January, 2022

iSamples Metadata Scheme

A core description scheme applicable to material samples from any of the partner systems.

_____

ESIP Winter Meeting

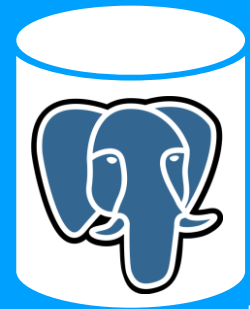Stephen M. Richard, smrTucson@gmail.com

iSamples Architecture

Domain sample registries

OpenContext   GEOME

iSamples in-a-Box (iSB) stack

Ingest process;
APIs from source systems

PostgreSQL database-
Cache raw records

Sitemap-enable harvesting

Transformation to Common schema

Full text search
Geospatial Index

API

API access by other apps

Record lookup
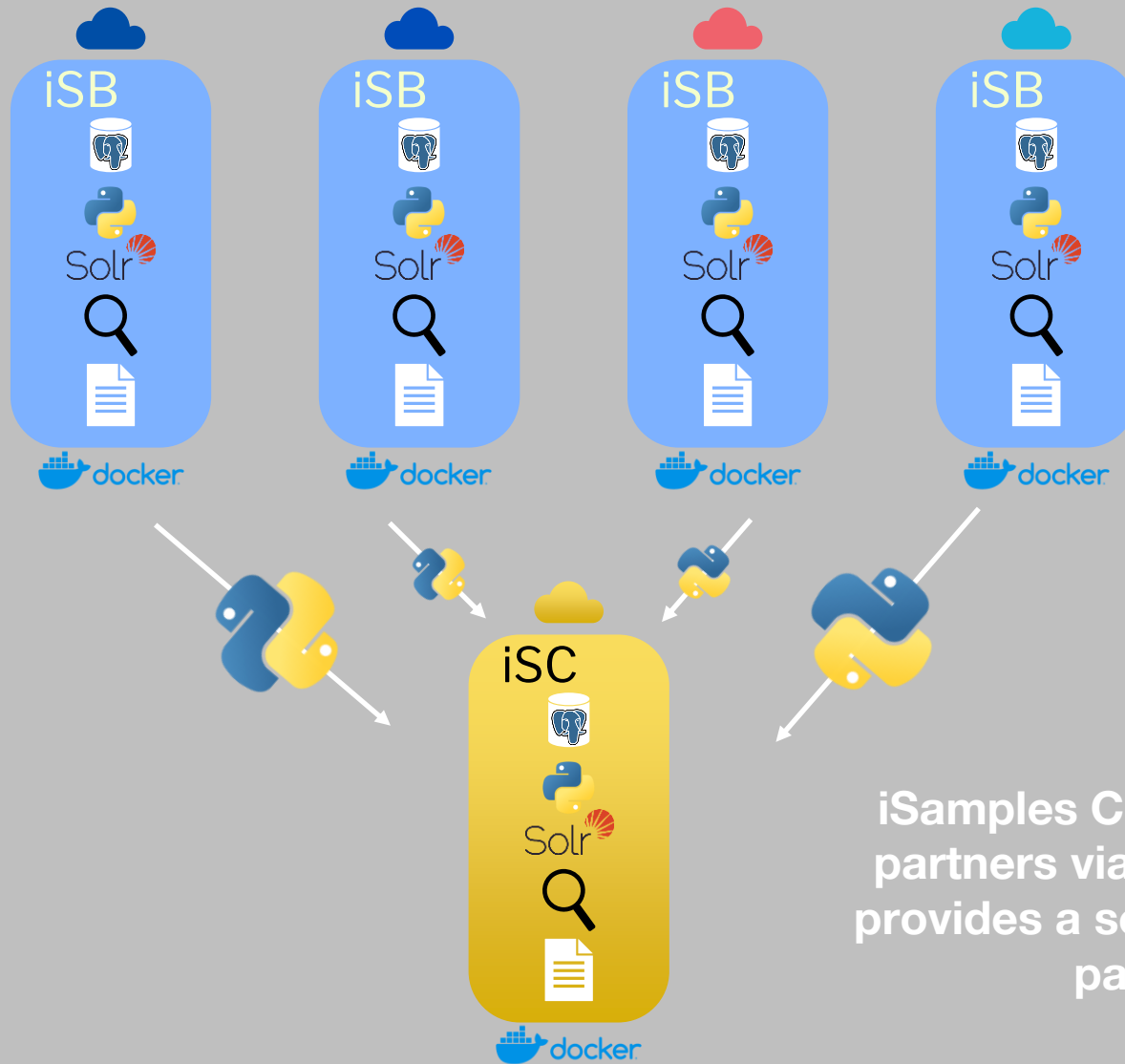
Search UI

Each partner can implement the iSB stack via Docker

Partners can customize any of the components in the iSB stack to meet their requirements

