

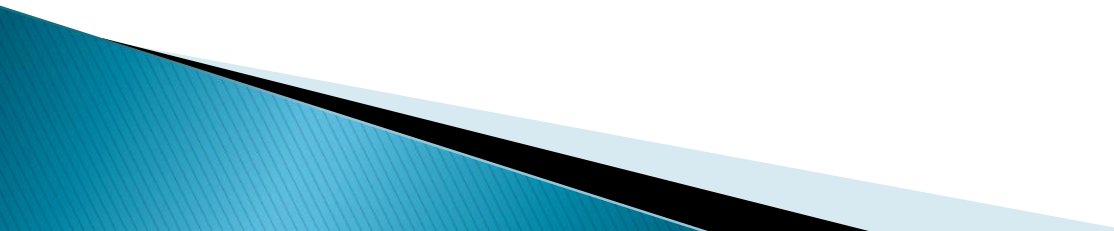
BioDialog Project: Activities & Outcomes at Ainshams university

Assoc. Prof. Ibrahim Fathy Moawad

Faculty of Computer and Information Sciences – Ain Shams University

Ain Shams BioDialog Project PI

Agenda

- ▶ BioDialog Project
 - ▶ Biodiversity Overview
 - ▶ Biodiversity Informatics
 - ▶ ASBIRG & Current Researches
- 

BioDialog Project

PI: Prof. Birgitta König-Ries



BioDialog Project


- ▶ Intercultural Dialog through **Biodiversity Informatics**: methods and techniques of managing biodiversity data.
 - ▶ Partners
 1. Germany: Jena University.
 2. Egypt: Ain Shams & Assuit Universities
 3. Tunisia: Safex University.
- 

BioDialog: Motivation

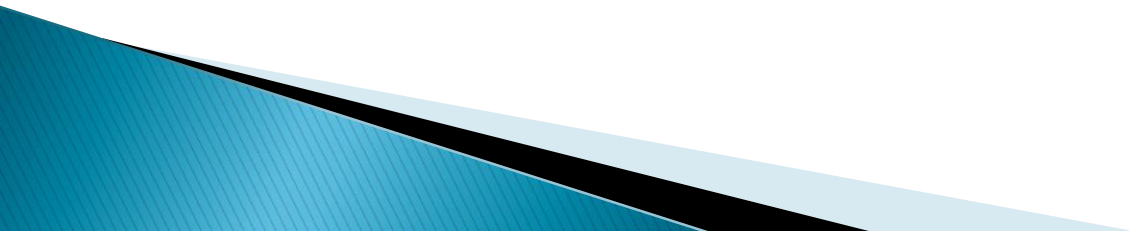
- ▶ A recent study: **the state of biodiversity informatics** for different countries (King, 2011).
 1. **The biodiversity potential (Biodiversity richness)**: physical, biological and environmental characteristics.
 2. **The capacity to generate biodiversity data records: raw data with high quality** (specimens, samples, observations).
 3. **The availability of technical infrastructure for hosting, managing and sharing** biodiversity data records.
- ▶ Order: Germany: 12, Egypt: 88 and Tunisia: 103.

Need for education & more research on biodiversity and on biodiversity informatics.

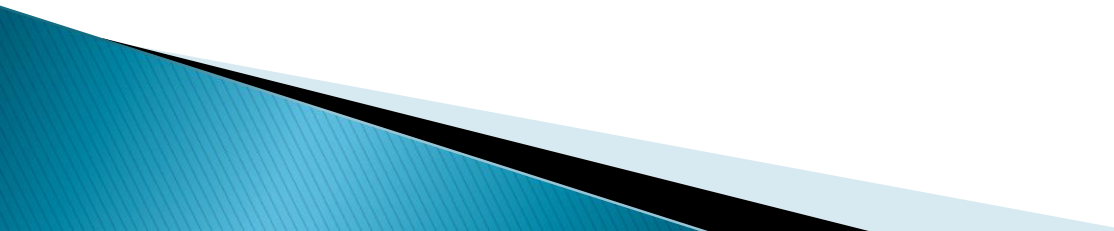
BioDialog: Objectives

- ▶ To establish a scientific exchange:
 1. **Understanding** biodiversity and biodiversity informatics practices in local context.
 2. Constructing a new **regional research network**.
 3. **Contributing** in the development of a **knowledge-based society**.
 4. **Bridging** between data management techniques and biodiversity research.
 5. **Awareness** of the importance of biodiversity is crucial to ensure its **global preservation**.
- 

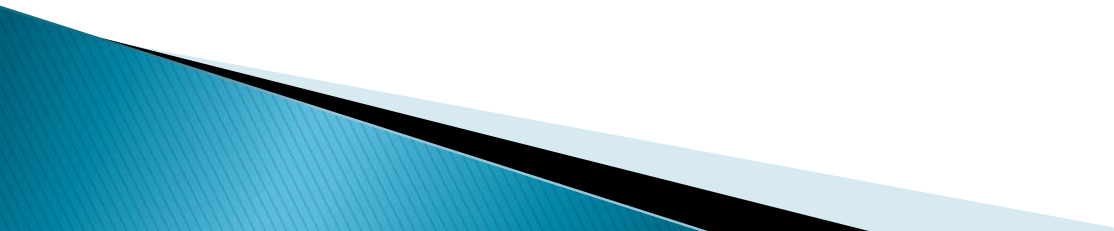
Biodiversity Overview



Biodiversity questions?

