align it flush with the left-side ruler. For unmounted tissues, provide a descriptive note indicating location. Examples: 17,29, envelope TCAD_TN1.</p>	TEXT (coordinates preferred)

Metadata Field	Filename Code (Table 3.2)	Status	Field Description	Data Type
<code>comment</code>	-	Optional	Free-text field for recording any additional notes relevant to the measurement, including observations about the instrument, session, specimen, tissue, or data quality that are not captured elsewhere in the metadata. Example: <i>Amazing specimen.</i>	TEXT
<code>measurementIndex</code>	IDX	Required	The measurement number index appended to the base filename (Part 3, Table 3.2: IDX) to properly associate each row of metadata with its single, corresponding measurement file. Example: 0001.	TEXT

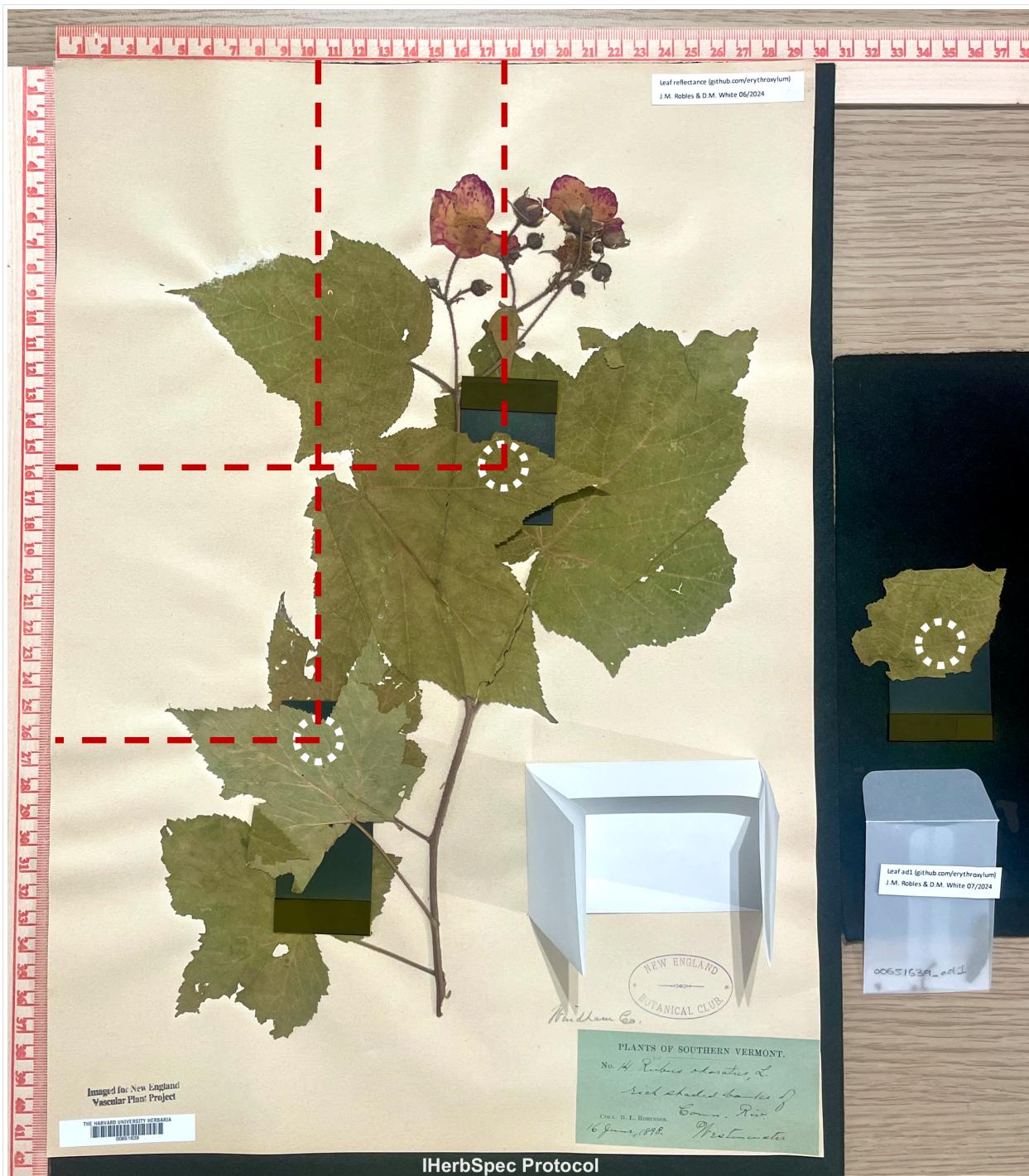


Fig. 4.2: Diagram of coordinate system for scoring `tissueLocation` in Table 4.3: Tissue Metadata. Herbarium sheet is placed on top of the benchtop black background with centimeter rulers at top and left sides for ‘x,y’ notation of measurement area (white dashed circles) in centimeters with 0,0 at the top left. Black background cards are placed under unglued portions of leaves. From left to right, tissue `TCAB_TN3` has the location 10,26, tissue `TCAD_TN2` has location 17,16 and tissue `TCAD_TN1` is stored with a label in a glassine envelope inside the packet with location `envelope_TCAD_TN1`. For reference, specimen `NEBC_00651639` metadata fields are proposed in [Appendix II](#). Specimen courtesy of the New England Botanical Society.

4.2 Controlled Vocabularies

Table 4.4: Developmental Stage Class Codes

Available codes for enumerating the required `tissueDevelopmentalStage` metadata [Table 4.3](#). Codes follow ‘CamelCase’ format with capitalized initial letters.

Code	Description
Young	Actively developing tissue that is not yet fully expanded; may appear thinner, lighter in color, or more pliable than mature tissue.
Mature	Fully developed and expanded tissue showing typical structural and color characteristics for the taxon; not visibly senescent.
Old	Senescent tissue showing visible signs of aging or decline, such as yellowing, darkening, curling, or drying.
Uncertain	The development stage has been assessed but cannot be confidently determined due to intermediate features, damage, or insufficient visual cues.
NotScored	Developmental stage was not assessed or recorded for this tissue.

Table 4.5: Target Class Codes.

Available codes for enumerating the required `targetClass` metadata [Table 4.3](#). Either the abbreviated code or the full CamelCase-formatted code (with an initial capital letter) may be used.

Code	Full code	Description
W	<code>WhiteReference</code>	White reference.
WC	<code>WhiteCalibratedReference</code>	White calibrated reflectance standard (see Section 5.3).
B	<code>BlackBackground</code>	Black background material, recorded when used as background for other target tissue measurements.
BC	<code>BlackCalibratedReference</code>	Black calibrated reflectance standard (see Section 5.3).
P	<code>Paper</code>	Herbarium sheet paper, recorded when used as background for other target tissue measurements.
AB	<code>LeafAbaxial</code>	Abaxial leaf surface.
AD	<code>LeafAdaxial</code>	Adaxial leaf surface.
LF	<code>Leaf</code>	Leaf surface. Applied when the abaxial and adaxial side cannot be differentiated or when leaves are terete or otherwise not bifacial (e.g. <i>Curio rowleyanus</i>).
PT	<code>Petal</code>	Petal.
IF	<code>Inflorescence</code>	Inflorescence.
BR	<code>Bract</code>	Bract.

Code	Full code	Description
FR	Fruit	Fruit. The specific tissue (e.g., exocarp, mesocarp) can be described in tissueNotes.
PSS	Photosynthetic-SucculentStem	Photosynthetic stem as in succulents like Cactus.
OB	OuterBark	Outer bark, rhytidome. Woody branch outer bark as in Hadlich et al. (2018).
IB	InnerBark	Phloem. Woody branch inner bark as in Hadlich et al. (2018).
HS	HerbaceousStem	Herbaceous stem; can be photosynthetic as in PSS.
WD	Wood	Wood.