iSamples Central harvests from the partners via a standard API. Central provides a search endpoint across all partner registries

iSB iSB iSB iSB

iSC

3

Metadata development approach

- Study existing sample descriptions from partner systems
    - OpenContext: Samples from archaeology sites
    - GEOME – Genomic samples
    - SESAR --- Earth Science samples
    - Smithsonian – various, mostly biological samples

- Review other related models
    - IGSN   GeoSciML   SESAR
    - ODM2   TDWG   W3C SSN

Identifiers

Thing in information system

Thing in the world

Material Sample

Metadata
Sample Description
('Digital Specimen')

Identifier

Identifier

Identifier for material sample dereferences on the Web to the sample metadata record

The metadata record is the digital representation of the material object, Also known as 'Digital Specimen'. It has a separate identifier from the physical sample.

5

Basic discovery

Sample Description

Core properties for cross-domain sample search

Identifier

A globally-unique, web-resolvable identifier for the sample. Various schemes supported (igsn, doi, ark...

Sample Label and Description (Text)

User-provided label - identifier and description. Critical for full text indexing and discovery

What is it

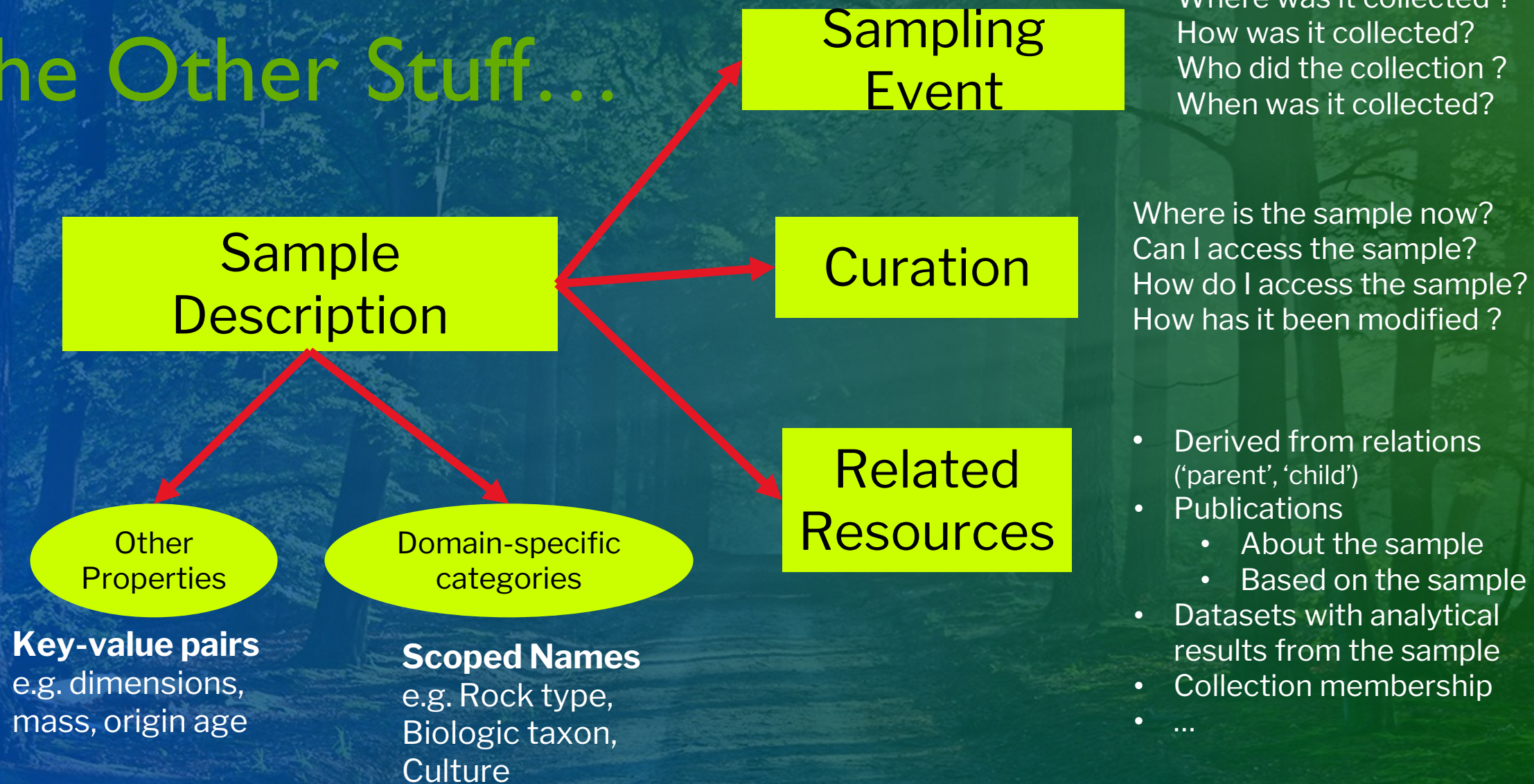**What is it composed of**

What is the sampled feature (Context)