- ▶ **Detection**: How much biodiversity is there?
 - ▶ **Emergence**: Why does Biodiversity exist?
 - ▶ **Consequences**: Whether Biodiversity matters for ecosystem functions and services?
 - ▶ **Conservation**: How can explore ways to safeguard biodiversity.
- 

Facts about Biodiversity

- ▶ Most estimates fall between **5 million and 30 million species** currently living on Earth.
 - ▶ Most living species are **microorganisms and tiny invertebrates**.
 - ▶ Roughly **1.75 million species** have been **formally described and given official names**.
 - ▶ Medicines: **118 of the top 150 prescription drugs** in America contain chemicals derived from plants, fungi and other species.
- 

Facts about Biodiversity

disappear
unnoticed

Estimated
50% loss
until
2200

10% described

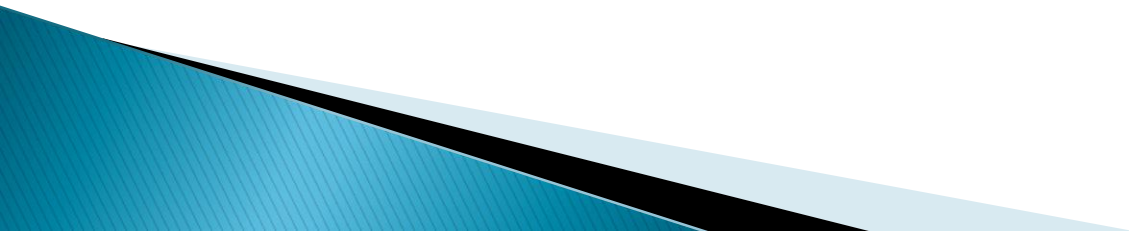


The Jena Experiment

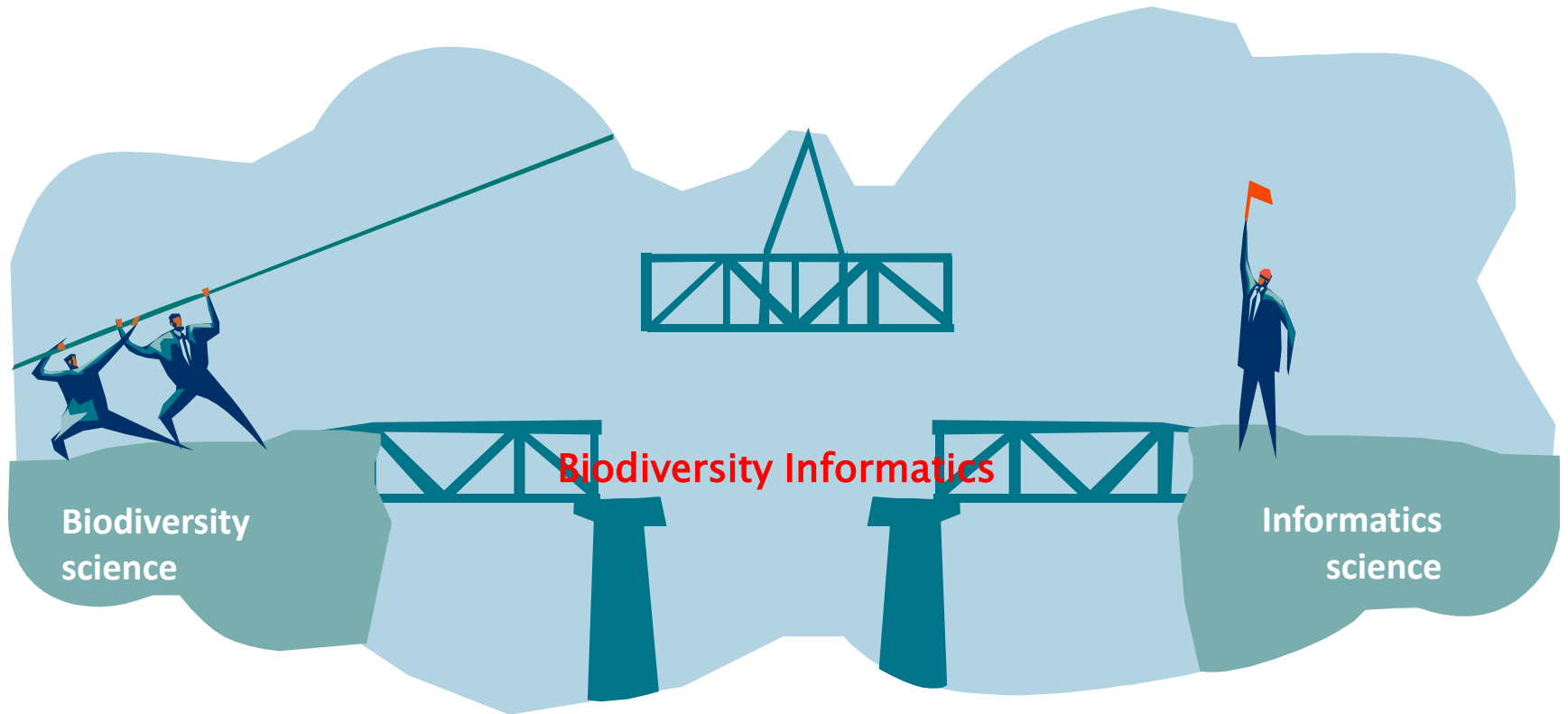


- ▶ One of **the longest-running biodiversity experiments** in Europe.
- ▶ Studying biodiversity effects in **experimental grassland communities** for more than 10 years.
- ▶ Investigation of **above-ground and below-ground consumers and processes**.
- ▶ Discover biodiversity **effects on ecosystem functioning**.