Table 4.6: Tissue Descriptor Codes

Code	Description
GoodPreservation	Tissue in the measurement area appears well preserved in color, structure, and texture (including original features from disease or herbivory) and shows minimal signs of degradation or breakage from pressing, drying, or storage. Tissues that are simply discolored may still be considered well preserved if other aspects of integrity are maintained.
MediumPreservation	Tissue in the measurement area shows moderate degradation, such as partial discoloration, wilting, or deformation. Some structural loss may be present, though not severe. Evidence of degradation from other parts of the specimen (e.g., mold elsewhere on the tissue) may support assigning this level of preservation, but such observations should be recorded in tissueNotes if not present within the measurement area. This is expected to be the most common preservation condition for herbarium specimens used in spectral measurement.
PoorPreservation	Tissue in the measurement area shows clear signs of degradation, including severe discoloration, wrinkling, deformation, or breakage. Mold, insect damage, or other signs of poor preservation may also be present. As with other flags, evidence of degradation outside the measurement area (e.g., mold elsewhere on the tissue) may support the assigned flag but should be recorded in tissueNotes if not directly observed in the measurement area. Note that natural discoloration tendencies of certain taxa should be considered when applying this flag (see Appendix II).
MidveinPresent	Target measurement area contains midvein or similarly prominent secondary venation.

Code	Description
OrganismPresent	Indicates that a visible organism (e.g., bryophyte, lichen, fungal structure) is present on or within the measurement area. This includes epiphyllous, endophytic, or other leaf-associated organisms, regardless of their ecological role (e.g., mutualistic, parasitic, or incidental). This flag serves as a general indicator and can be used in conjunction with more specific organism flags below.
BryophytePresent	Indicates that a visible bryophyte (e.g., moss, liverwort, or hornwort) is present on or within the measurement area.
LichenPresent	Indicates that a visible lichen thallus or fragment is present on or within the measurement area.
FungusPresent	Indicates that fungal structures are visible in the measurement area (e.g., hyphae, mycelium, fruiting body) and are presumed to be pre-mortem associates, such as endophytes or pathogens active while the plant was alive. This flag can overlap with PathogenPresent or MoldPresent.
PathogenPresent	Target measurement area contains necrotic tissue or other signs of pathogenic infection. Can be used with FungusPresent when fungal pathogens are suspected or known. This flag is based on visible tissue symptoms, not molecular confirmation.
MoldPresent	Indicates that the measurement area shows signs of post-mortem fungal growth (e.g., surface mold, fuzz, bloom), likely resulting from poor drying or storage conditions. In practice, mold may be difficult to distinguish from other fungal growth without microscopic or culture analysis. Use judgment and note uncertainty in tissueNotes if needed.
HerbivoryPresent	Target measurement area contains herbivory.
AlcoholPresent	Target measurement area was preserved with ethanol or other alcohol.
PreservativePresent	Target measurement area contains chemical preservative contamination excluding alcohol preservatives (e.g., diatoms, formaldehyde).
BurnPresent	Target measurement area contains burned tissue.
DebrisPresent	Target measurement area contains non-specific material not described in other codes (e.g., dust, soil particles, insect parts, fibers, etc.) that may interfere with clean measurements. Can be elaborated in the tissueNotes field.

Table 4.7: Background and White Reference Class Codes

Available codes for enumerating the required `tissueBackgroundClass` metadata [Table 4.3](#). Either the abbreviated code or the full CamelCase with initial capital letter jformatted code may be used.

Code	Full code	Description
W	WhiteReference	White reference.
B	BlackBackground	Black background (<4% reflectance).
P	PaperBackground	Herbarium sheet paper.
O	OtherBackground	Other background material (must be described in metadata).

4.3 Guidelines for Data Archiving and Sharing

- For each specimen, a copy of **all unprocessed spectral files and associated metadata should be archived** by the herbarium or institution that owns the specimen, or otherwise managed in accordance with institutional data storage practices. Wherever possible, these data should remain **co-located with specimen records** or stored in infrastructure (e.g., institutional servers, cloud services, or collection management systems) that supports long-term stewardship, identifier stability, and integration with specimen metadata.
- In addition to institutional archiving, projects are strongly encouraged to **deposit unprocessed spectral data files and metadata in a persistent, open-access repository** that issues DOIs and supports versioning.

The **IHerbSpec Dataverse** (<https://dataverse.harvard.edu/dataverse/iherbspec/>) is a community-curated, thematic repository designed specifically to serve herbarium spectral data and metadata. It provides free access, dataset-level versioning, rich metadata support, and curation aligned with the IHerbSpec Protocol. Datasets deposited in the IHerbSpec Dataverse can be linked to specimen records in primary digitization platforms (e.g., Symbiota, Specify, Tropicos) using stable identifiers and resource relationships, enabling integration across systems while allowing spectral data to be managed in infrastructure optimized for these data types.

- The IHerbSpec Dataverse supports a **project-based data sharing model**, suitable for laboratory studies, collaborative research projects, and digitization efforts that span multiple collections. Datasets may include raw spectra, metadata spreadsheets, README documentation, processed spectra, and associated trait data. Metadata are reviewed prior to release to promote compliance with the IHerbSpec Protocol and to facilitate interoperability and reuse.
- To enhance discoverability and specimen-level integration, datasets archived in external repositories (including the IHerbSpec Dataverse) may be linked to specimen records using standardized associations (e.g., Darwin Core Resource Relationship terms such as `isReferencedBy`). These links can be prepared in bulk and ingested by specimen portals, allowing users to search for specimens associated with spectral data while the files themselves remain hosted in the most appropriate repository.
- To facilitate reuse and analysis, projects may also share **tabular data files** (e.g., `.csv`) containing samples in rows and spectral bands and metadata fields in columns. If processed spectra are included (e.g., interpolated, splice-corrected, continuum-removed), all applied processing steps should be clearly documented.
- When projects include **additional trait data** alongside spectral measurements, contributors should ensure that trait definitions, units, and measurement methods are clearly documented

and standardized where possible. Ongoing community efforts will be required to align spectral data with broader trait data standards and repositories.

- Contributors should consider data governance needs, including **sensitive specimen information, traditional knowledge considerations, access restrictions, and long-term curation responsibilities**. The IHerbSpec community is actively developing governance, curation, and review practices to support responsible data sharing consistent with FAIR and CARE principles.
- The IHerbSpec community is also working toward the development of **software tools to automatically validate metadata, filenames, and dataset structure** against the IHerbSpec Protocol during data submission, and to enable future parsing of bulk datasets into specimen-specific subsets for downstream integration and reuse.
- See an example of spectral data and metadata sharing here:
<https://doi.org/10.7910/DVN/LXPHBC> —

Navigate the protocol

- Proceed to [Part 5](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Part 5 – Instrumentation and Materials Guidelines

5.1 Precautions

5.1.1. Carefully handle backgrounds and white references (e.g., Spectralon®):

- Use laboratory gloves to handle black backgrounds and white references and tweezers to handle herbarium tissues.
- Avoid touching and dirtying the reflective surfaces.
- Avoid blowing or cleaning backgrounds and reference standards with canned air, which can cause chemical contamination (other options, such as O2 Hurricane, avoid this contamination).

5.1.2. Avoid burning target tissues:

- High light intensities in optical probes can burn tissues, especially thin leaves! Use low light intensity and minimum measurement duration based on manufacturer's guidance to avoid this.
- Avoid measuring tissues with glue on them, whenever possible.
- Place a black background under the leaf or target tissue to avoid spectral contamination from the herbarium mounting paper whenever possible. Background surfaces, except for non-reflective black, will influence reflectance spectra (see Fig. 1.1).

5.2 Spectra Data Files

5.2.1. Output unprocessed percent reflectance data to files, if possible.

- Percent reflectance should be present in the file at the wavelengths directly measured by the instrument.
- If instruments and software output transformed data by default (e.g., Spectral Evolution PSR+ output is interpolated at 1 nm resolution), refer to user manuals to change settings to output percent reflectance values at measured wavelengths instead.

