Study Design and Data Collection

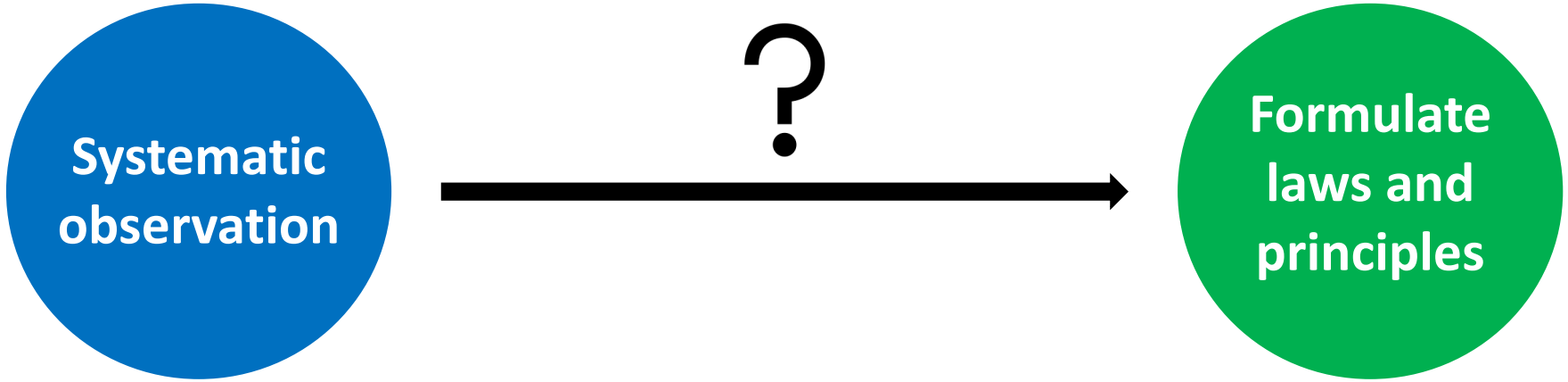
Ecological and Epidemiological Modeling in Madagascar

Centre ValBio, Ranomafana, January 2020

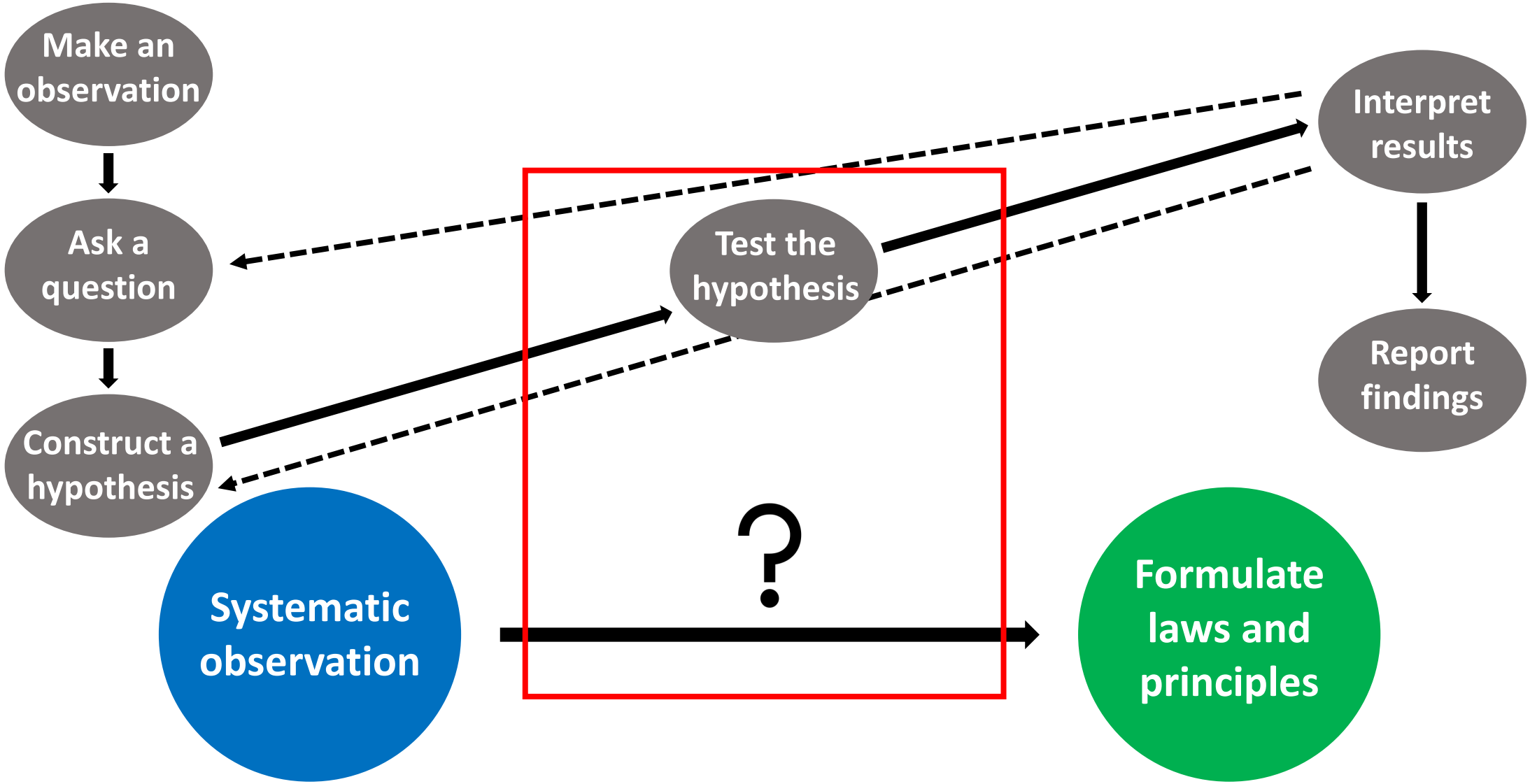
What *is* science?

the **systematic observation** of natural events and conditions in order to **discover facts** about them and to **formulate laws and principles** based on these facts

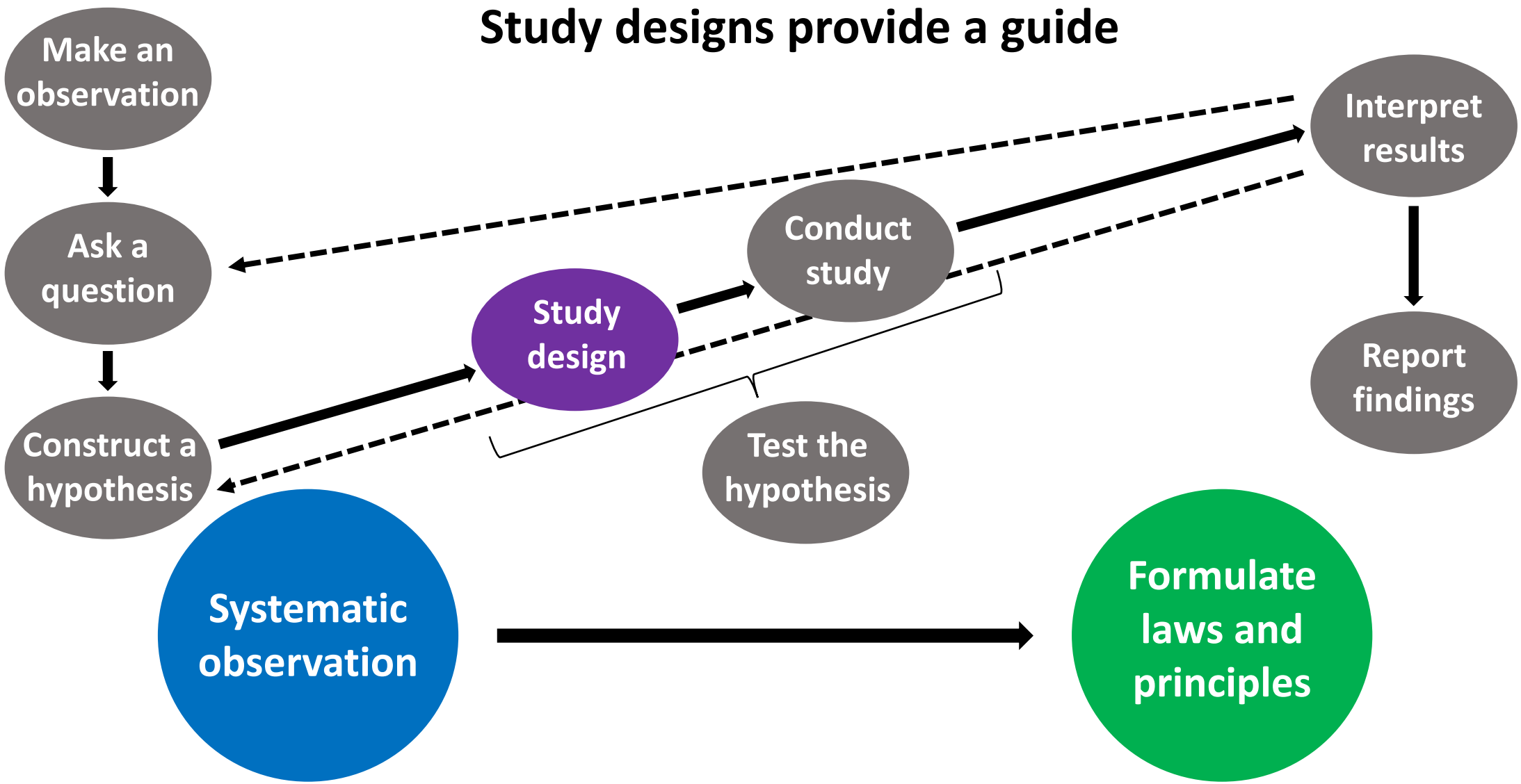
-- Academic Press Dictionary of Science and Technology



How to *do* science?



How to *do* science?



Goals for this lecture

- Outline the study design process
- Understand why study design helps us “do” science
- R tutorial: sample and data organization with *Ekipa Fanihy*

Study Design

Study Design

1. Define the **research question**: What are we trying to answer?

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?
5. Acknowledge **limitations**: What can we actually infer from our data?

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?
5. Acknowledge **limitations**: What can we actually infer from our data?
6. Outline a **data organization plan**: How should we organize our data?