Keywords

Specimen type

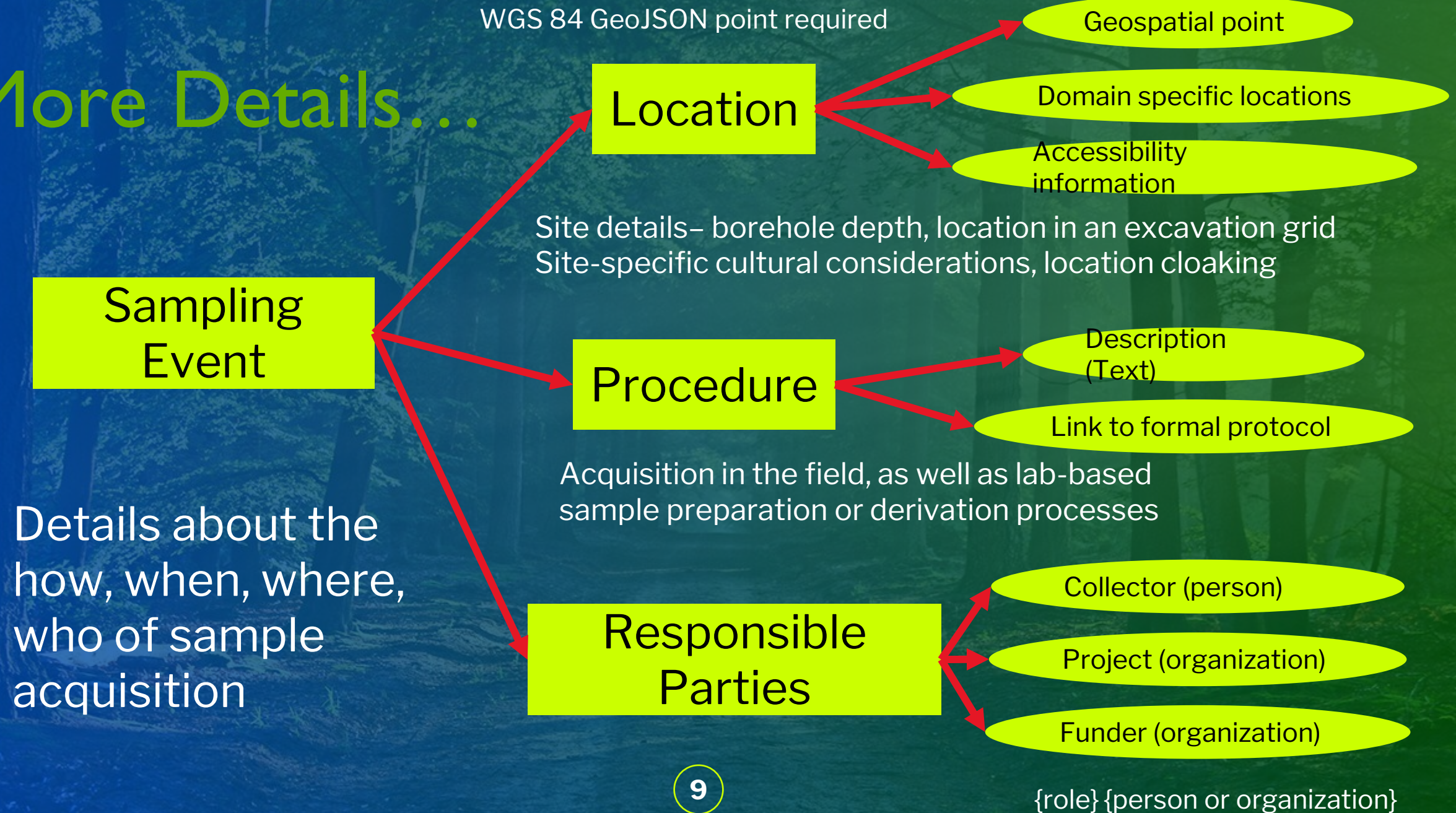Material type

Sampled feature type

Core controlled vocabularies

6

The Other Stuff…

Sample Description

Sampling Event

Where was it collected ?
How was it collected?
Who did the collection ?
When was it collected?

Curation

Where is the sample now?
Can I access the sample?
How do I access the sample?
How has it been modified ?

Related Resources

- Derived from relations ('parent', 'child')
- Publications
  - About the sample
  - Based on the sample
- Datasets with analytical results from the sample
- Collection membership
- …

Other Properties

Key-value pairs
e.g. dimensions, mass, origin age

Domain-specific categories

Scoped Names
e.g. Rock type, Biologic taxon, Culture

7

Links

- Github Repository: https://github.com/isamplesorg/metadata/
  - **Current JSON schema**: iSamplesSchemaBasic0.3.json
  - **Vocabularies**: /vocabulary subdirectory
    - Decision trees:
      - SpecimenTypeDecisionTreev2.pdf
      - MaterialTypeDecisionTreev3.pdf
      - SampledFeatureDecisionTreeV4.pdf
    - Turtle (.ttl) files are skos versions.

- **Card sorting** exercise to test vocabularies
  - Specimen type: https://x12745x3.optimalworkshop.com/optimalsort/3x1kqnv2
  - Material type: https://x12745x3.optimalworkshop.com/optimalsort/ud4300c4-0
  - Sampled feature/context: https://x12745x3.optimalworkshop.com/optimalsort/qf1254qw

More Details…

Curation

Where has the sample been, where is it now, can I get a split, what has happened to the sample…?

Storage location — Where is the sample now?

Responsible parties — Who do I contact about the sample?