5.2.2. File formats differ among spectroradiometer manufacturers, models, and softwares.

- Make sure that files include percent reflectance values.

5.3 Instrumentation and Materials Quality Control

5.3.1. Instrument maintenance:

- Follow manufacturer SOP for spectroradiometer maintenance. If possible, send the instrument to the manufacturer at regular intervals for maintenance (e.g., every two years).
- Have fiber optic cables checked by the manufacturer annually and replace damaged fiber optic cables. You can test the integrity of the cables by shining a flashlight through one end of the fiberoptic and observing the light transmitted out the other side.

5.3.2. For background and white reference measurement requirements, see [Section 1.2.2](#).

- Background materials are not spectrally neutral and exhibit distinct reflectance profiles that can introduce systematic, wavelength-dependent biases into leaf spectral measurements ([Fig. 5.1](#)). These effects are especially pronounced in the near-infrared and shortwave infrared regions, underscoring the importance of consistent background selection and documentation for reproducible and comparable spectral data.
- White reference scans are typically not saved as a separate file, but the white reference digital numbers (raw sensor values) should be included in each target spectrum file (e.g., `.sig` format). These values are used by the instrument software to calculate reflectance from raw radiance.
- Black background materials are required to have <4% reflectance across all wavelengths (see Fig. 1.1; e.g., IR Flock Sheet from [Musou Black USA](#)).
- If measuring leaves on herbarium paper on a benchtop, place a black mat under the herbarium sheets to prevent contaminating reflectance from the benchtop (see Fig. 4.2). Black felt (3–5 mm thick) is a practical and cost-effective choice for this purpose.

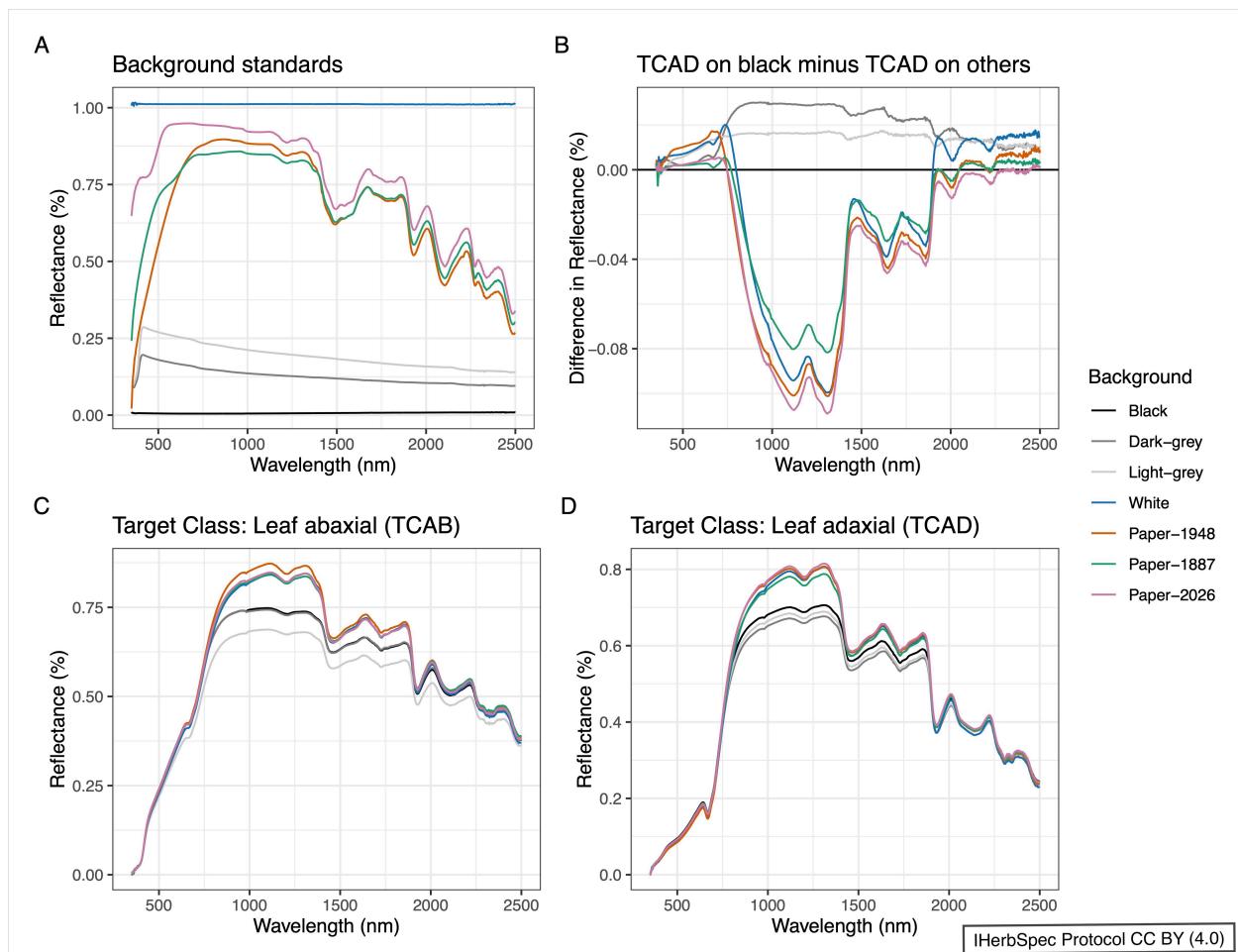


Fig. 5.1. Reflectance of different background materials and their effects on target tissue spectra. **(A)** Reflectance spectra of common background materials used in herbarium spectral measurements, including black (Musou IR Flock), dark grey, light grey (grey matte paint on canvas), white (Spectralon 99%), and three herbarium mounting paper types of different ages. Each background exhibits a distinct spectral profile across the visible–shortwave infrared range (400–2500 nm). **(B)** Differences between adaxial leaf spectra measured on black background (TCAD_BGB) with spectra of the same leaf surface measured on the other backgrounds (see legend). Deviations from zero indicate wavelength-dependent background-induced biases, which are greatest in the near-infrared (NIR; c. 700–1100 nm) and shortwave infrared I regions (SWIR 1; 1100–2000 nm). **(C)** Reflectance spectra of leaf abaxial surface (TCAB) measured on different backgrounds, illustrating how background reflectance can influence absolute reflectance values across spectral regions. **(D)** Mean reflectance spectra of leaf adaxial surface (TCAD) measured on different backgrounds, showing similar background-dependent effects. All measurements were made with a Spectral Evolution NaturaSpec spectroradiometer using a handheld herbarium leaf probe.

5.3.3. Target measurements of background material (black or paper) must be linked to tissue measurements using filename conventions ([Part 3](#)).

- This supports quality control of background conditions and enables development of spectral unmixing approaches to isolate tissue reflectance.

5.3.4. Wear gloves and carefully handle all tissues, white reference, black backgrounds, and calibrated reflectance standards to keep them clean (see 5.1 Precautions).

5.3.5. Calibrated reflectance standards:

- Institutions should consider purchasing calibrated reflectance standards, such as Spectralon® Calibrated (SL Standard) Diffuse Reflectance Standards. These should be kept clean and separate from ‘working’ white references, and used for long-term assessment of instrument performance, optical drift, and the condition of reference materials (e.g., white references and black backgrounds).
- These are supplied with a calibration certificate and serial number that verifies reflectance values across wavelengths.
- Reflectance standards are white (e.g., Spectralon® 99%) and black (e.g., Spectralon® 2%) at a minimum, and can also include gradations of gray.
- It is recommended that standards be measured at regular intervals depending on the level of activity (e.g., once per session; consult spectroradiometer manufacturer), and archived with the filename conventions provided in [Table 3.1](#).