Study Design

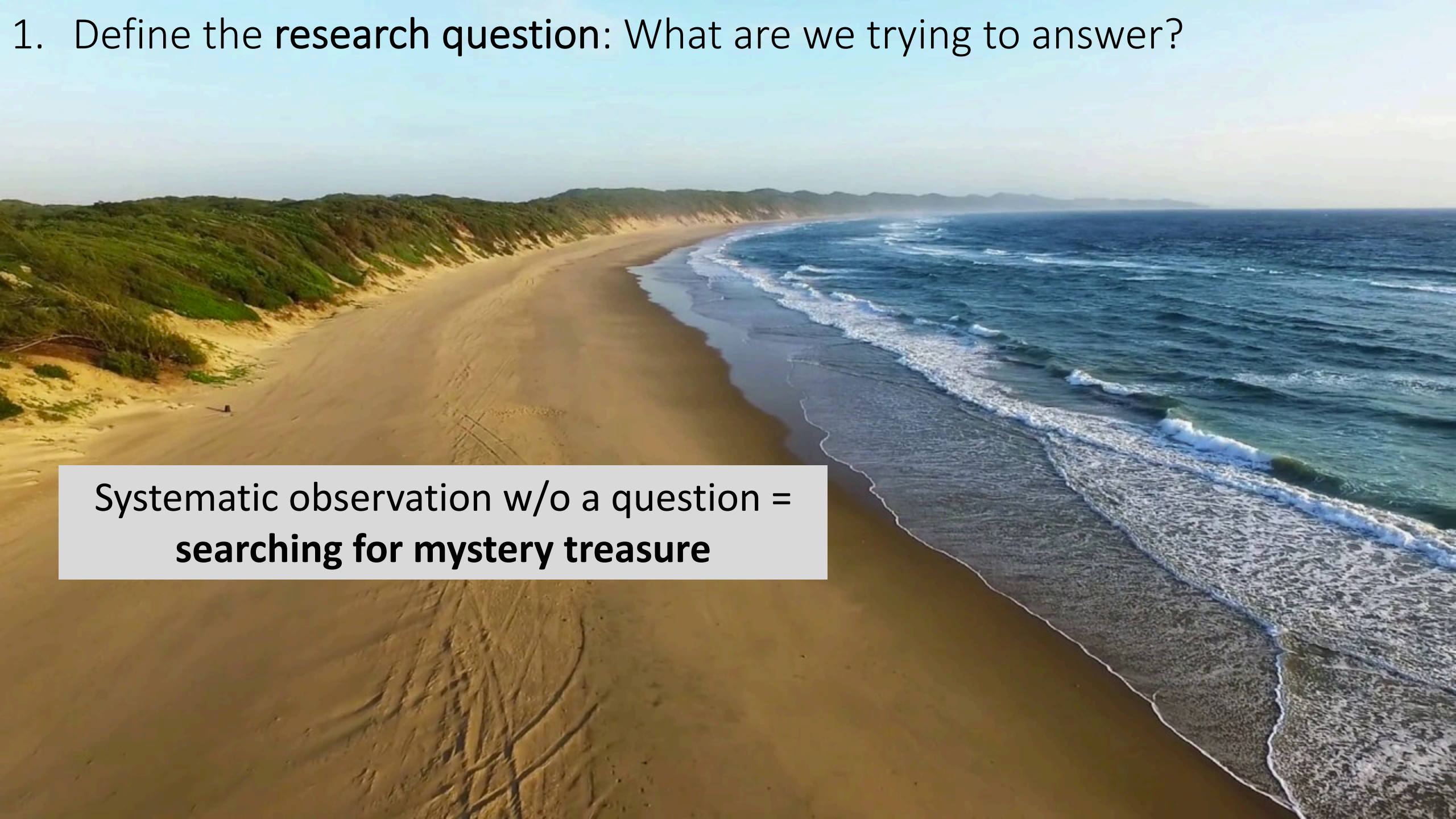
1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?
5. Acknowledge **limitations**: What can we actually infer from our data?
6. Outline a **data organization plan**: How should we organize our data?
7. Be **flexible**: How can we prepare for potential/unanticipated challenges?

1. Define the **research question**: What are we trying to answer?

Mahaliana

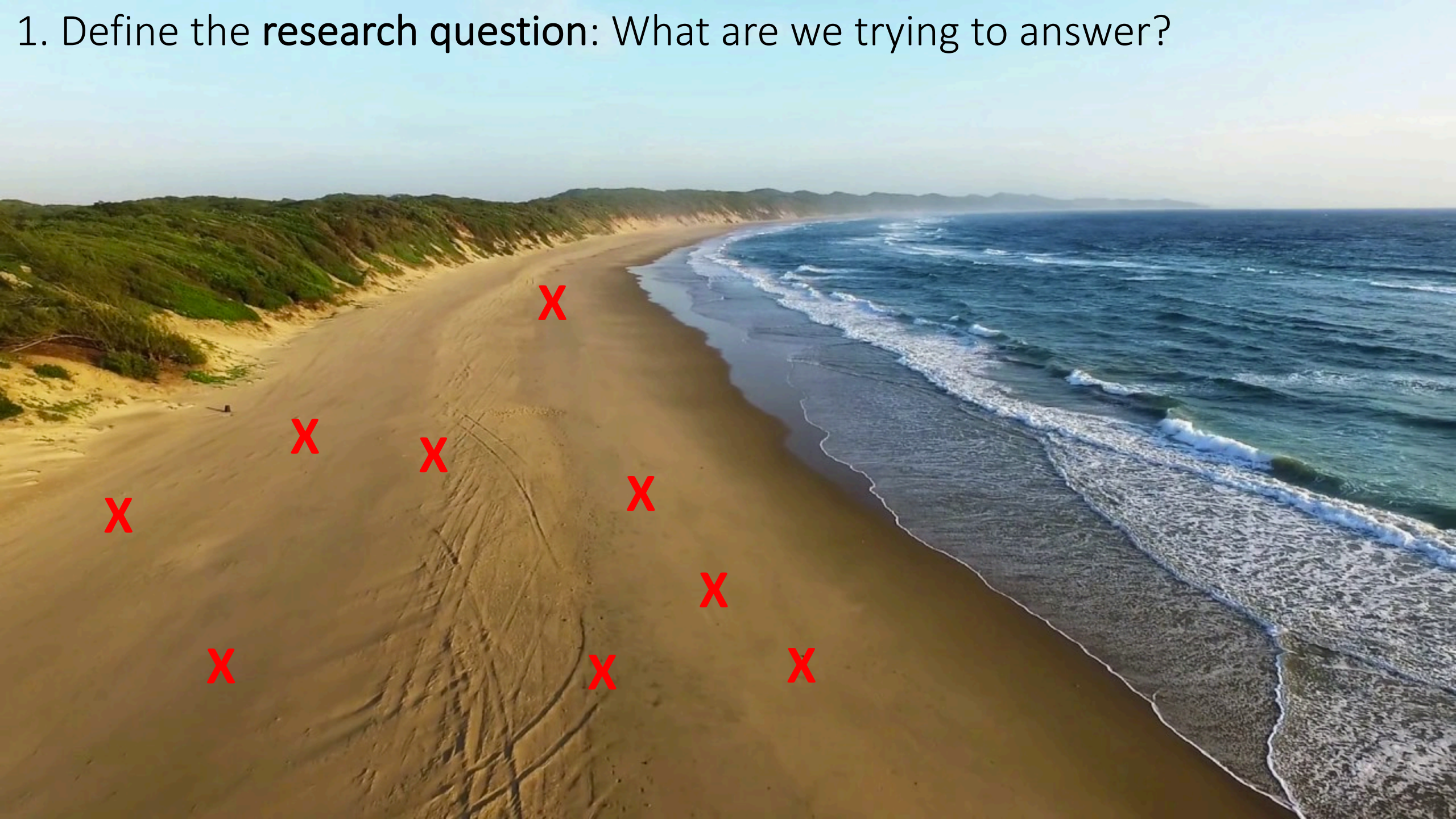
IT ALWAYS STARTS WITH A QUESTION

1. Define the **research question**: What are we trying to answer?



Systematic observation w/o a question =
searching for mystery treasure

1. Define the **research question**: What are we trying to answer?



1. Define the **research question**: What are we trying to answer?

1. Define the **research question**: What are we trying to answer?

1. Define the **research question**: What are we trying to answer?

1. Define the **research question**: What are we trying to answer?

Every component of study design hinges on the research question

- **Research question** = the target
- **Study design** = an effective and efficient route to answering the question

1. Define the **research question**: What are we trying to answer?

Every component of study design hinges on the research question

- **Research question** = the target
- **Study design** = an effective and efficient route to answering the question

Good research questions are:

1. Define the **research question**: What are we trying to answer?

Every component of study design hinges on the research question

- **Research question** = the target
- **Study design** = an effective and efficient route to answering the question

Good research questions are:

- Clear
- Focused and testable
- Original
- Based on previous observations

1. Define the **research question**: What are we trying to answer?

- Formulate a **hypothesis**
- Develop a **model** to demonstrate your hypothesis

Every component of study design hinges on the research question

- **Research question** = the target
- **Study design** = an effective and efficient route to answering the question

Good research questions are:

- Clear
- Focused and testable
- Original
- Based on previous observations

1. Define the **research question**: What are we trying to answer?

How do bat populations maintain virulent human-infection viruses?

1. Define the **research question**: What are we trying to answer?

How do bat populations maintain virulent human-infection viruses?

broad

Focused/ testable

What is the **force of infection** of henipaviruses in Madagascar *Eidolon dupreanum* populations?

2. Define the **sample type**: What data do we need to answer our question?

What is **data**?

2. Define the **sample type**: What data do we need to answer our question?

What is **data**?

= **evidence** to support a **claim**

Force of infection = Rate at which bats become infected

2. Define the **sample type**: What data do we need to answer our question?

What is **data**?

= **evidence** to support a **claim**

Data:

- Age data
- Serology

Force of infection = Rate at which bats become infected

2. Define the **sample type**: What data do we need to answer our question?

What is **data**?