Biodiversity Informatics



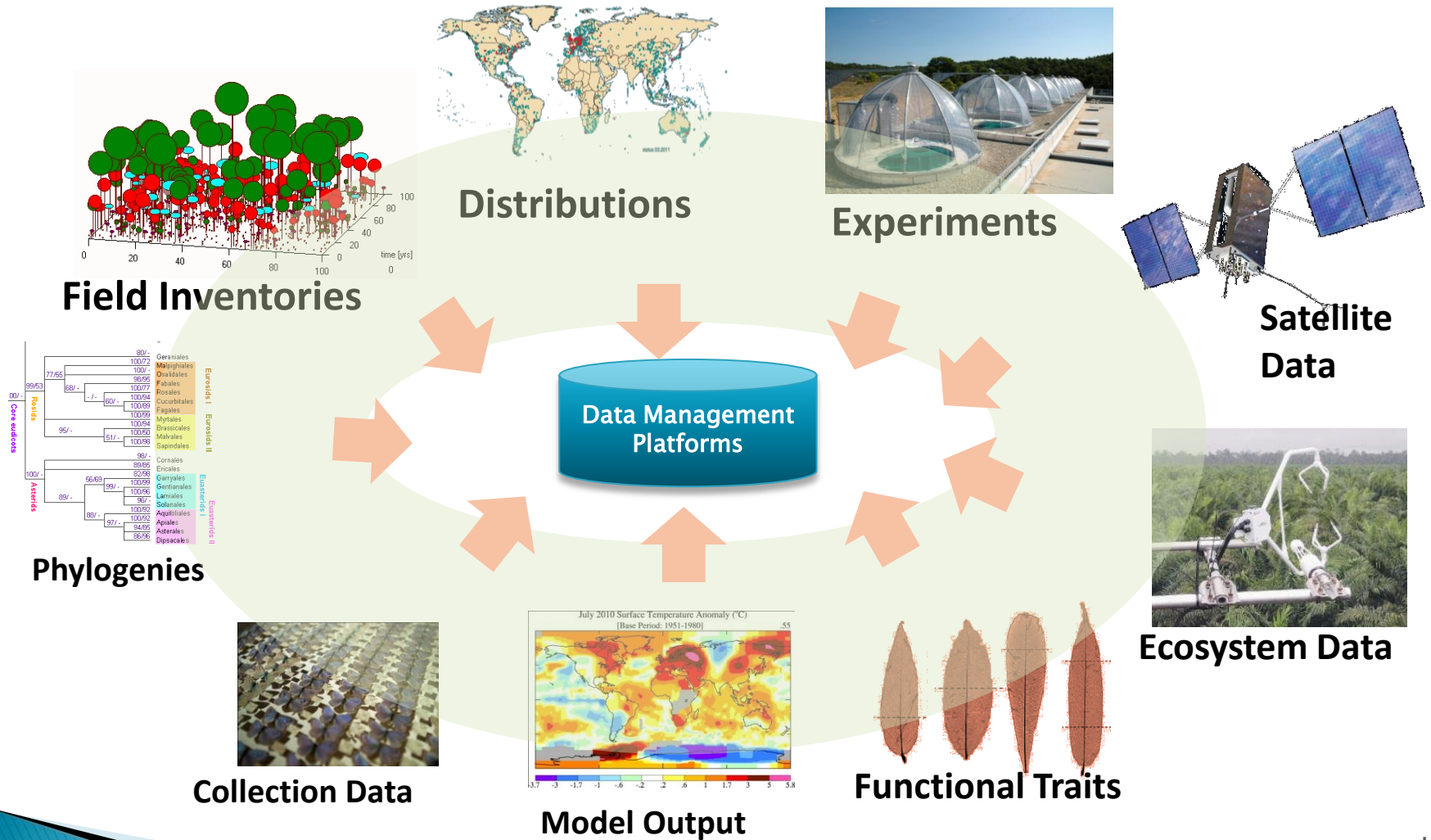
What is Biodiversity Informatics?



using informatics tools and applications to manage, disseminate, analyse, share, publish and discover biodiversity data & information.

Big Challenge

find and integrate all these data types



Have **YOU** been in this Situation?