5.4 Specimen and Tissue Annotation

5.4.1. Metadata recommendation:

- IHerbSpec recommends recording the `tissueLocation` metadata field (see [Table 4.3](#)) to document the position of measured tissues on the specimen.

5.4.2. Consult with herbarium collections managers to define a suitable specimen annotation protocol that aligns with institutional policies.

5.4.3. Recommended specimen annotation practices:

- Attach a project annotation label to the herbarium sheet (e.g., “Leaf spectral reflectance; DM White (2025)”).
- Use archival-quality materials for labels and annotations, such as acid-free paper.
- Use pencil for direct annotation of measured tissues on sheets.
- For loose tissues (from packets or envelopes):
 1. Place the measured tissue in a glassine envelope, which is archival and thin.
 2. Include a label inside the envelope indicating the `targetTissueClass` and `targetTissueId` (see [Table 4.3](#); e.g., “Tissue TN1”).
 3. Store the envelope in the specimen packet or attach it to the herbarium sheet (see Fig. 4.2).

Navigate the protocol

- Proceed to [Part 6](#)

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Part 6 – Selecting Tissues for Spectral Measurement

6.1 Sources of Variation in Herbarium Plant Tissues

Spectral data from herbarium specimens are influenced by two major sources of variation.

The first is the natural **biological variation** of phenotypes constructed through genotype-by-environment interactions that biologists seek to understand. This includes differences in developmental stage (e.g., young, mature, senescent), within-individual morphological variation (e.g., sun vs. shade leaves), and biotic influences such as herbivory or pathogens.

The second source is **herborization-related variation**—herborization being defined as the process of preserving plant tissues in an herbarium, from the time of collection through long-term storage. This includes collection, pressing, drying, application of preservatives, and subsequent storage conditions. Each of these steps will affect the spectral data to some extent.

For example, differences in technique (e.g., bagged vs. immediately pressed specimens; use of ethanol) and environmental factors during drying and storage can lead to discoloration, dehydration, tissue distortion, and contamination from biotic or abiotic agents such as fungus, glue, or insecticides. Among these, the drying protocol might be particularly influential: specimens not dried efficiently—especially in warm, humid conditions—are prone to rot and degradation, which significantly affects spectral quality. Over time, temperature and humidity fluctuations may accelerate tissue degradation.

As a result, herbarium specimens exhibit a wide range of tissue conditions, reflecting both the biology of the living organism and the varied preservation environments they experience. **Documenting these sources of variation as metadata is essential**, as they can be correlated with model accuracy (Kühn et al. 2024, White et al. 2025) and can be leveraged to improve the performance of predictive models.

When available, contextual details from specimen labels—such as the use of ethanol or other treatments—should be captured in metadata. As it is rare to see such details, botanists should be encouraged to include preservation protocols on specimen labels.

The metadata fields `hasGlue`, `hasNonGlueContamination`, `measurementFlags`, and `tissueNotes` ([Table 4.3](#)), are designed to record the condition of the tissue within the target measurement area at the time of measurement. These fields are essential for downstream filtering and analysis of spectra.

Examples of specimen tissue metadata records are provided in [Appendix II](#).

General Guidelines for Tissue Selection

Note: Much of this text is tailored to measuring leaf tissues, but can generally be applied to other tissues.

1. Quality over quantity.

- Under the goal of broad digitization of herbarium tissues, spectral data quality could be considered more important than specimen quantity. Targeting higher quality, less degraded specimens will yield more informative spectral data.
- If the majority of tissues on a specimen are degraded, damaged, or contaminated, that specimen should generally not be measured unless it holds particular research value.

2. Select specimens based on their general tissue condition:

- For digitization projects focused on broad taxonomic coverage, specimens should be evaluated as whole units when determining suitability for spectral measurement. Next, technicians should prioritize tissues that are both suitable for measurement and representative of the overall condition of the specimen, rather than selecting isolated tissues that are in unusually good condition. This approach helps ensure alignment between the general specimen condition and the resulting spectral data, and could avoid biases based on assumptions of exceptional tissue conditions. If the majority of tissues on a specimen are degraded, damaged, or contaminated, that specimen should generally not be measured unless it holds particular research value. Conversely, if the specimen consists mostly of clean, mature, fully expanded tissues, it is appropriate to proceed with measurement—while also considering additional measurements, when feasible, to capture variation such as developmental stage or damage from herbivory and scoring the appropriate metadata for these additional measurements.
- *Caveat:* while the overall condition of the specimen informs whether it should be measured, metadata on developmental stage, contamination, and tissue condition should always record the characteristics of the specific measurement area of the tissue—not the specimen as a whole. This ensures metadata accuracy and supports high-resolution filtering and analysis. Information about the specimen can be recorded in the `tissueNotes` field ([Table 4.3](#)).

3. Score metadata with care.

- Metadata collection is a critical component for understanding factors affecting spectral data quality and will provide confidence during data aggregation.
- Herbaria and research teams should train technicians in specimen selection strategies and scoring metadata to reduce subjectivity.

4. Prioritize:

- Tissue measurements on black backgrounds.
 1. Detached tissues of appropriate size in fragment packets allow full assessment of the presence of glue and access to both leaf surfaces.
 2. For mounted specimens, taped or sewn tissues can have black backing inserted beneath (technicians should look for evidence of old glue).
- Specimens with mature tissues (e.g., fully expanded leaves) that are clean of biotic or abiotic contaminants.

- Tissues that fill the probe measurement area, even if midvein is measured.
- Leaves with flat, complete surfaces and intact laminae.

5. Avoid:

- Tissues with glue.
- Tissues attached to herbarium paper such that no black background can be inserted beneath the tissue.
- Tissues that are not pressed flat (e.g., deformed, folded, crumpled, or abnormal leaves).
- Tissues contaminated by biotic or abiotic agents (e.g., algae, lichen, fungus, chemical preservatives, glue).
- Damaged tissues (e.g., by herbivory, pathogens, herborization phenomena).
- Midribs and major venation in the leaves, or other types of vasculature or intrusions in target tissue.
- If larger tissues are available, avoid smaller tissues that don't cover the entire optical probe measurement area.

6. Measuring small tissues:

Measuring taxa with naturally small tissues presents unique challenges, but several strategies can help obtain representative spectral measurements with minimal error.

- Center the probe on multiple areas of the small tissue and collect several measurements, then average the values to represent the tissue's spectral profile.
- Create a mosaic by arranging multiple small tissue units side by side to fill the measurement area. Avoid overlapping tissues, as stacking can alter light scattering and distort reflectance values (Fig. A1).
- Consider optical setup with small measurement area: For projects in the instrument selection phase, Spectral Evolution offers a custom optical probe with a 2-mm measurement area compatible with their NaturaSpec and PSR spectroradiometers.
- Avoid using narrow angle lenses as these might distort the spectrum (Fig. A1).

7. Multiple individual plants within one herbarium sheet:

- Each individual plant on a sheet should be demarcated with a unique `specimenId` ([Table 4.2](#)).