= **evidence** to support a **claim**

Sample:
- Teeth
- Serum

→

Data:
- Age data
- Serology

→

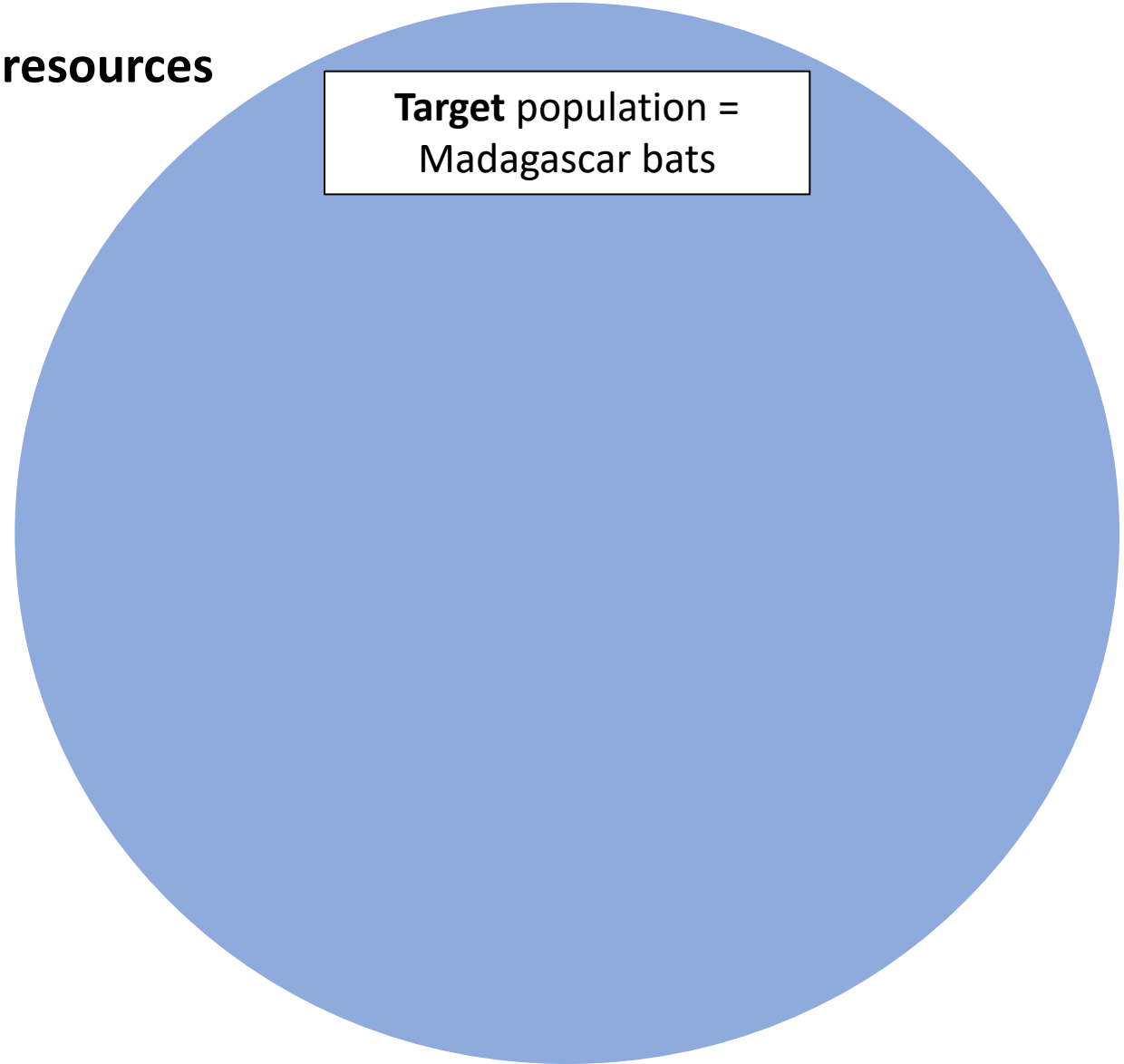
Force of infection = Rate at which bats become infected

3. Identify a **system**: Where can we collect our data?

Choosing a study population that allows you to **answer your research question**:

- **effectively**
- **tractably**—time, money, and effort are **limited resources**

Target population:
Want to make inferences about



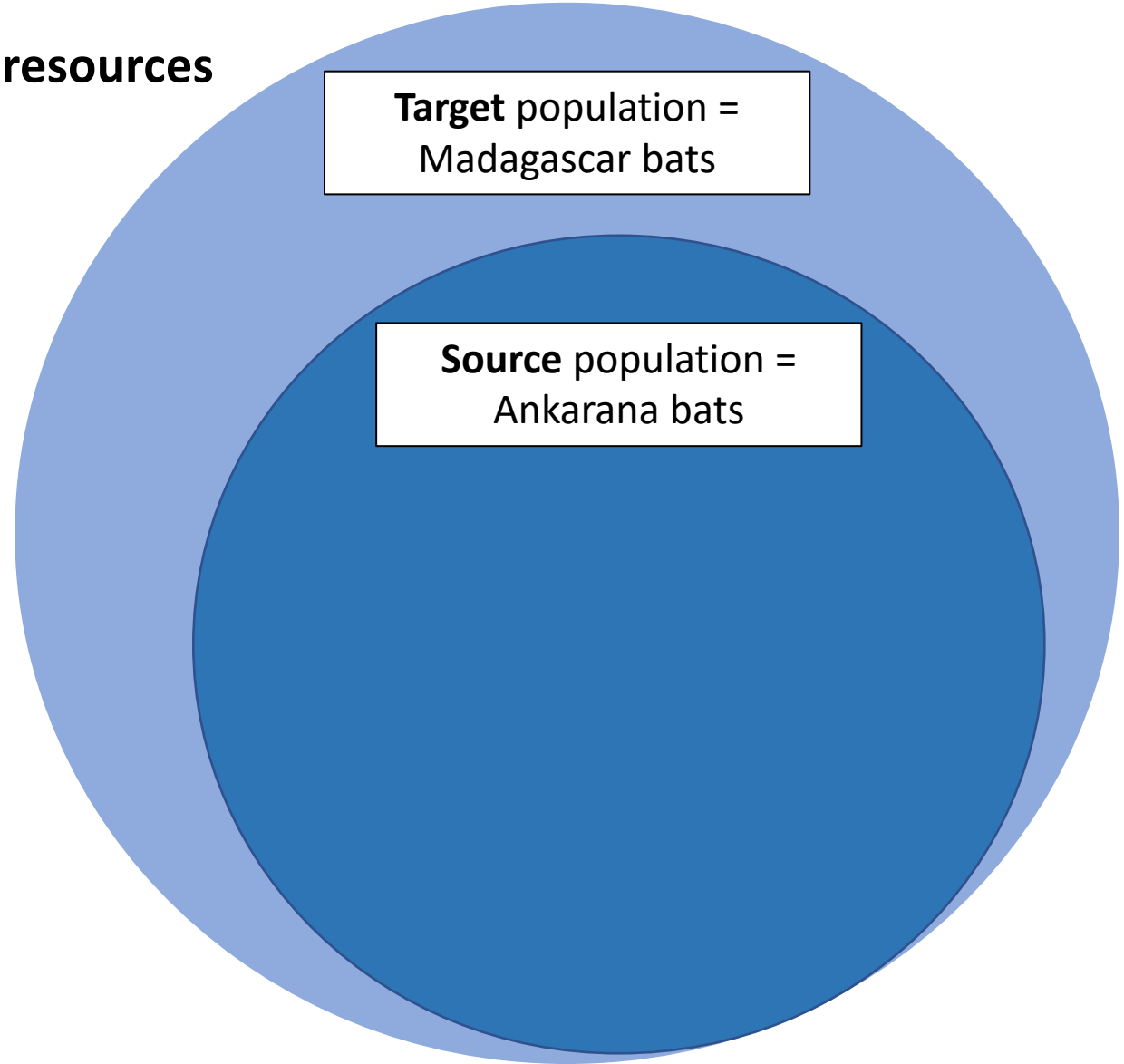
3. Identify a **system**: Where can we collect our data?

Choosing a study population that allows you to **answer your research question**:

- **effectively**
- **tractably**—time, money, and effort are **limited resources**

Target population:
Want to make inferences about

Source population:
Choosing study population



3. Identify a **system**: Where can we collect our data?

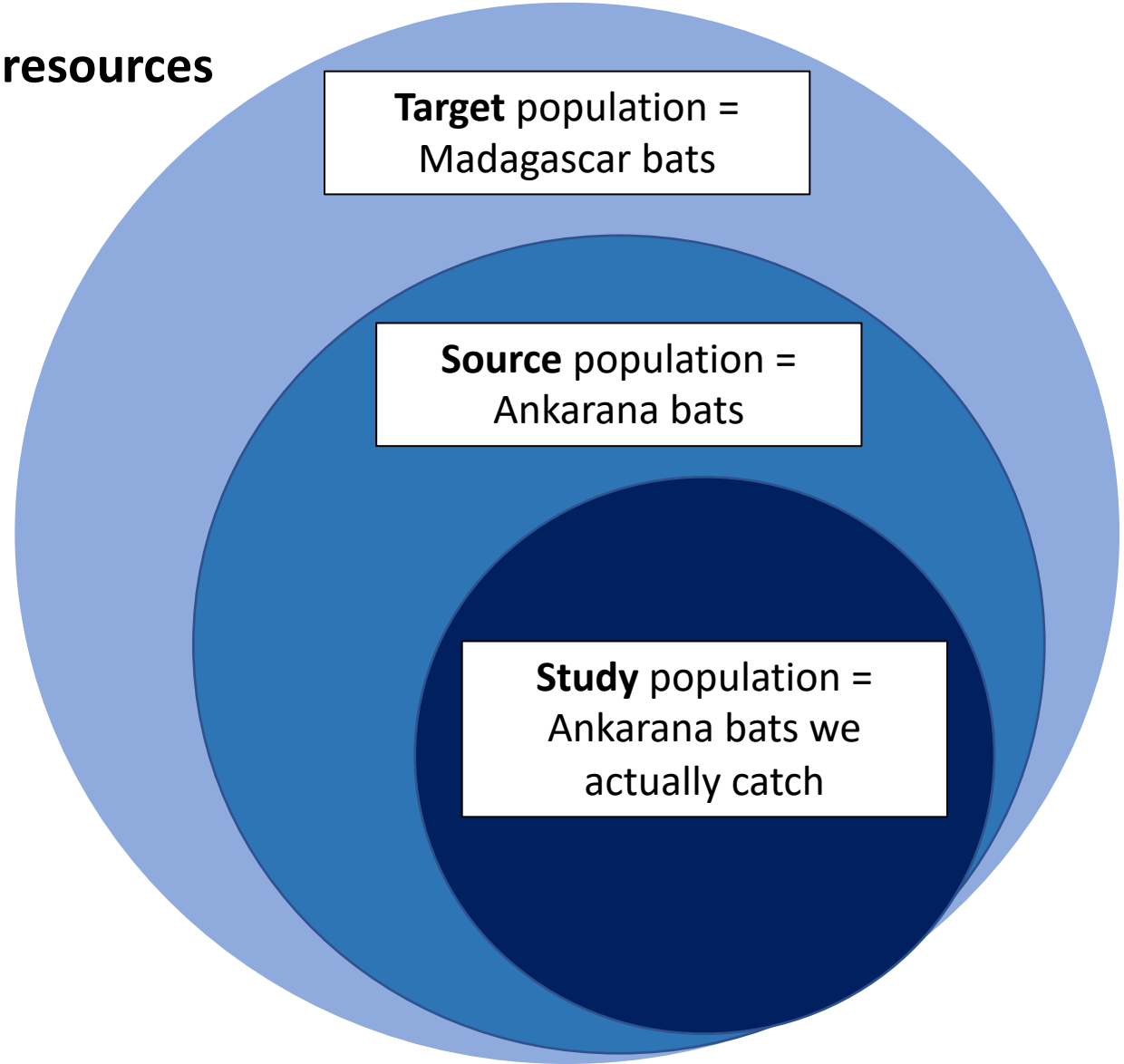
Choosing a study population that allows you to **answer your research question**:

- **effectively**
- **tractably**—time, money, and effort are **limited resources**

Target population:
Want to make inferences about

Source population:
Choosing study population

Study population:
Sampled individuals



3. Identify a **system**: Where can we collect our data?

Choosing a study population that allows you to **answer your research question**:

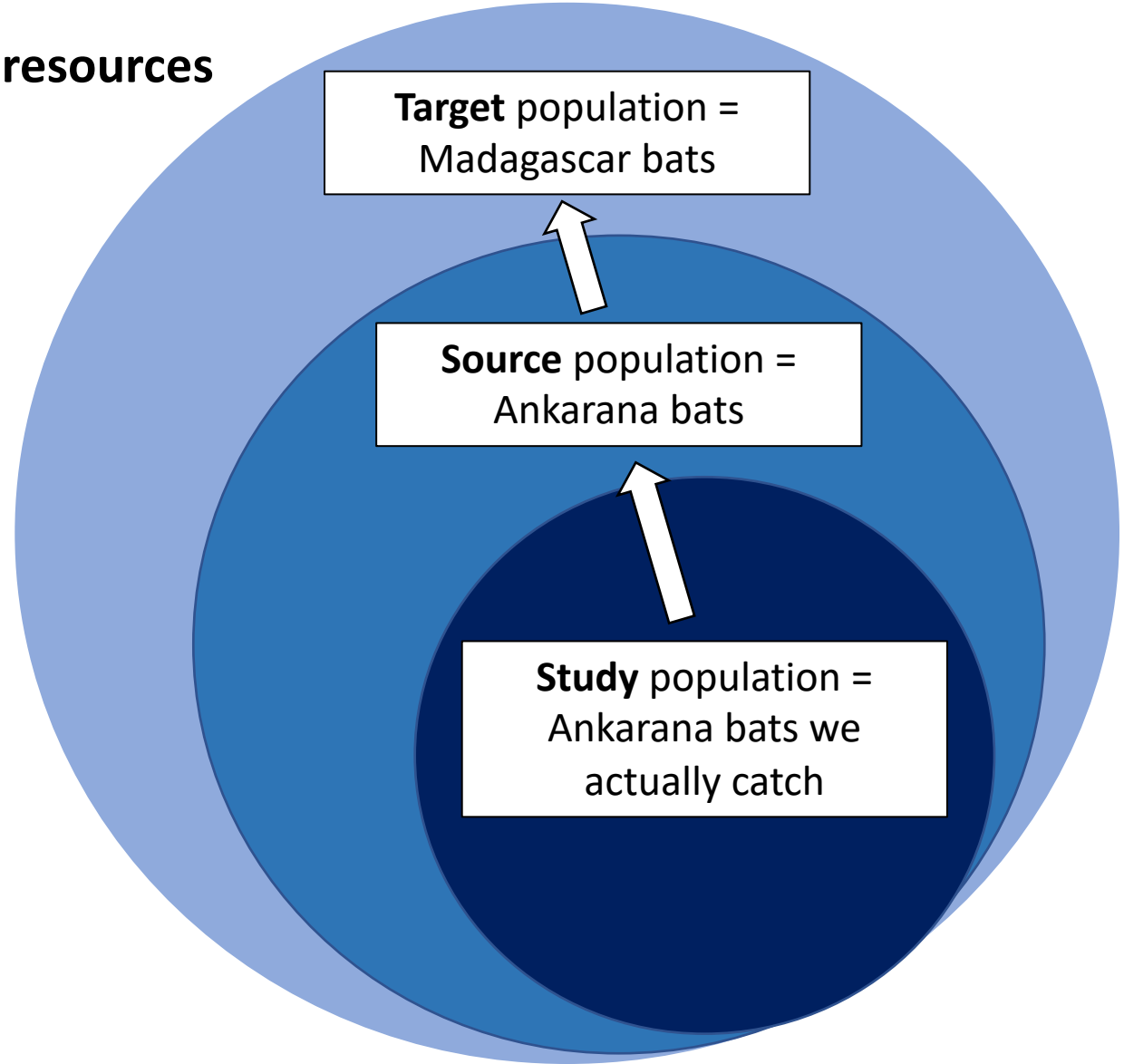
- **effectively**
- **tractably**—time, money, and effort are **limited resources**

Target population:
Want to make inferences about

Source population:
Choosing study population

Study population:
Sampled individuals

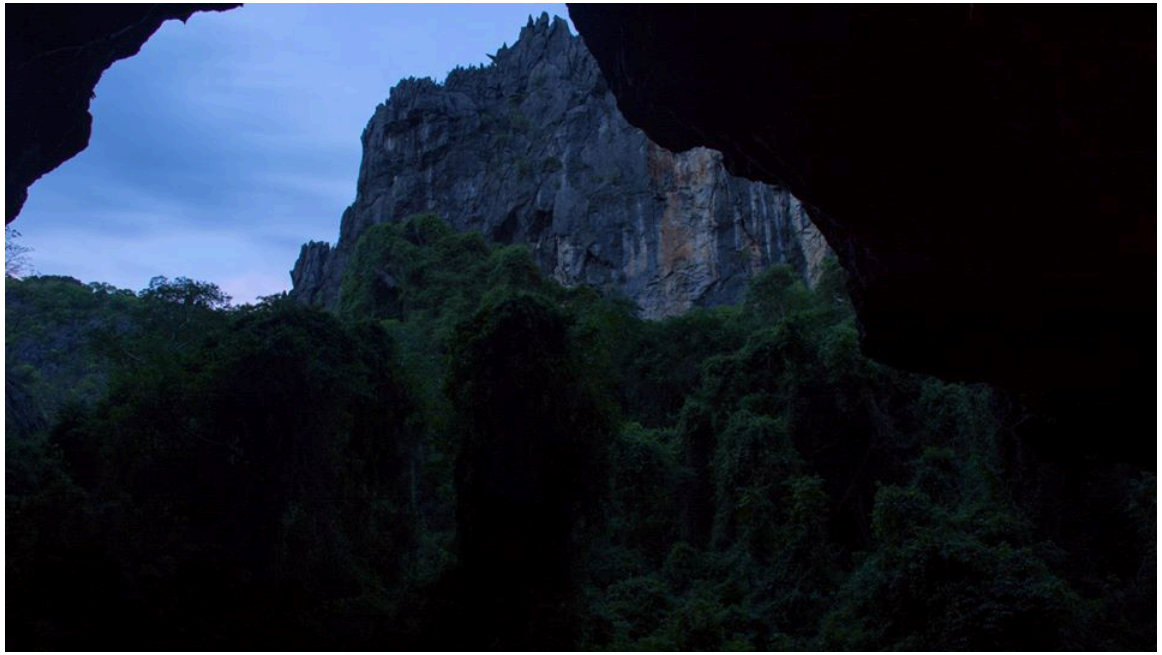
Well designed studies allow us to make inference about the target population



3. Identify a **system**: Where can we collect our data?

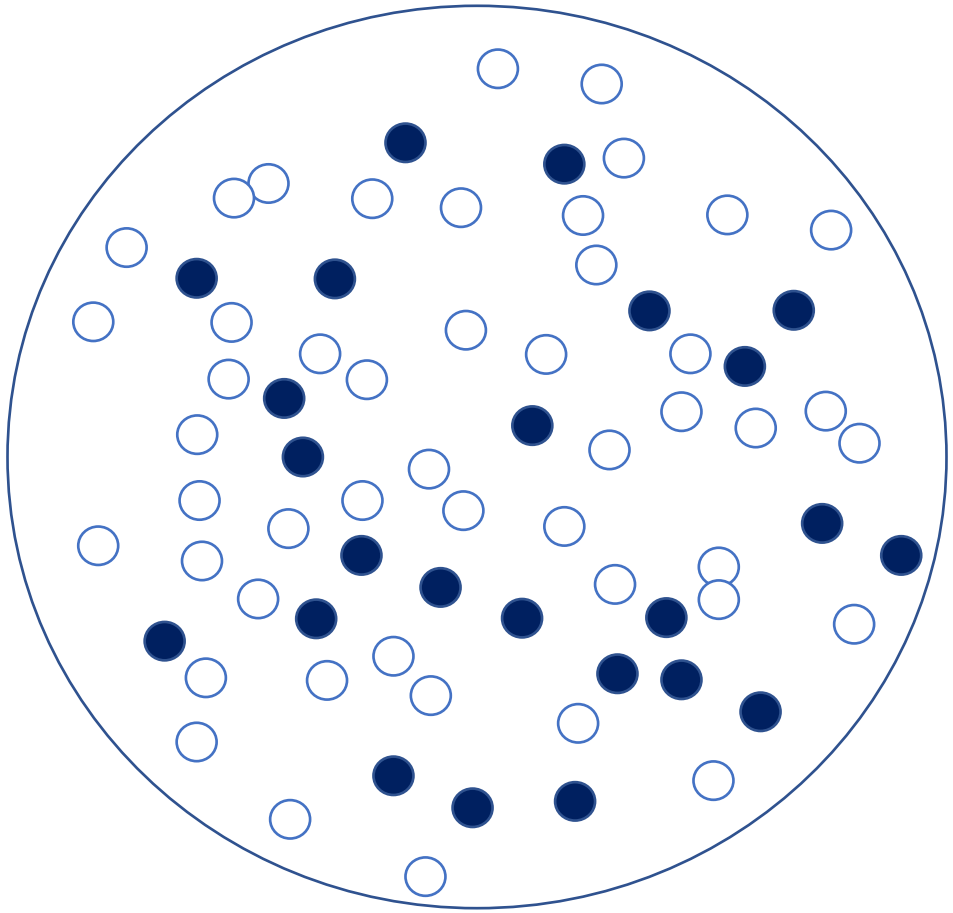
Choosing a study population that allows you to **answer your research question**:

- **effectively**
- **tractably**—time, money, and effort are **limited resources**



4. Choose a **sampling scheme**: How should we collect our data?

The **study** population should be **randomly selected** from the **source** population



Choose a **sampling scheme**: How should we collect our data?

Observational

- Descriptive
- Cross-sectional
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

There are several **study types** to choose from…
but not all types will be able to answer your question

Choose a **sampling scheme**: How should we collect our data?

Observational

- Descriptive
- Cross-sectional
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

Observational vs. Experimental?