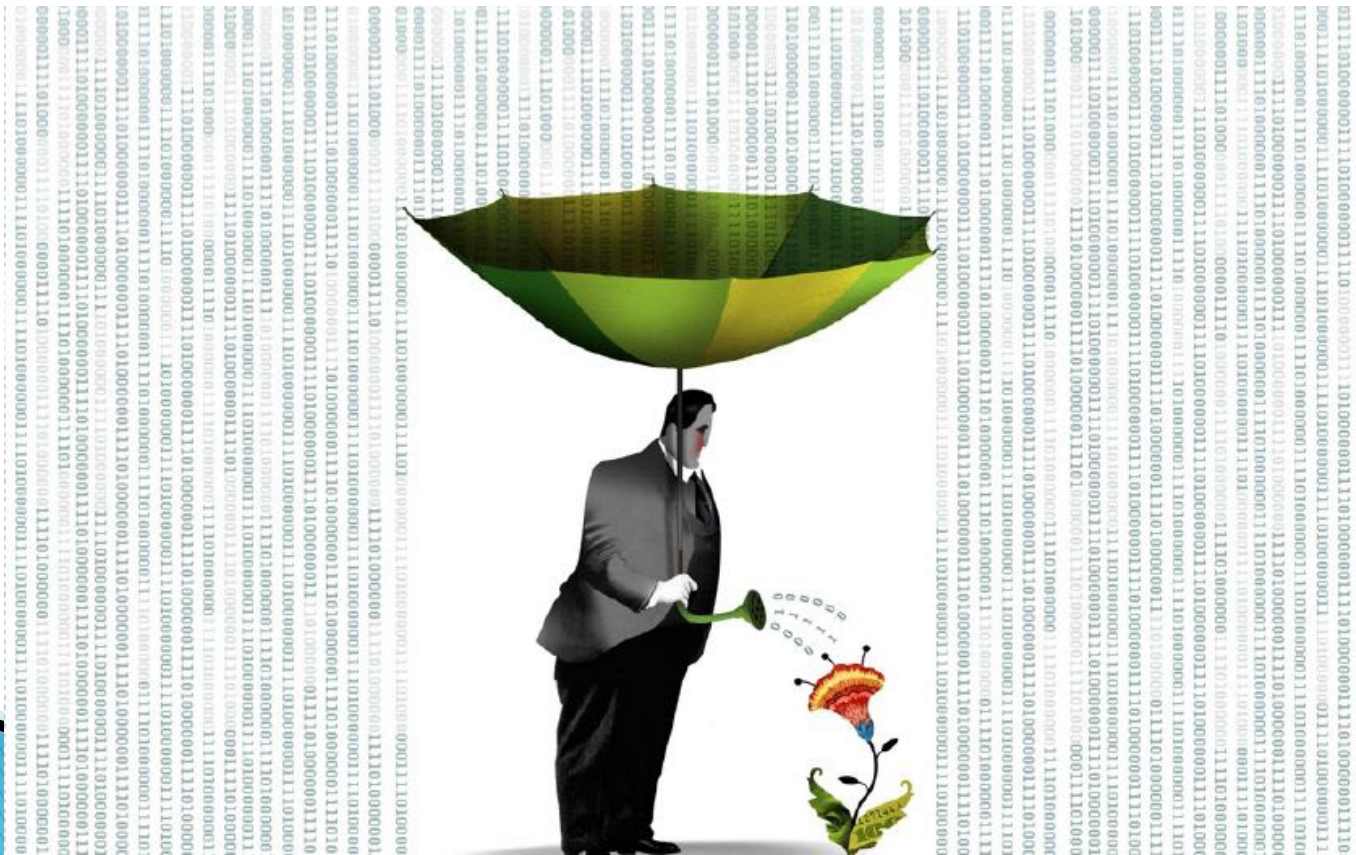
- Data not **discoverable**
- Data not **understandable** (unexplained variables)
- Difficulty to use **diverse** datasets for analysis (unstandardised,..)
- Datasets with **errors**
- No information about **data sharing** and **use policy**
- Data **lost**
- Data not **re-usable**
- No proper **storage** facility

FRUSTRATED !!!

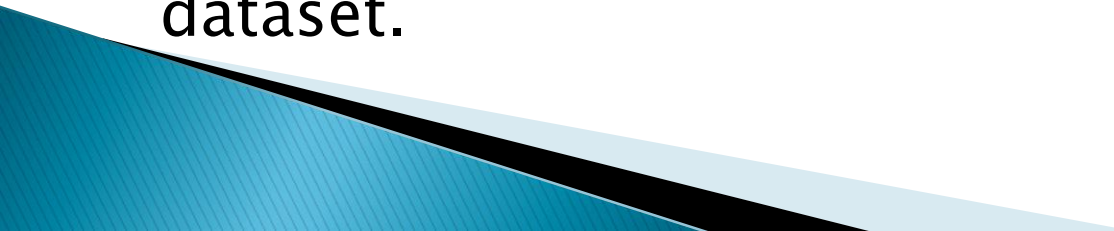


Biodiversity Informatics

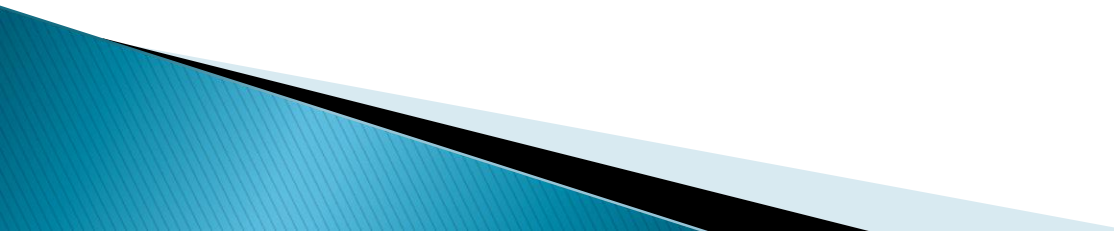
- ▶ Huge Amount, fast change, uncertainty and heterogeneity of data -> **new mechanisms in order to collect, store, access, and reuse these data.**