Decision tree for tissue selection

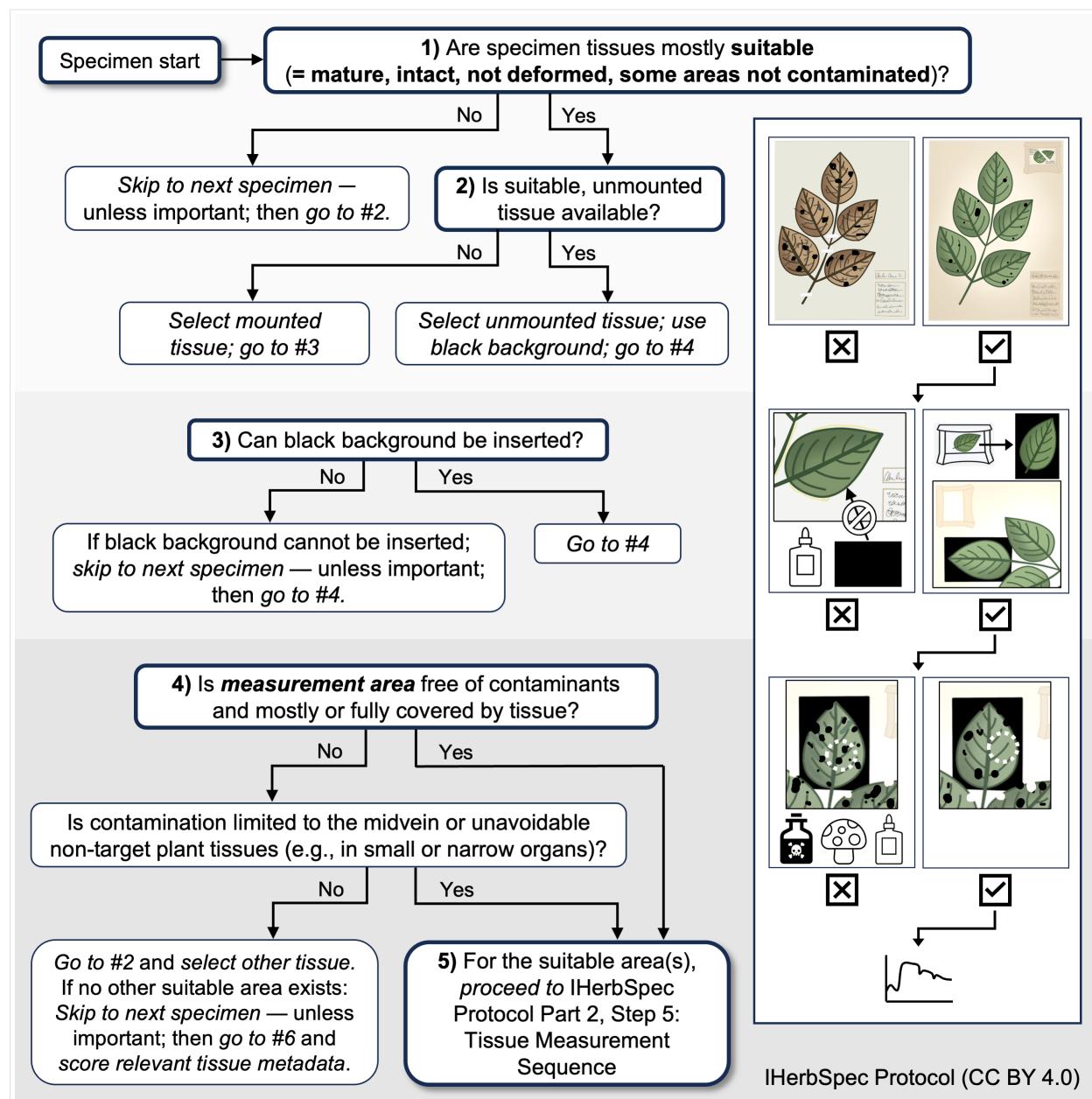


Fig. 6.1. Decision tree diagram for target tissue selection.

Navigate the protocol

Browse additional parts and appendices:

- [**Part 1**](#) – Protocol Overview
- [**Part 2**](#) – Measurement and Metadata Workflow
- [**Part 3**](#) – Filename Conventions and Formats
- [**Part 4**](#) – Metadata and Databasing
- [**Part 5**](#) – Instrumentation and Materials Guidelines
- [**Part 6**](#) – Selecting Tissues for Spectral Measurement
- [**References**](#)
- [**Appendix I**](#) – Number of Measurements per Tissue
- [**Appendix II**](#) – Tissue Metadata for Quality Control

References

- Cavender-Bares J, White DM, Ahlstrand NI, Austin M, Bastianelli D, Bazan S, Boughalmi K, Cardinal-McTeague W, Chacón-Madrigal E, Couvreur TLP, Davis C, Durgante FM, Grace OM, Q JAG, Hansen K, Hernández-Leal MS, Hopkins MJG, Jackson R, Kothari S, Lee AK, Léveillé-Bourret É, Pinto-Ledezma J, Casaverde NLQ, Meireles JE, Neto-Bradley B, Nichodemus CO, Ree RH, Schmull M, Soltis DE, Soltis PS, Tuomisto H, Ustin S, Vasconcelos CC. 2025. **Next-generation specimen digitization: capturing reflectance spectra from the world's herbaria for modeling plant biology across time, space, and taxa** *New Phytologist*. <https://doi.org/10.1111/nph.70645>.
- Durgante FM, Higuchi N, Almeida A, Vicentini A. 2013. **Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy**. *Forest Ecology and Management* 291:240–248. <https://doi.org/10.1016/j.foreco.2012.10.045>.
- Hadlich HL, Durgante FM, dos Santos J, Higuchi N, Chambers JQ, Vicentini A. 2018. **Recognizing Amazonian tree species in the field using bark tissues spectra**. *Forest Ecology and Management* 427:296–304. <https://doi.org/10.1016/j.foreco.2018.06.002>.
- Kühn P, Proß T, Römermann C, Wesche K, Brügelheide H. 2024. **Using near-infrared spectroscopy to predict nitrogen and phosphorus concentrations of herbarium specimens under different storage conditions**. *Plant Methods* 20:19. <https://doi.org/10.1186/s13007-024-01146-x>.
- Mersni J, Bazan S, Baby E, Bonnal V, Bastianelli D. 2025. **Protocol for NIR spectra acquisition on herbarium specimen**. *Agritrop*. <https://doi.org/10.18167/agritrop/00880>.
- Neto-Bradley BM, Bonnet P, Goëau H, Joly A, Cavender-Bares J, Coomes DA. 2025. **Using reflectance spectra and PI@ntNet to identify herbarium specimens: a case study with *Lithocarpus***. *New Phytologist*. <https://doi.org/10.1111/nph.70258>.
- White DM, Cavender-Bares J, Davis CC, Guzmán Q JA, Kothari S, Robles JM, Meireles JE. 2025. **Seeing herbaria in a new light: leaf reflectance spectroscopy unlocks trait and classification modeling in plant biodiversity collections**. *New Phytologist*. <https://doi.org/10.1111/nph.70357>.

Navigate the protocol

Browse additional parts and appendices:

- **Part 1** – Protocol Overview
- **Part 2** – Measurement and Metadata Workflow
- **Part 3** – Filename Conventions and Formats
- **Part 4** – Metadata and Databasing

- **Part 5** – Instrumentation and Materials Guidelines
- **Part 6** – Selecting Tissues for Spectral Measurement
- **References**
- **Appendix I** – Number of Measurements per Tissue
- **Appendix II** – Tissue Metadata for Quality Control

Appendix I – Considerations Regarding the Number of Measurements per Specimen

Considerations

Spectral measurements of plant tissues are highly sensitive to both instrument configuration and small spatial differences across target tissue. Changes in the optical probe geometry in relation to the tissue—such as tilting, poor contact with the leaf surface, partial coverage of the measurement area (Fig. A1), or micro-contaminants like debris—can introduce technical errors. In addition, leaf surface microtopography, anatomical variation, and alterations from herborization (e.g., drying distortion, discoloration, or degradation) all contribute to variation in reflectance spectra (Cavender-Bares et al. 2025).