4. Choose a **sampling scheme**: How should we collect our data?

Observational

- **Descriptive**
- Cross-sectional
- Longitudinal
- Ecological

Experimental

- Randomized Control Trial (RCT)

Descriptive studies = observational research that describes the **characteristics** of a population

- focus on the **what** instead of the **why**

Good for **generating hypotheses**, especially when data is limited

- **Bat survey:** we need to find the bats before we can study their viruses

Not all studies are hypothesis-driven…
but research really does always start with a question

4. Choose a **sampling scheme**: How should we collect our data?

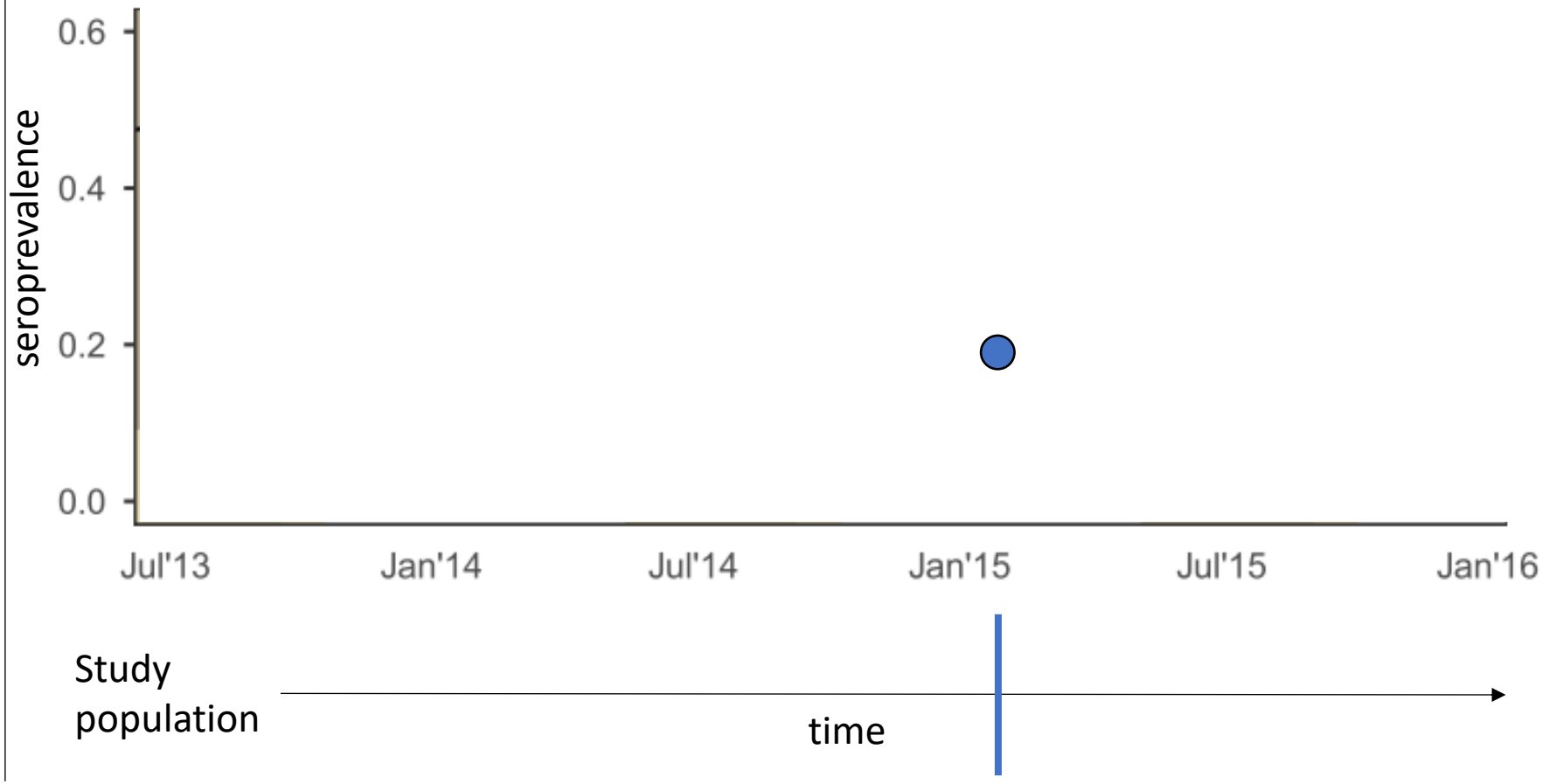
Observational

- Descriptive
- **Cross-sectional**
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

Cross-sectional studies = snapshot of the **population at a particula point in time**



4. Choose a **sampling scheme**: How should we collect our data?

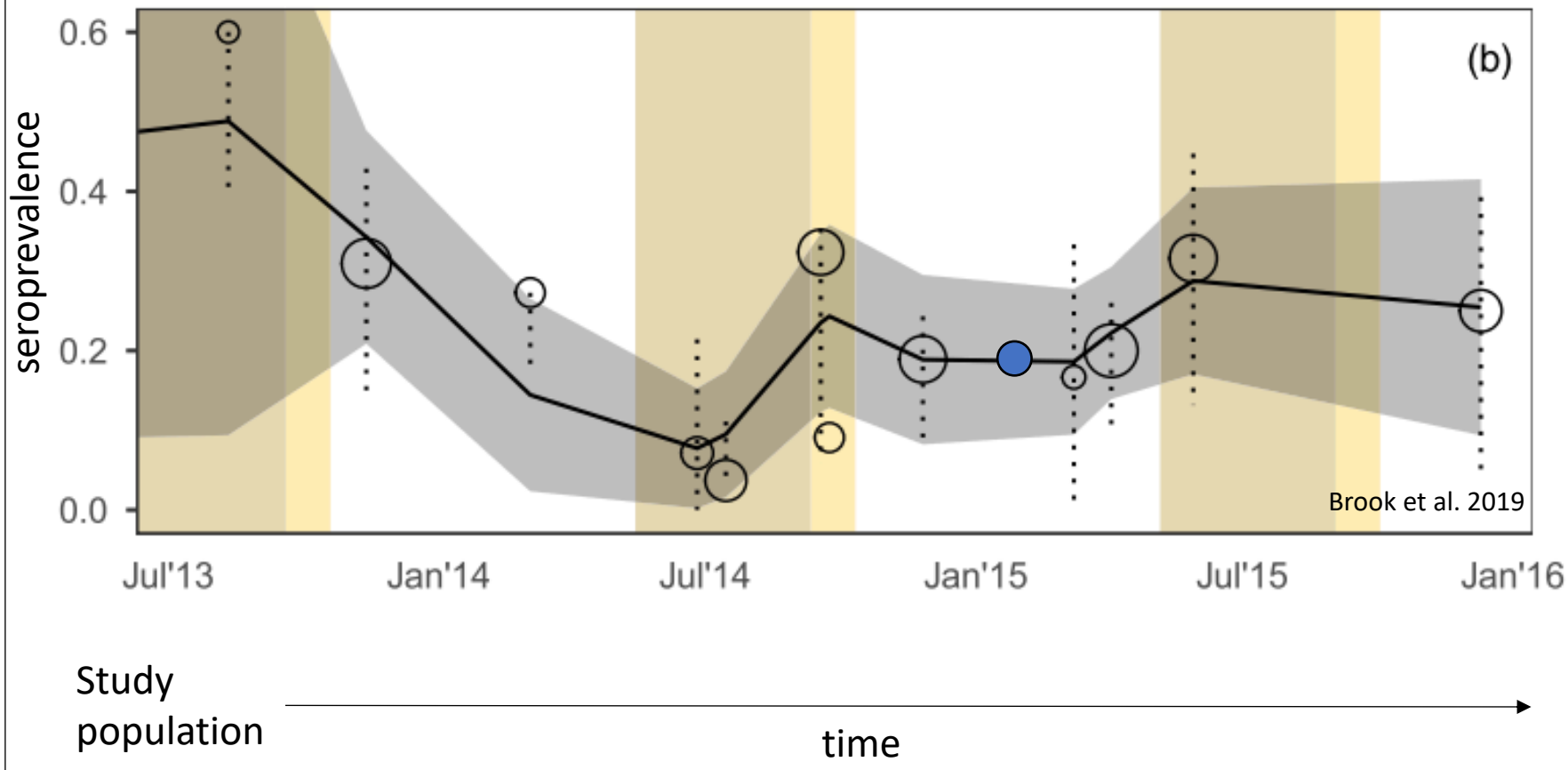
Observational

- Descriptive
- **Cross-sectional**
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

Cross-sectional studies = snapshot of the **population at a particula point in time**



4. Choose a **sampling scheme**: How should we collect our data?

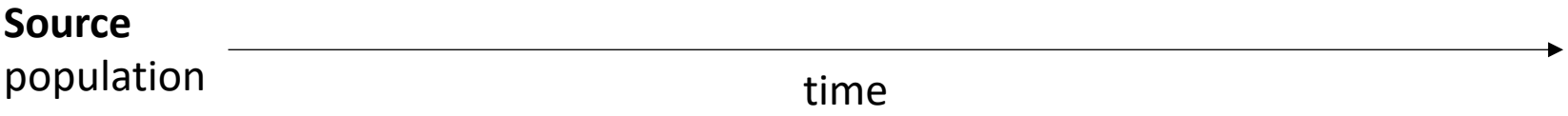
Observational

- Descriptive
- Cross-sectional
- **Longitudinal**
- Ecological

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

Longitudinal studies = follow a population over a **period of time**



- **Cohort** studies = follow a group of individuals over a **period of time**

4. Choose a **sampling scheme**: How should we collect our data?

Observational

- Descriptive
- Cross-sectional
- Longitudinal
- **Ecological**

Experimental

- Experimental Ecology
- Randomized Control Trial (RCT)

Unit of **comparison** = **populations** instead of individuals

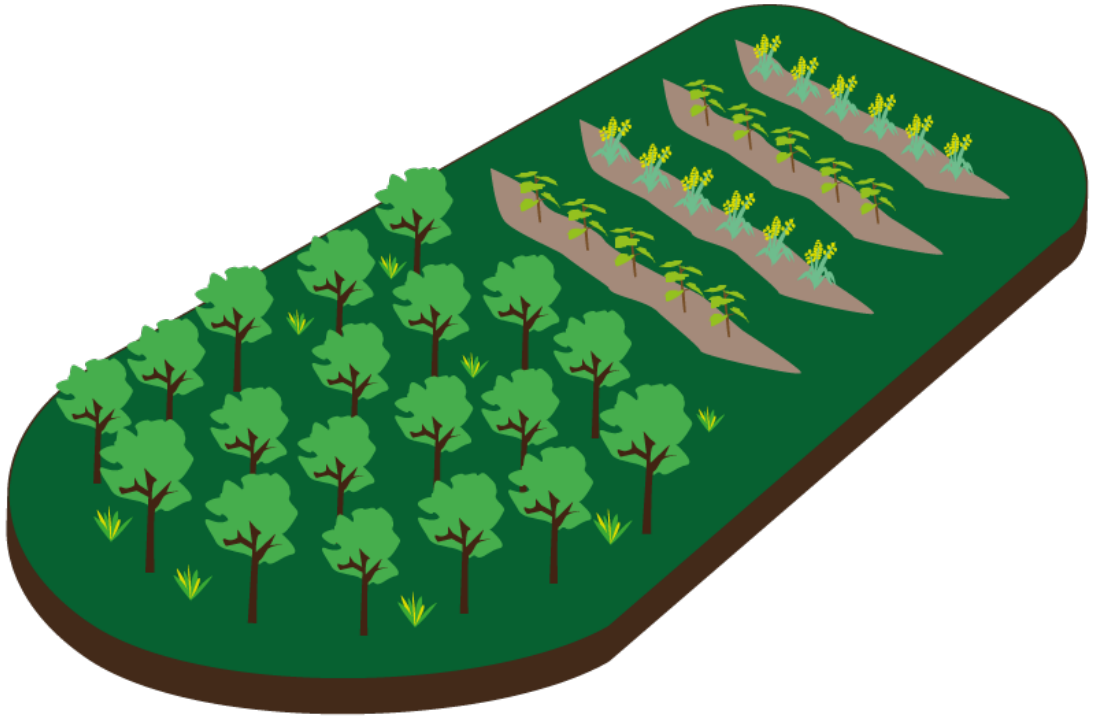
4. Choose a **sampling scheme**: How should we collect our data?