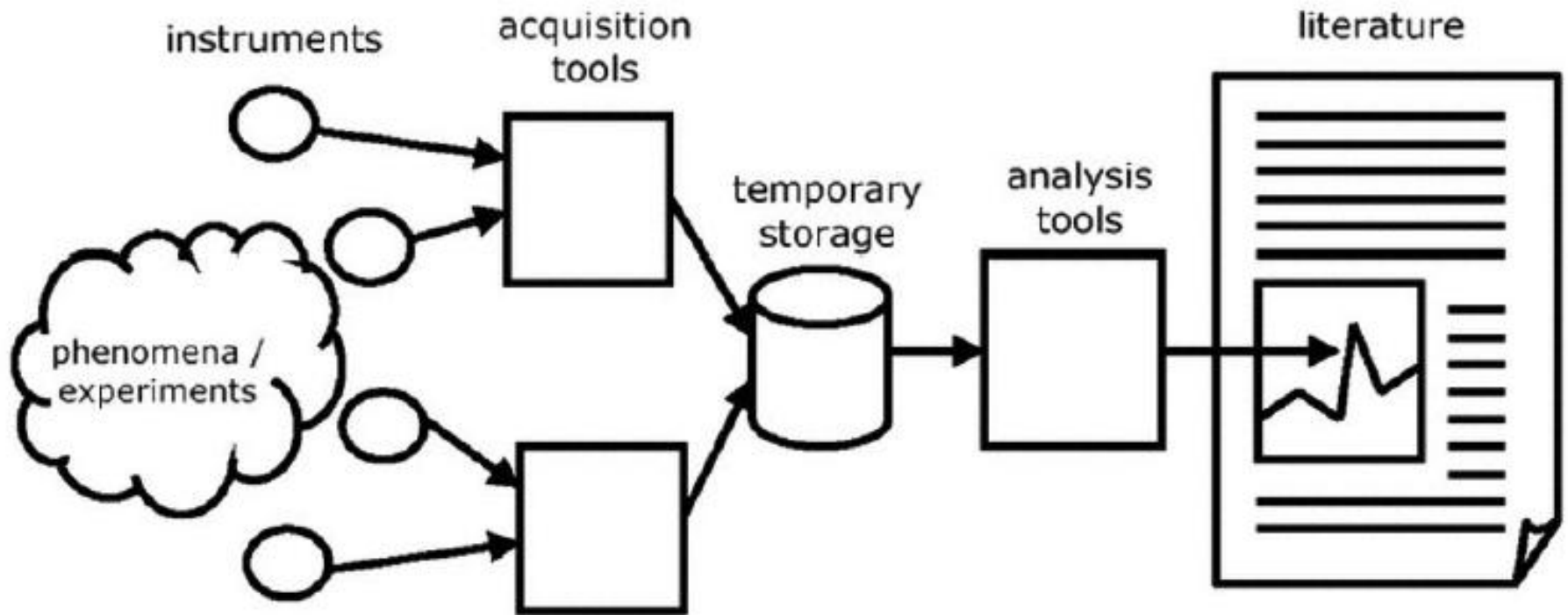
Biodiversity Informatics Objectives

- ▶ The use of information technology (IT) to **support biodiversity research**.
 - ▶ **Organizing knowledge** about individual biological organisms and **the ecological systems** they form.
 - ▶ **Providing access** to the data available on recorded observations for each species.
 - ▶ **Understanding the uncertainties** associated with each dataset.
- 

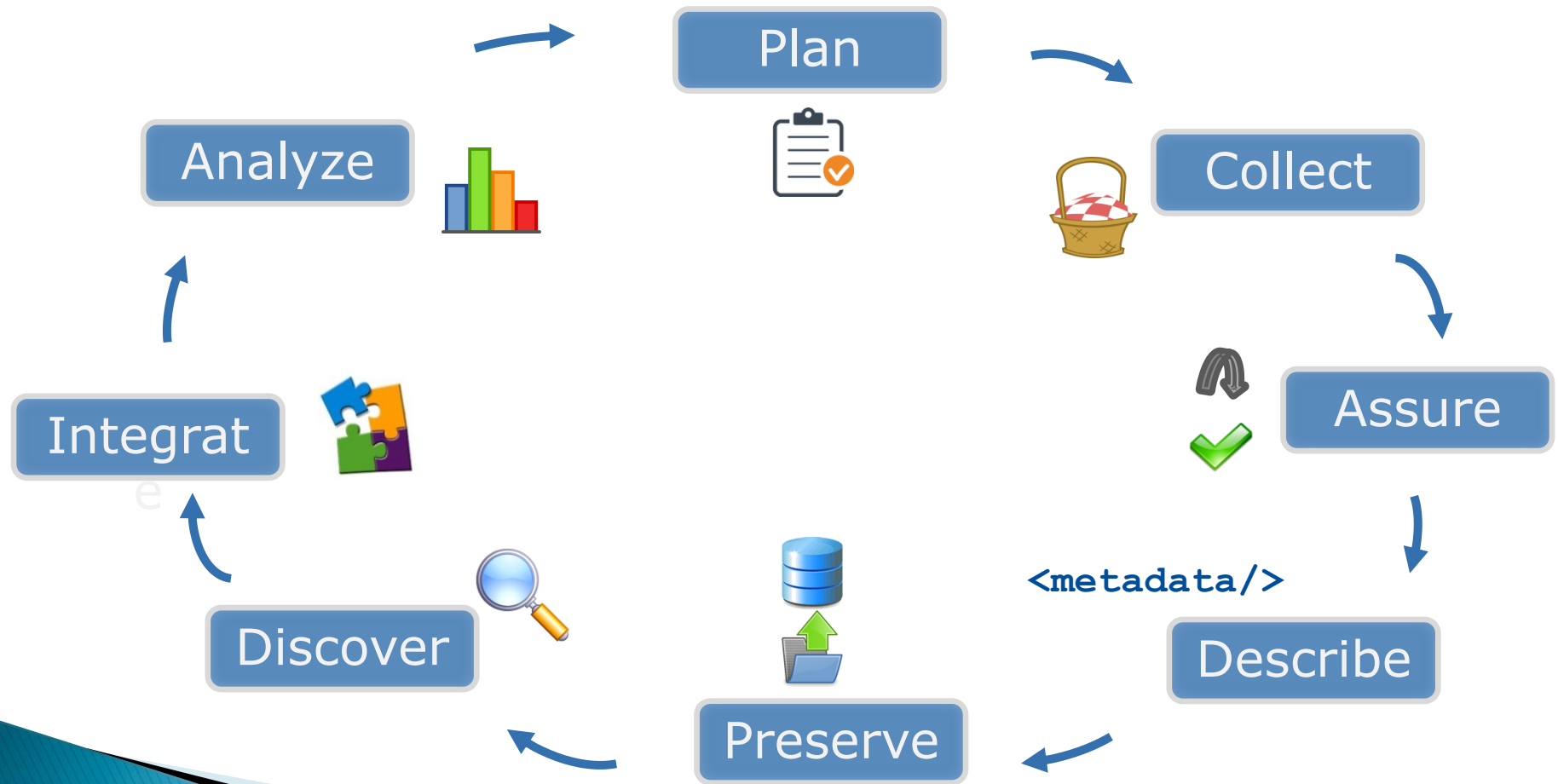
Biodiversity Informatics Example Applications