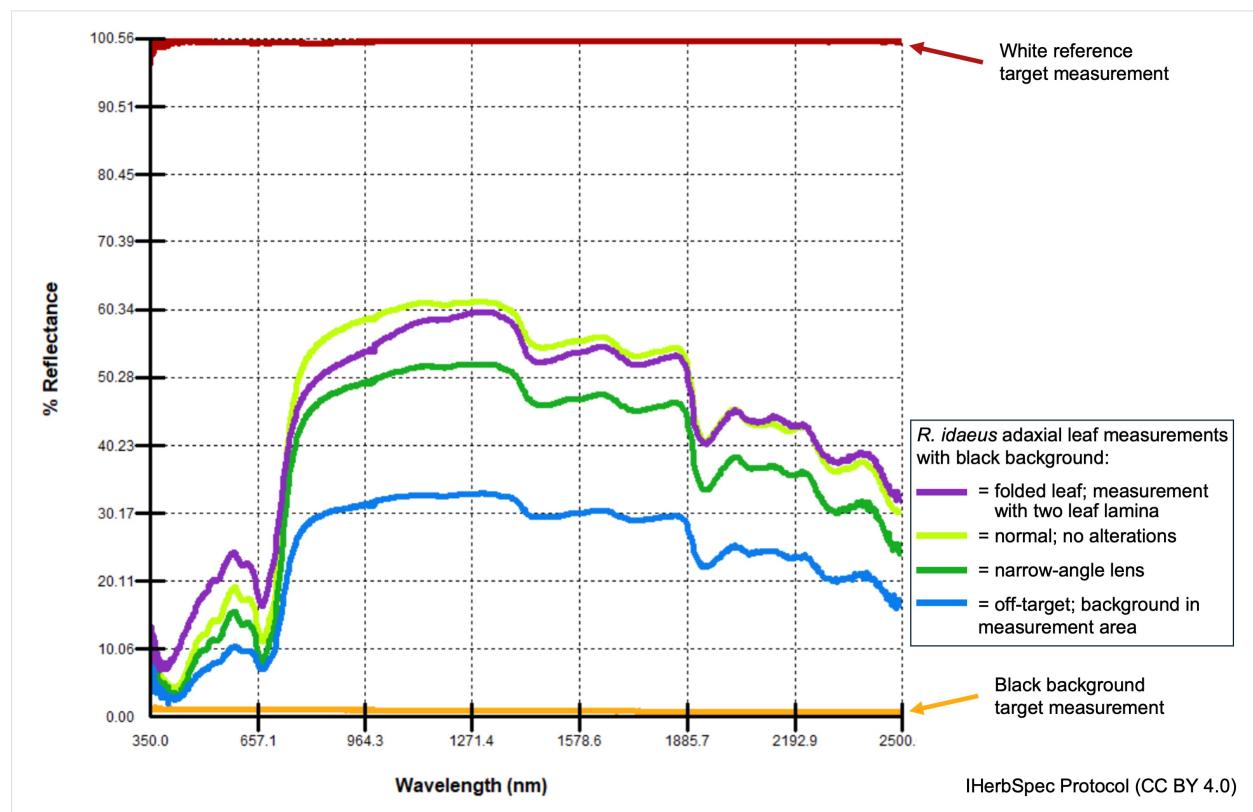


Fig. A1: Variability in spectral measurements due to instrument configuration and technical error. The following spectral measurements are shown: white calibration standard (red), black background (orange), standard adaxial measurement with black background (light green), standard adaxial measurement but with narrow-angle lens following recalibration (dark green), adaxial measurement where the tissue did not fully cover the optical field of view (i.e., `backgroundInMeasurement = TRUE`; blue), and a folded leaf (purple). All measurements were made of the same tissue using an SVC HR-1024i spectroradiometer.

These considerations support the IHerbSpec Protocol requirement of collecting a minimum of three representative measurements per adaxial and abaxial leaf surface, when suitable tissues for both surfaces are available. Three measurements are necessary to calculate a mean spectrum and its variance. The protocol further recommends collecting five measurements per surface to achieve a more robust spectral characterization and to meet the threshold for improved performance in classification models. The protocol allows for users to make these minimum or higher measurements across one or many leaves. For other tissue types, at least three measurements are recommended, with more encouraged when feasible. Below, we elaborate justification for this recommendation.

Several studies have demonstrated improved model performance when multiple measurements per individual specimen are averaged, including those by Durgante et al. (2013; Figs. A2, A3), Neto-Bradley et al. (2025), and S. Bazan et al. (Unpubl.). The IHerbSpec Protocol's recommendation of five measurements is based on empirical evidence that model accuracy gains tend to stabilize beyond five to ten measurements per tissue (Fig. A2), although the optimal number may vary depending on tissue type, instrument sensitivity, and modeling goals. Beyond classification, IHerbSpec members are investigating how replicate measurements affect the accuracy and robustness of trait prediction models, evaluating both the benefits of averaging spectra and the variation in predicted trait values from unaveraged measurements of the same tissue.