Observational

- Descriptive
- Cross-sectional
- Longitudinal
- Ecological

Experimental

- **Experimental Ecology**
- Randomized Control Trial (RCT)



4. Choose a **sampling scheme**: How should we collect our data?

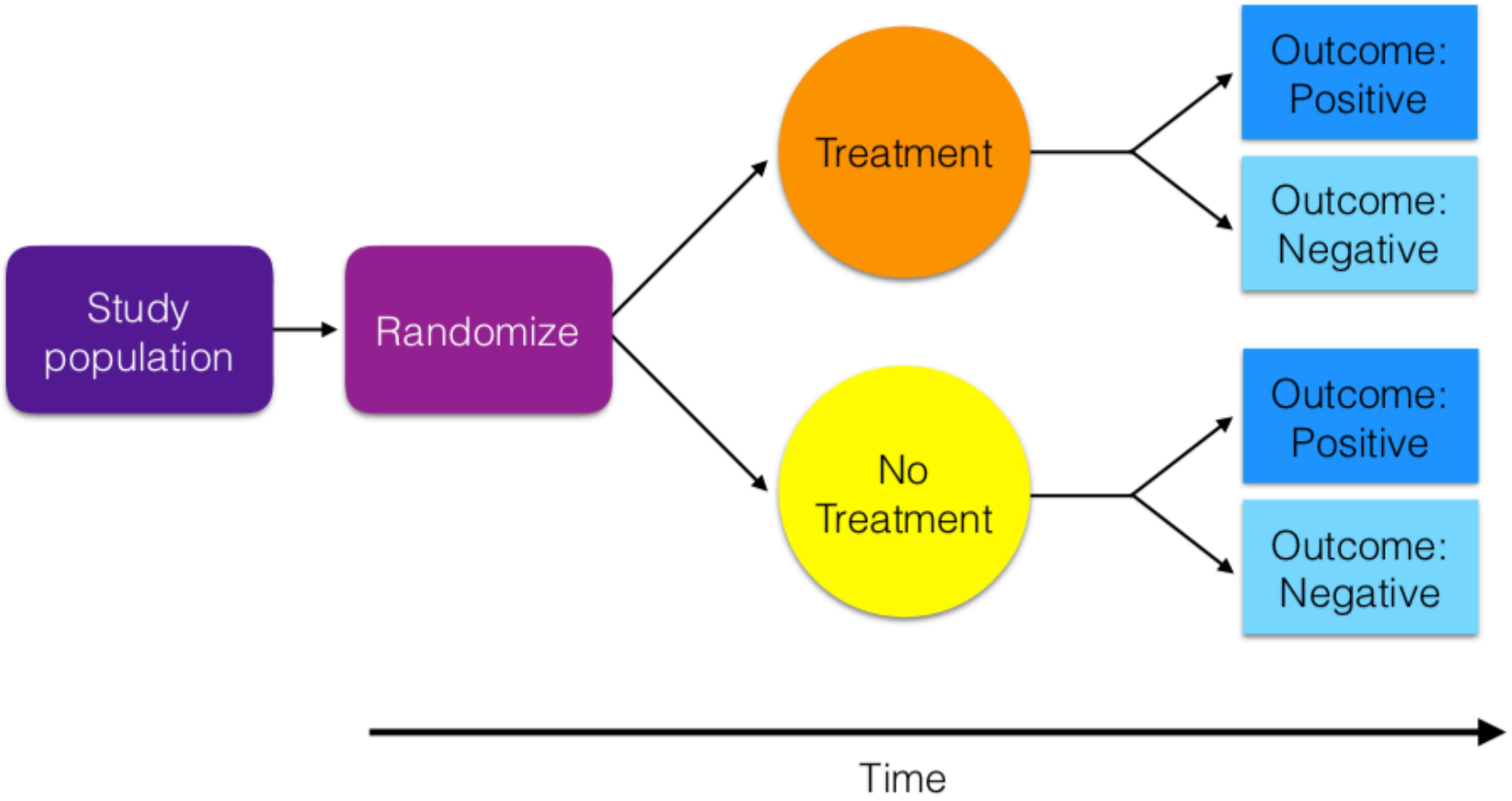
Observational

- Descriptive
- Cross-sectional
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- **Randomized Control Trial (RCT)**

Randomized Control Trial = subjects are randomly allocated into groups (test and control) to receive or not receive a treatment



4. Choose a **sampling scheme**: How should we collect our data?

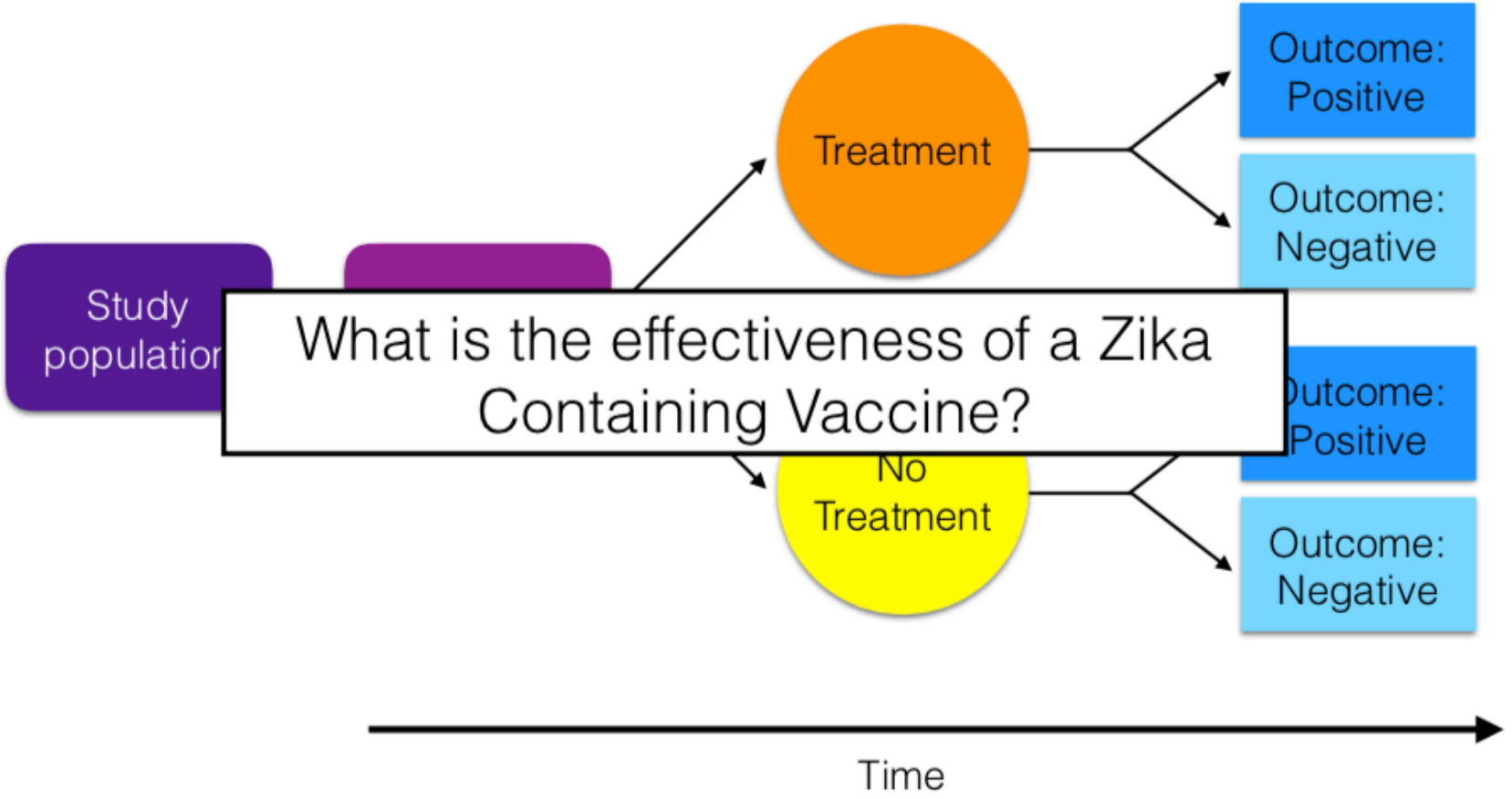
Observational

- Descriptive
- **Cross-sectional**
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- **Randomized Control Trial (RCT)**

Randomized Control Trial = subjects are randomly allocated into groups (test and control) to receive or not receive a treatment