1. Prediction of distributions of **known and unknown species**.
 2. Prediction of geographic and ecological distribution of **infectious diseases**.
 3. Prediction of **species' invasions**.
 4. Assessment of impacts of **climate change on biodiversity**.
- 

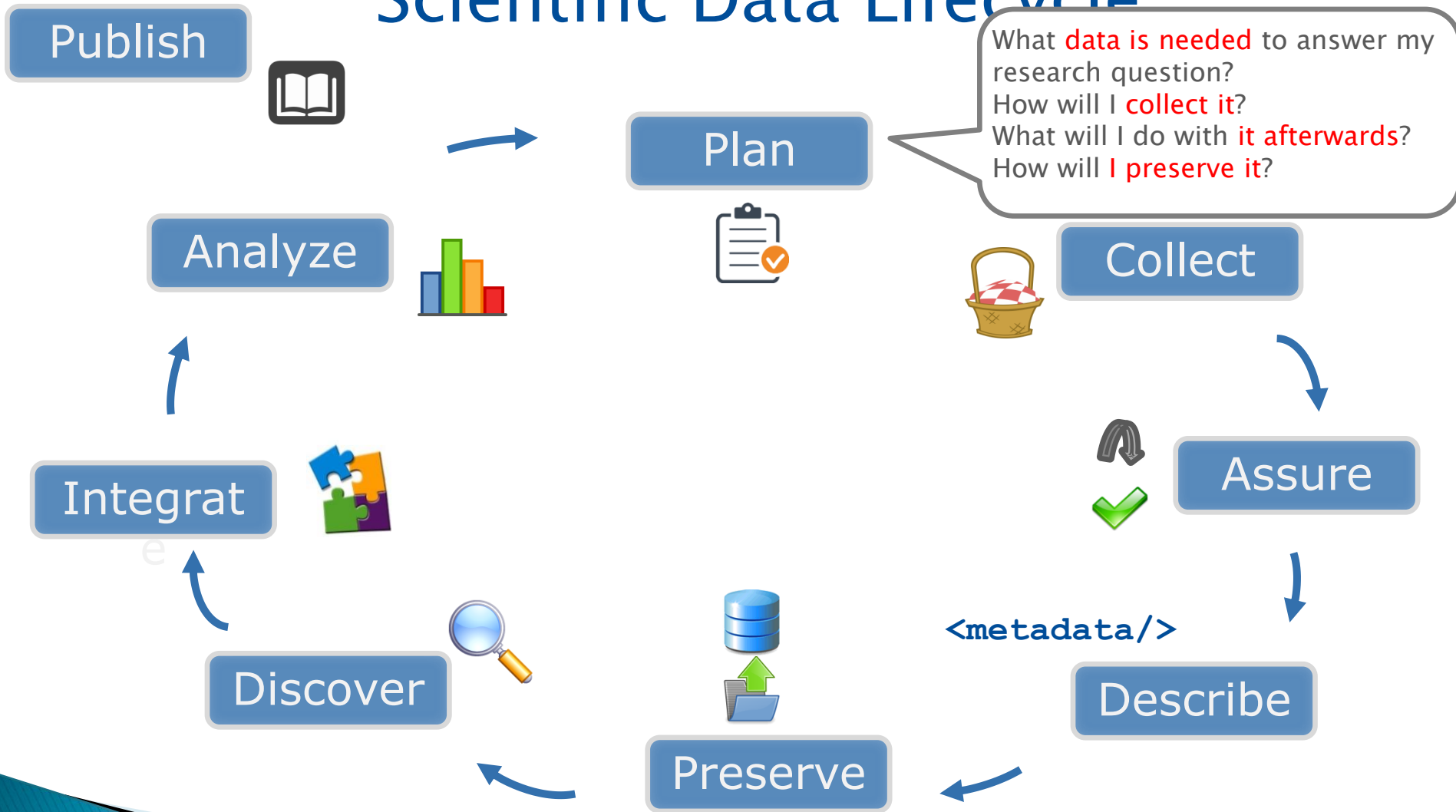
Publish & Forget



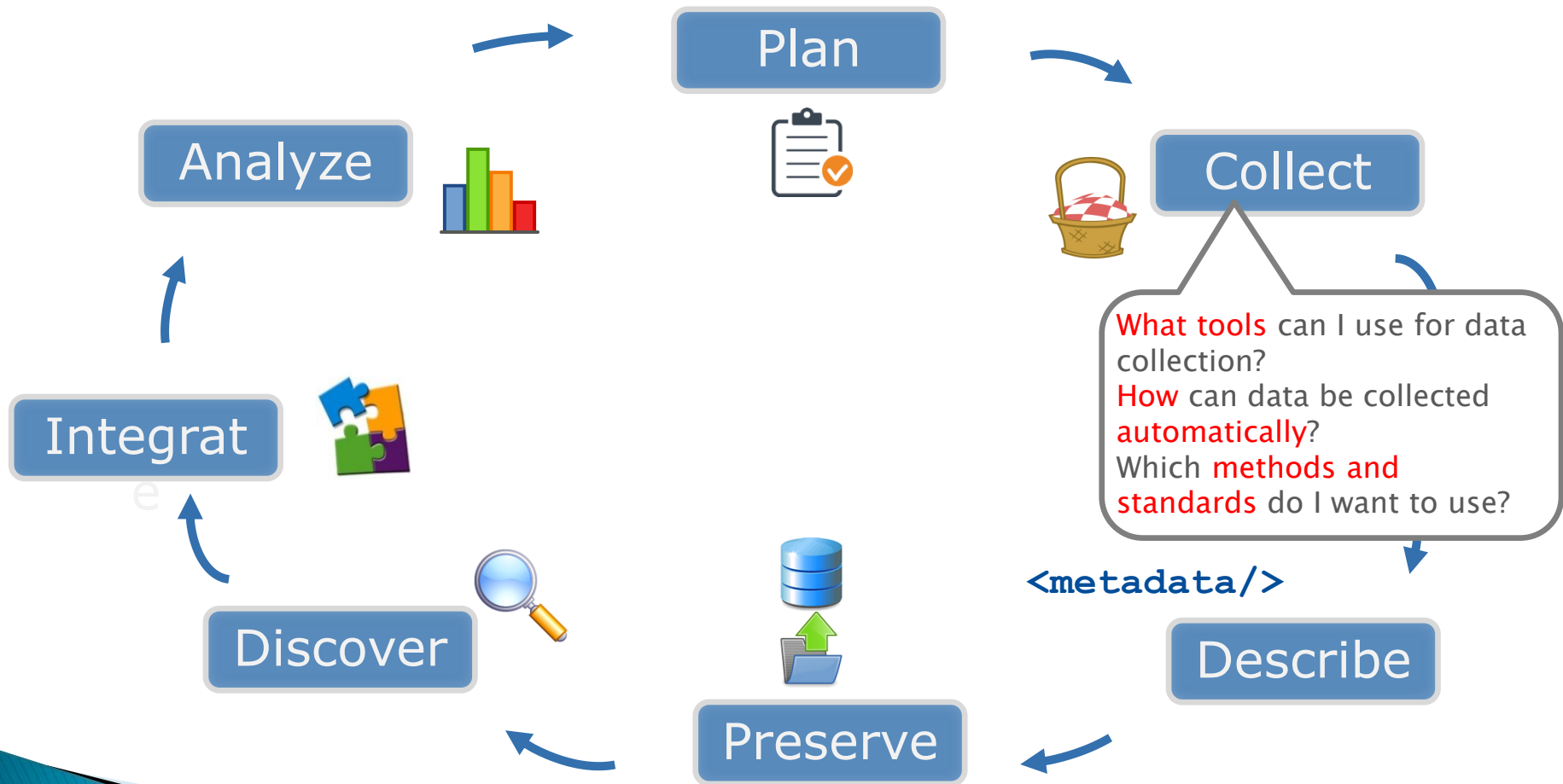
Scientific Data Lifecycle is a conceptual tool which helps to understand the **different steps that data follow from data generation to knowledge creation**



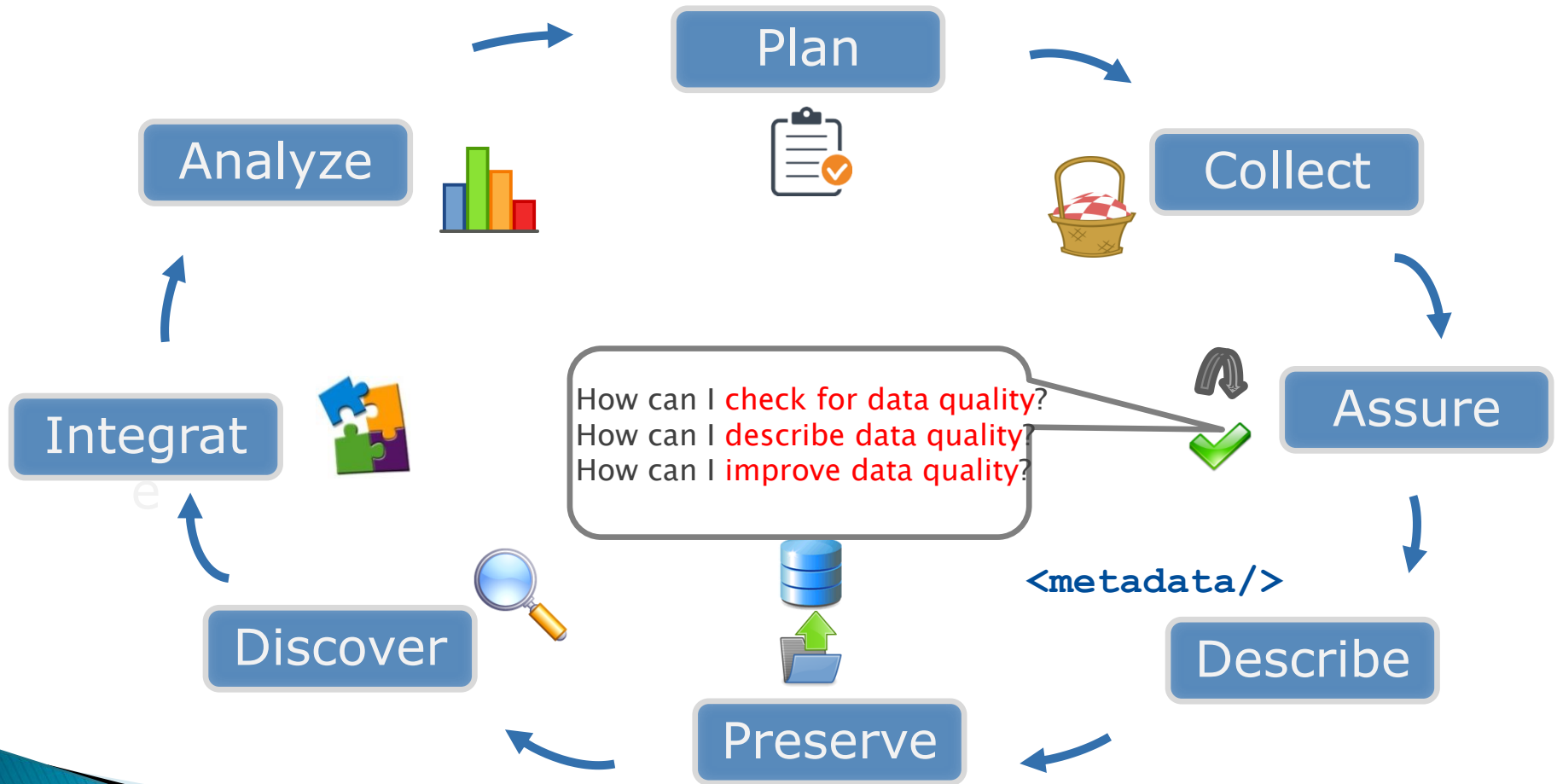
Scientific Data Lifecycle



Scientific Data Lifecycle



Scientific Data Lifecycle



Scientific Data Lifecycle – Assure

- Monitor and maintain **quality**
- **Data cleaning**: correct measurement errors
- ▶ Objective:
**Accessible, accurate, complete, consistent, relevant,
comprehensive, easy to read and interpret**