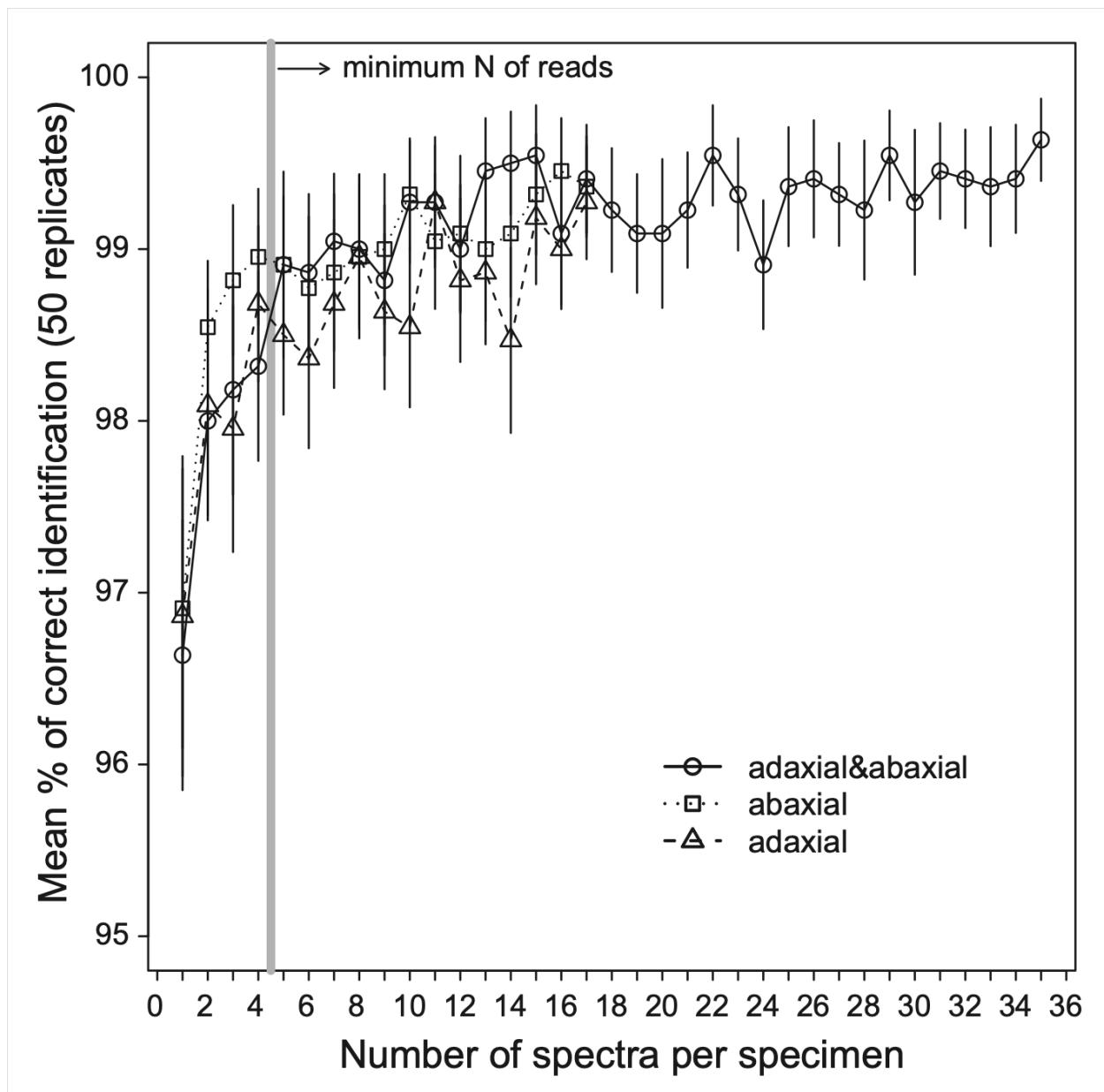


Fig. A2: Accuracy of taxonomic discrimination as a function of the number of averaged measurements per specimen. Three tissue class datasets are analyzed: combined adaxial and abaxial spectra (circles), abaxial only (squares), and adaxial only (triangles). Each point represents the average classification accuracy from 50 replicate linear discriminant analysis models, using randomized subsampling and test set validation. Error bars indicate 95% confidence intervals. The vertical gray line marks the minimum number of spectral measurements ($n = 5$), below which accuracy significantly decreases (Tukey test, $P < 0.05$). Data were collected using an FT-NIR spectrometer (1000–2500 nm) from 10 species in Lecythidaceae, including eight *Eschweilera* and two *Corythophora* species. Accuracy increases notably between one and five spectra per specimen, after which performance stabilizes, indicating diminishing returns beyond five measurements. This figure is reproduced from Durgante et al. (2013) with permission from the publisher.

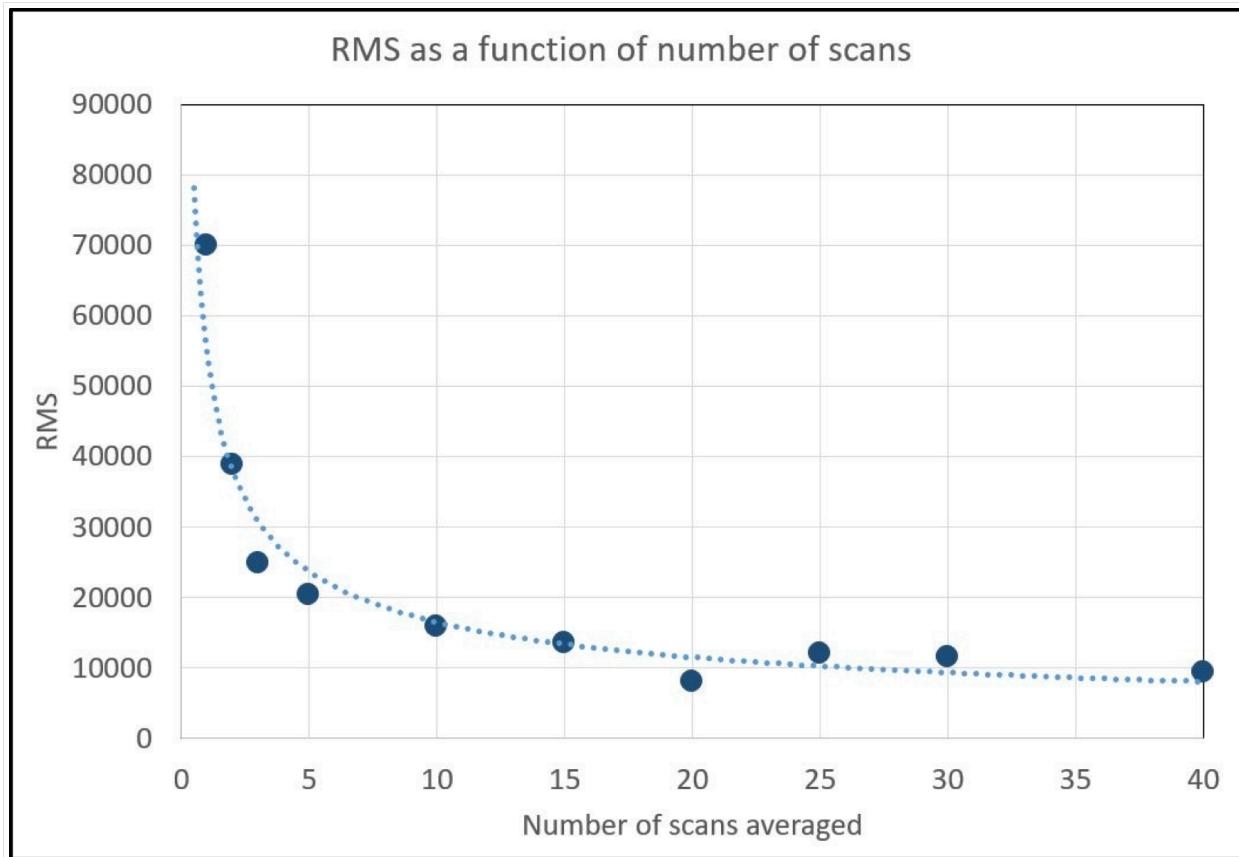


Fig. A3: Spectral variability as a function of the number of measurements averaged per measurement. Here, a total of 100 spectral measurements were collected from a single *Annona* sp. leaf using an ASD LabSpec Pro spectrometer over a black background. Subsets of these measurements were then averaged in groups ranging from 2 to 40 to simulate different measurement strategies. The plot shows the root mean square (RMS) distance between replicate spectra as a function of the number of measurements averaged, illustrating that spectral variability decreases substantially as more measurements are combined. Notably, averaging fewer than five measurements did not adequately capture intra-tissue spectral variation. This analysis supported the required minimum of 10 measurements per specimen for the formal protocol currently implemented by teams at CIRAD and IRD in France (Mersni et al. 2025).

Another important recommendation in the IHerbSpec Protocol is to measure both adaxial and abaxial leaf surfaces. These surfaces differ in anatomy and function due to evolutionary adaptations for light capture, gas exchange, and environmental stress. Such structural differences are consistently reflected in their spectral signatures (see Fig. 1.1), and we recommend measuring both surfaces—when suitable tissue is accessible—to capture this biologically meaningful variation.

Within the context of a large-scale macroevolutionary project focusing on Annonaceae (ERC GLOBAL project; PI Thomas Couvreur), researchers aimed to measure spectra on one specimen per species with the intent to capture as much variation as possible for each specimen. In accordance with the protocol established at the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), researchers decided to make 20 spectral measurements per specimen to develop their spectral library for 1700 Annonaceae species preserved in over

30 different herbaria.

This decision was informed by preliminary experiments demonstrating that spectral stability improves with larger numbers of measurements (Fig. A3). The choice of 20 spectra was also appropriate because of variation in the number and size of leaves, allowing for multiple measurements of multiple leaves or single large leaves that filled an entire herbarium sheet. It also allowed for poor-quality spectra to be discarded based on downstream filtering criteria while still retaining a large number of measurements per specimen.

Overall, given the sensitivity of spectral data, research projects are encouraged—when feasible—to collect more than five measurements per tissue, including across different tissue units within an individual. This enables further assessment of intra-individual spectral heterogeneity and its impact on downstream analyses.

The marginal cost of additional measurements is relatively low across instruments, with a single measurement typically completed within 1 to 10 seconds. Nevertheless, determining the appropriate number of measurements remains a critical design consideration, particularly in large-scale digitization projects involving significant resource investment.

Navigate the protocol

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control

Appendix II – Representing Tissue Conditions in Metadata for Quality Control

This appendix provides a brief discussion and examples of the IHerbSpec approach for recording metadata pertaining to sources of biological and herborization variation that affect spectral data.

Scoring metadata for tissue conditions

As explained in [Section 6.1 Sources of Variation](#), we aim to capture biological variation that is scientifically meaningful, but the degraded nature of herbarium specimens requires a careful assessment and annotation of tissue quality and sources of contamination. The Tissue Metadata fields of, `developmentalStage`, `hasGlue`, `hasNonGlueContamination`, `measurementFlags`, and `tissueNotes` [Table 4.3](#) are designed to support quality control of spectral data and enable downstream filtering.