4. Choose a **sampling scheme**: How should we collect our data?

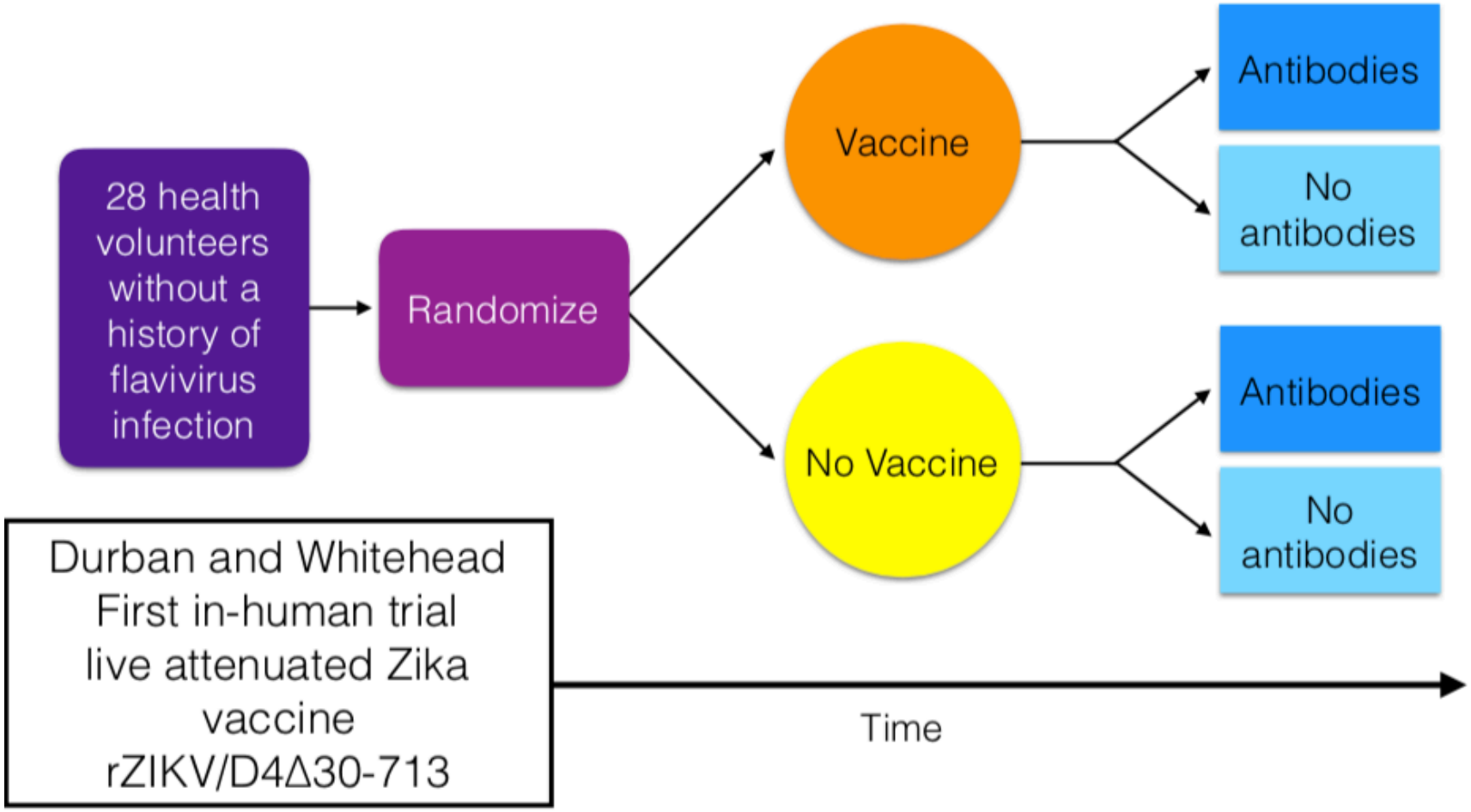
Observational

- Descriptive
- **Cross-sectional**
- Longitudinal
- Ecological

Experimental

- Experimental Ecology
- **Randomized Control Trial (RCT)**

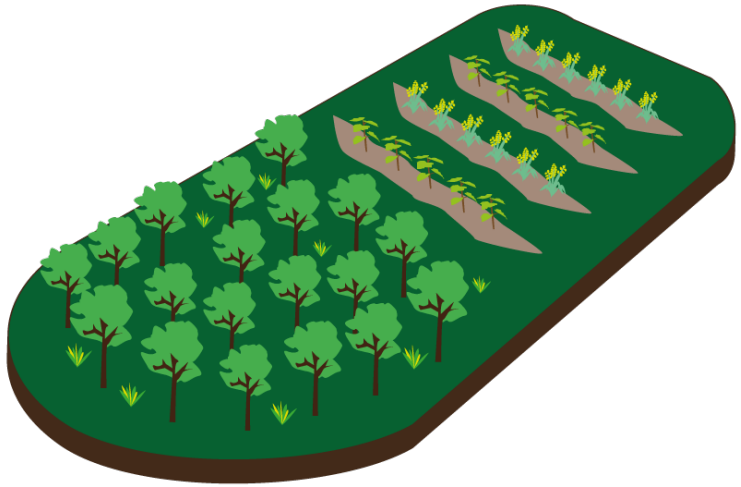
Randomized Control Trial = subjects are randomly allocated into groups (test and control) to receive or not receive a treatment

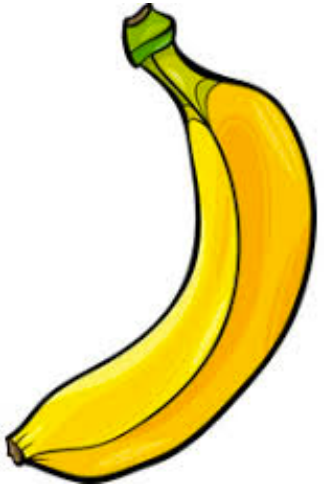


5. Acknowledge **limitations**: What can we actually infer from our data?

5. Acknowledge **limitations**: What can we actually infer from our data?

Systematic observation → **Formulate laws and principles**





5. Acknowledge **limitations**: What can we actually infer from our data?

Sample:
- Teeth
- Serum

→

Data:
- Age data
- Serology

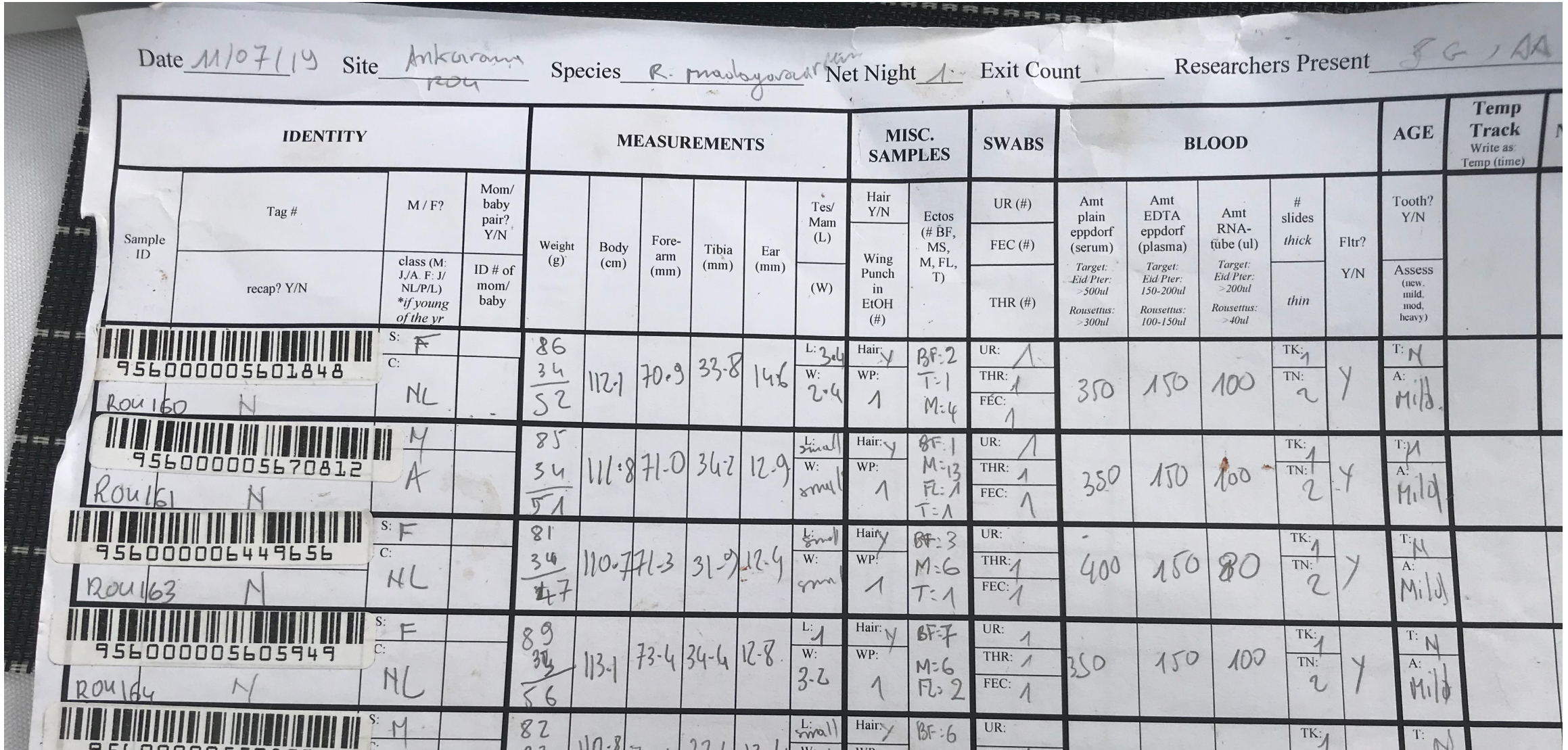
→

Force of infection = Rate at which bats become infected

6. Outline a **data organization plan**

6. Outline a **data organization plan**

Designing your **1) field datasheet**



Date 11/07/19 Site Ankarana RS4 Species R. madagascariensis Net Night 1 Exit Count _____ Researchers Present SG, AA

IDENTITY				MEASUREMENTS						MISC. SAMPLES		SWABS	BLOOD					AGE	Temp Track Write as: Temp (time)	
Sample ID	Tag # / recap? Y/N	M / F? / class (M: J/A. F: J/ NL/P/L) *if young of the yr	Mom/ baby pair? Y/N / ID # of mom/ baby	Weight (g)	Body (cm)	Fore-arm (mm)	Tibia (mm)	Ear (mm)	Tes/ Mam (L) / (W)	Hair Y/N / Wing Punch in EtOH (#)	Ectos (# BF, MS, M, FL, T)	UR (#) / FEC (#) / THR (#)	Amt plain eppdorf (serum) Target: Eid/Pter: >500ul Rousettus: >300ul	Amt EDTA eppdorf (plasma) Target: Eid/Pter: 150-200ul Rousettus: 100-150ul	Amt RNA-tube (ul) Target: Eid/Pter: >200ul Rousettus: >40ul	# slides thick / thin	Fltr? Y/N	Tooth? Y/N / Assess (new, mild, mod, heavy)		
9560000056018 48 ROU160	S: F C: NL	N		86 34 52	112.1	70.9	33.8	14.6	L: 3.6 W: 2.4	Hair: Y WP: 1	BF: 2 T: 1 M: 4	UR: 1 THR: 1 FEC: 1	350	150	100	TK: 1 TN: 2	Y	T: N A: Mild		
9560000056708 12 ROU161	S: M C: A	N		85 34 51	111.8	71.0	34.2	12.9	L: small W: small	Hair: Y WP: 1	BF: 1 M: 13 FL: 1 T: 1	UR: 1 THR: 1 FEC: 1	350	150	100	TK: 1 TN: 2	Y	T: N A: Mild		
9560000064496 56 ROU163	S: F C: NL	N		81 34 47	110.7	71.3	31.9	12.4	L: small W: small	Hair: Y WP: 1	BF: 3 M: 6 T: 1	UR: THR: 1 FEC: 1	400	150	80	TK: 1 TN: 2	Y	T: N A: Mild		
9560000056059 49 ROU164	S: F C: NL	N		89 34 56	113.1	73.4	34.4	12.8	L: 1 W: 3.2	Hair: Y WP: 1	BF: 7 M: 6 FL: 2	UR: 1 THR: 1 FEC: 1	350	150	100	TK: 1 TN: 2	Y	T: N A: Mild		
9560000056[illegible]	S: M C:			82	110.8				L: small	Hair: Y	BF: 6	UR:				TK: 1		T:		