Access constraints — Can I see the sample? Handle the sample?
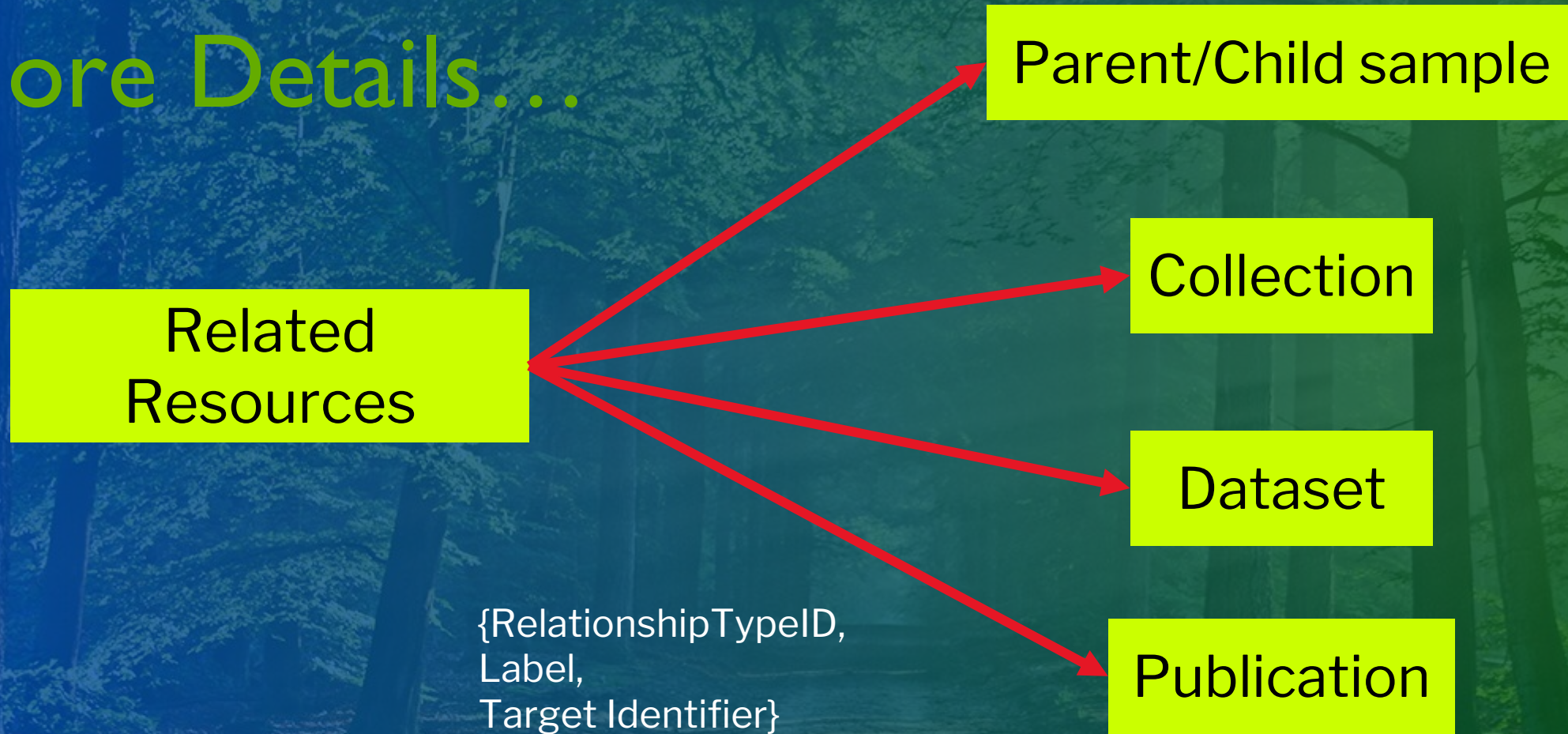
Usage constraints — What kind of experiments/ measurements can be done with the sample?

Modifications — Preservation actions, Destructive analyses, lost splits, contamination

10

More Details…

Related Resources

→ Parent/Child sample

→ Collection

→ Dataset

→ Publication

{RelationshipTypeID,
Label,
Target Identifier}

11

Material Type Vocabulary

- What is the sample composed of – what is the 'substance'

- Intended to be high-level, for cross-domain usage

- Use for faceted search in user interface

12

- Material
  - Any anthropogenic material
    - Anthropogenic material
    - Anthropogenic metal material
  - Any ice
    - Frozen water
  - Biogenic non-organic material
  - Dispersed media
  - Fluid material
    - Gaseous material
    - Liquid water
    - ⇨ Non-aqueous liquid material
  - Natural Solid Material
    - Mineral
    - Mixed soil, sediment or rock
    - Particulate
    - Rock
    - Sediment
    - Soil
  - Organic material

Specimen Type Vocabulary

- What kind of object is the sample? The thing I would see in a drawer at the museum.

- Intended to be high-level, for cross-domain usage

- Use for faceted search in user interface

13

- Physical specimen
  - Any aggregation specimen
    - Aggregation
    - Anthropogenic aggregation
    - Biome aggregation specimen
  - Any biological specimen
    - Biome aggregation specimen
    - Organism part
    - Organism product
    - Whole organism specimen
  - Artifact
  - Fossil
  - Liquid or gas
  - Other solid object
  - Research product
    - Analytical preparation
    - Experiment product

Sampled Feature type Vocabulary

- High level categories for the context— What is the sample supposed to represent?

- Intended to be high-level, for cross-domain usage

- Use for faceted search in user interface

14

- Any sampled feature
  - Anthropogenic environment
    - Active human occupation site
      - Experiment setting
      - Laboratory or curatorial environment
    - Site of past human activities
  - Earth environment
    - Atmosphere
    - Earth interior
    - Earth Surface
      - Lake, river, or stream bed
      - Marine water body bottom
      - Subaerial surface environment
    - Glacier environment
    - Subsurface fluid reservoir
    - Water body
      - Marine environment
      - Terrestrial water body
  - Extraterrestrial environment