The codes for enumerating the `developmentalStage` fields are provided in [Table 4.4](#), and the codes for `measurementFlags` are described in [Table 4.6](#).

Technicians should also be aware that some taxa (e.g. Myrtaceae) or tropical collections in general may naturally exhibit discoloration or non-flat, deformed surfaces (e.g., *Gasteranthus atratus*). Such characteristics do not necessarily indicate tissue degradation and poor-quality spectral data. In addition, some taxa exhibit more pronounced discoloration during herborization as a result of their biology, and such discoloration should not be interpreted as evidence of poor preservation condition. Some traits also covary; for example, young leaves wilt and become discolored faster and are more challenging to press and dry flat, so they may more often be scored as moderately or poorly preserved.

As a reminder, **measurementFlags pertain only to the measurement area**, not the whole tissue unit (e.g., leaf). Since suitable measurement areas of tissues (see decision tree in [Fig. 6.1](#)) will not contain any contaminants, pathogens, or other damage, technicians should not be enumerating any codes in the `measurementFlags` field. This could differ in projects collecting measurements pertaining to extended uses of the specimen, e.g., symbionts, degradation, etc., that might be measuring tissue areas with these other features.

The specimens and metadata examples in [Table A1](#) have been selected to guide consistent scoring of these fields during spectral measurement and promote metadata consistency across projects. Note that the images do not specify the exact measurement area corresponding to these metadata fields (as is done in the IHerbSpec Metadata Spreadsheet for specimen NEBC_00651639 shown in main text [Fig. 2.2](#)).

Table A1: Examples for Scoring of Tissue Conditions.

This table provides example metadata records for herbarium specimens, illustrating how to score select Tissue Metadata [Table 4.3](#) related to condition and contamination. The fields

`developmentalStage`, `hasGlue`, `hasNonGlueContamination`, and `measurementFlags` refer specifically to the measurement area. Broader observations about the specimen or other tissues can be recorded in `tissueNotes` (see [Table 4.3](#)). Linked specimen images are included, though they do not indicate the precise measurement area corresponding to each metadata record.

This table can be downloaded on the iherbspec.github.io/protocol.

herbariumCode	specimenID	developmentalStage	hasGlue	hasNonGlueContamination	measurementFlags	tissueNotes
A_00631088		mature	TRUE	FALSE	MediumPreservation	Discolored
A_00672503		mature	FALSE	TRUE	GoodPreservation	PathogenPresent
A_00746288		mature	TRUE	FALSE	MediumPreservation	Discolored. Glue on loose leaf in measurement area.
A_00772865		young	FALSE	FALSE	PoorPreservation	Measured leaves uneven, mottled, discolored, breakage
A_2613405		mature	FALSE	FALSE	MediumPreservation	
A_2614928		mature	uncertain	TRUE	PoorPreservation	OrganismPresent
AAU_A.S.Barfod48062		mature	FALSE	TRUE	MediumPreservation	MidveinPresent
AAU_I.A.Chacon1108		mature	FALSE	TRUE	MediumPreservation	AlcoholPresent
AAU_R.C.Moran6003		mature	FALSE	TRUE	GoodPreservation	AlcoholPresent
ALF_035514		mature, young	FALSE	FALSE	GoodPreservation	
ECON_00338371		young	FALSE	FALSE	MediumPreservation	Discolored, wrinkled. Specimen burnt but not in target area.
GH_00611866		mature	FALSE	FALSE	MediumPreservation	Discolored specimen
INPA_142198		mature	FALSE	FALSE	MediumPreservation	Discolored, herbivory on specimen
INPA_179322		mature	FALSE	FALSE	GoodPreservation	
INPA_184434		mature	FALSE	FALSE	MediumPreservation	Herbivory on specimen, necrotic leaf
INPA_187932		mature	FALSE	FALSE	MediumPreservation	Burnt leaves on specimen
INPA_203125		mature	FALSE	FALSE	MediumPreservation	Discolored, wrinkled leaves
INPA_218399		mature	TRUE	FALSE	PoorPreservation	Discolored, herbivory, wrinkled and burnt leaves
INPA_264413		mature	FALSE	FALSE	GoodPreservation	

herbariumCode	specimenID	developmentalStage	hasGlue	hasNonGlue	measurementFlags	tissueNotes
INPA_267597		mature	FALSE	FALSE	GoodPreservation	
INPA_275373		mature	FALSE	FALSE	MediumPreservation	Discolored, mold, grayish spots on older leaves specimen
INPA_278882		mature	FALSE	TRUE	MediumPreservation	AlcoholPresent
INPA_4628		mature	FALSE	TRUE	PoorPreservation	PreservativePresent
INPA_59178		mature	TRUE	TRUE	PoorPreservation	PathogenPresent
INPA_72939		mature	TRUE	TRUE	PoorPreservation	FungusPresent
INPA_97794		mature	FALSE	FALSE	MediumPreservation	Wrinkled leaves
INPA264208		mature	FALSE	FALSE	MediumPreservation	Measurement area not flat. Herbivory on sheet
M_PU681750		mature	FALSE	uncertain	MediumPreservation	Discolored; Alcohol preservation suspected.
MIN_332436		mature	FALSE	FALSE	MediumPreservation	MidveinPresent
MIN_370477		mature	FALSE	FALSE	PoorPreservation	Discolored, wrinkled leaves
MIN_588200		mature	FALSE	FALSE	GoodPreservation	
NEBC_00634726		young	FALSE	FALSE	PoorPreservation	Measurement area leaves uneven, discolored
NEBC_00636882		uncertain	FALSE	TRUE	PoorPreservation	PathogenPresent
NEBC_00651639		mature	FALSE	FALSE	GoodPreservation	
NEBC_00677096		mature	FALSE	TRUE	MediumPreservation	PathogenPresent
NEBC_00695035		mature	FALSE	FALSE	GoodPreservation	
NEBC_00746092		mature	FALSE	TRUE	MediumPreservation	PathogenPresent
NEBC_00898559		mature	FALSE	FALSE	GoodPreservation	
NEBC_02618198		mature	FALSE	FALSE	MediumPreservation	Dark spots on measured leaf
NEBC_02618743		mature	FALSE	FALSE	GoodPreservation	
NY_00042623		mature	TRUE	FALSE	GoodPreservation	
NY_00043385		mature	FALSE	TRUE	MediumPreservation	PathogenPresent
NY_00194824		mature	FALSE	FALSE	GoodPreservation	
NY_00402473		mature	TRUE	FALSE	MediumPreservation	MoldPresent
NY_00709210		mature	FALSE	TRUE	MediumPreservation	PathogenPresent
NY_01163859		mature	TRUE	FALSE	MediumPreservation	Discolored; herbivory on specimen
NY_02498899		mature	TRUE	FALSE	MediumPreservation	Discolored
NY_02499746		mature	FALSE	TRUE	PoorPreservation	MoldPresent

herbariumCode	specimenID	developmentalStage	hasGlue	hasNonGlue	measurementFlags	tissueNotes
P_00391795		mature	FALSE	uncertain	MediumPreservation	wrinkled leaves; possible chemical preservative.
P_05198525		mature	FALSE	TRUE	PoorPreservation	FungusPresent
P_05198660		mature	FALSE	TRUE	MediumPreservation	AlcoholPresent

Navigate the protocol

Browse additional parts and appendices:

- [Part 1](#) – Protocol Overview
- [Part 2](#) – Measurement and Metadata Workflow
- [Part 3](#) – Filename Conventions and Formats
- [Part 4](#) – Metadata and Databasing
- [Part 5](#) – Instrumentation and Materials Guidelines
- [Part 6](#) – Selecting Tissues for Spectral Measurement
- [References](#)
- [Appendix I](#) – Number of Measurements per Tissue
- [Appendix II](#) – Tissue Metadata for Quality Control