6. Outline a **data organization plan**

Designing your 1) field datasheet

2) database structure

Compiling data the “long way”

- identifying information stored in columns

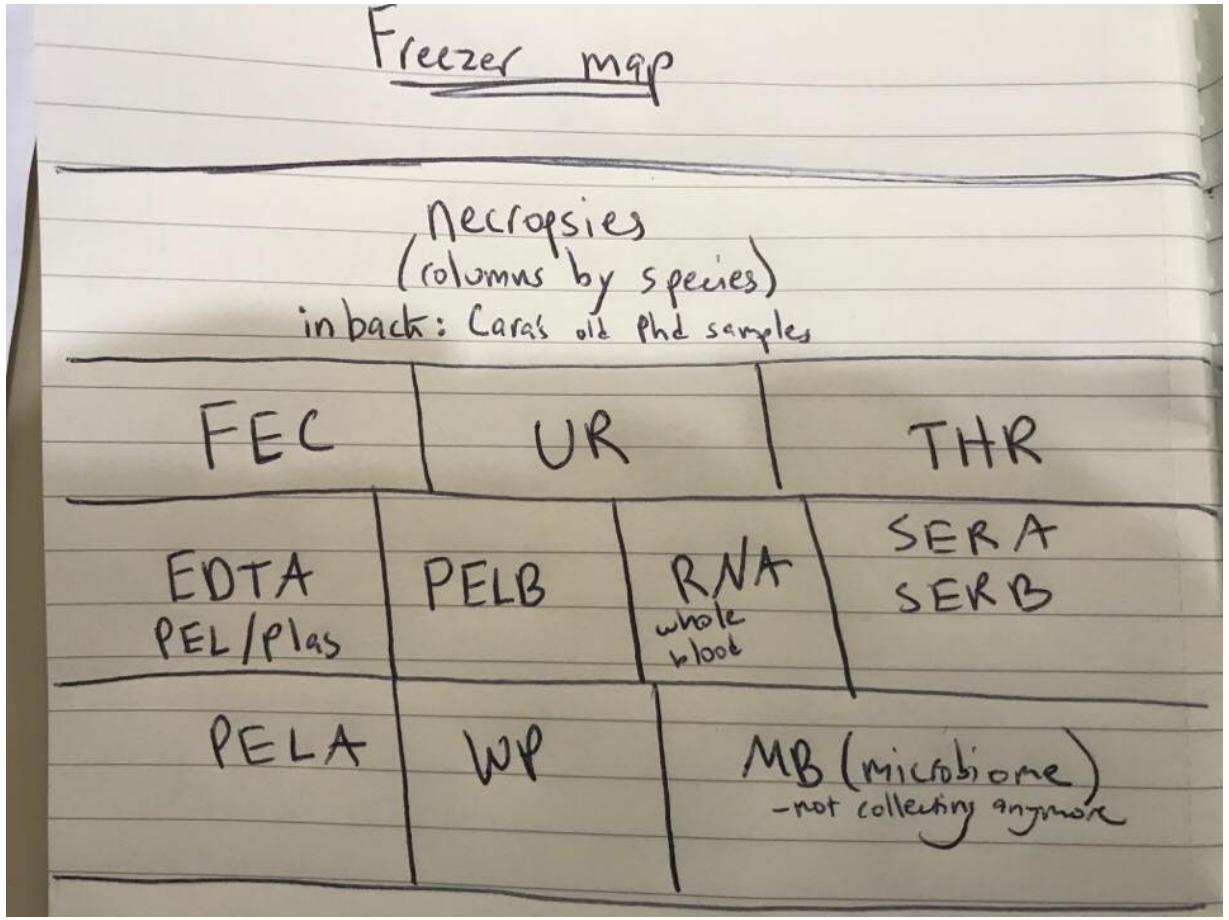
SITE DATA				Identification	
Roost Site	Researchers	Date	Net Night	Bat Species	Sample ID
Lakato	CB, CR, LA	8/22/13	2	Eidolon dupr	LKT-001
Lakato	CB, CR, LA	8/22/13	2	Eidolon dupr	LKT-002
Lakato	CB, CR, LA	8/22/13	2	Eidolon dupr	LKT-003
Lakato	CB, CR, LA	8/22/13	2	Eidolon dupr	LKT-004
Lakato	CB, CR, LA	8/22/13	2	Eidolon dupr	LKT-005
Marovitsika	CB, CR, AR	11/4/13	1	Pteropus rufi	MARO1
Marovitsika	CB, CR, AR	11/4/13	1	Pteropus rufi	MARO2
Marovitsika	CB, CR, AR	11/4/13	1	Pteropus rufi	MARO3
Marovitsika	CB, CR, AR	11/4/13	1	Pteropus rufi	MARO4
Marovitsika	CB, CR, AR	11/4/13	1	Pteropus rufi	MARO5
Marovitsika	CB, CR, AR	11/6/13	3	Pteropus rufi	MARO6
Marovitsika	CB, CR, AR	11/7/13	4	Pteropus rufi	MARO7
Marovitsika	CB, CR, AR	11/7/13	4	Pteropus rufi	MARO8
Marovitsika	CB, CR, AR	11/7/13	4	Pteropus rufi	MARO9
Marovitsika	CB, CR, AR	11/8/13	5	Pteropus rufi	MARO10
Marovitsika	CB, CR, AR	11/8/13	5	Pteropus rufi	MARO11
Marovitsika	CB, CR, AR	11/8/13	5	Pteropus rufi	MARO12
Marovitsika	CB, CR, AR	11/8/13	5	Pteropus rufi	MARO13
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB1
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB2
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB3
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB4
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB5
Ambakoana	CB,CR, AR, R	11/15/13	1	Pteropus rufi	AMB6

6. Outline a **data organization plan**

Designing your 1) field datasheet

2) database structure

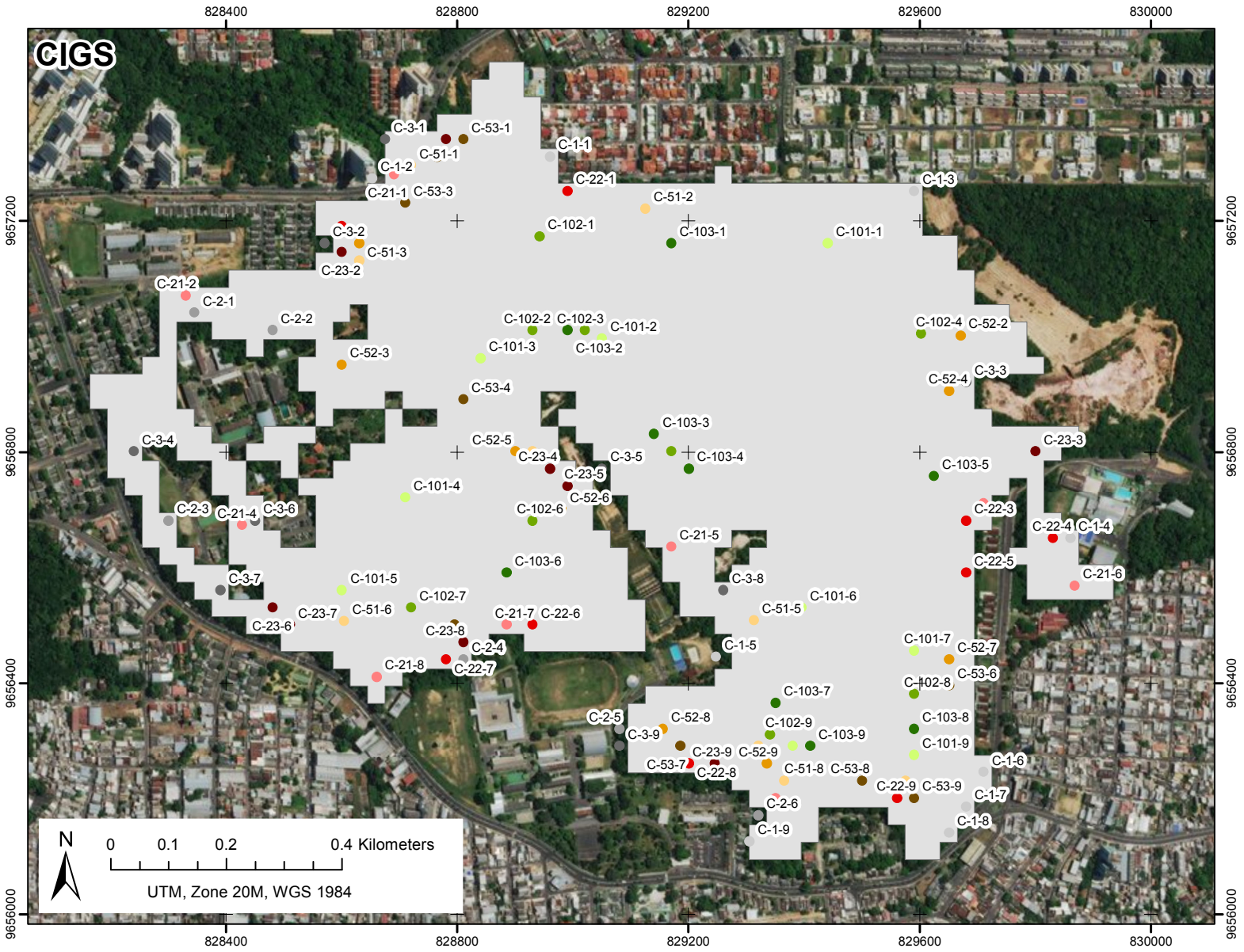
3) sample storage system



Box Number	Sample Type	Species	Reagent	SampleID	Date Stocka	Notes
PR-FEC-1	Feces	Asio madaga	UTM	AMAK_001	4/25/18	
PR-RNA-BLD	Whole Blood	Asio madaga	RNA-protect	AMAK_001	4/25/18	
PR-THR-1	Throat	Asio madaga	UTM	AMAK_001	4/25/18	
PR-SER-A-1	Serum A	Asio madaga	raw	AMAK_001	4/25/18	
PR-SER-B-1	Serum B	Asio madaga	raw	AMAK_001	4/25/18	
PR-PEL-A-1	Blood Pellet	Asio madaga	raw	AMAK_001	4/25/18	
PR-PEL-B-1	Blood Pellet	Asio madaga	raw	AMAK_001	4/25/18	

7. Be **flexible**

7. Be **flexible**



Study Design: Examples from E2M2?

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?
5. Acknowledge **limitations**: What can we actually infer from our data?
6. Outline a **data organization plan**: How should we organize our data?
7. Be **flexible**: How can we prepare for potential/unanticipated challenges?