Scientific Data Lifecycle – Assure

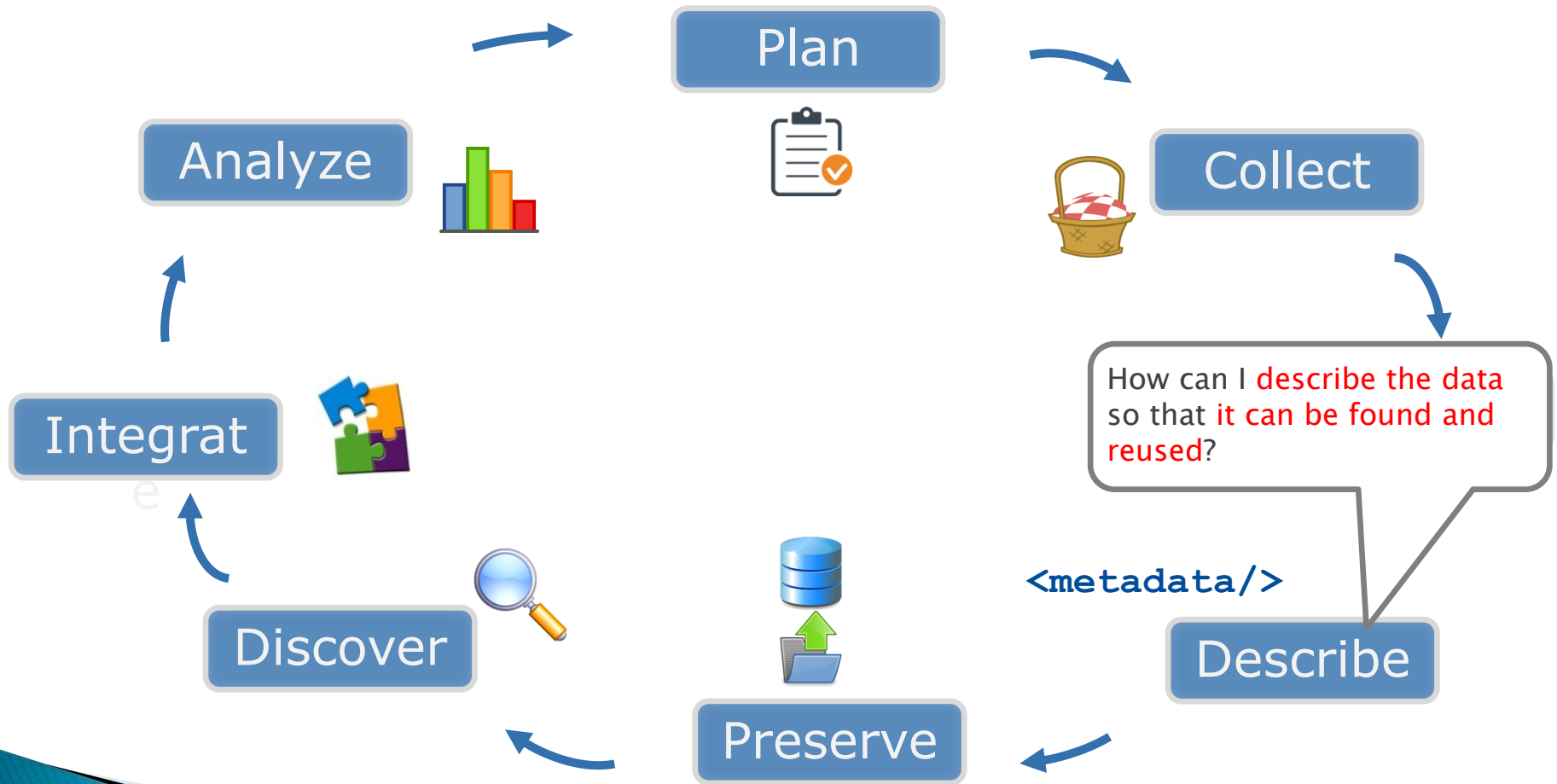
▶ Just Error Examples

The screenshot shows an Excel spreadsheet with two tables. The first table (rows 1-6) has columns A-F and H-N. The second table (rows 12-16) has columns A-F. A callout box on the right lists errors: inconsistent data format, column names, order of columns, different spelling/capitalization, spaces in site names, code used for site names but spelled out for others, and text and numbers in the same column.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Site	Date	Plot	Species	Weight	Acult		Rodent Trapping 3/15/2010						
2	DeepWell	2/13/2010		1 DIPO	12.1	j		Site	Plot	Adult	RodentSp	Weight		
3	Deep Well	Feb-10		2 Pero	13.22	j		DW		1 y	Pero	12		
4	rioSalado	2/13/2010	1a	pero	16	N		RS		2 j	PERO	escaped <15		
5	riuSladu	"	1*	CleGap	18.92	gut away		RS		3 ri	Clegap	91		
6				Mean1	15.06									
7														
8														
9														
10														
11														
12	Rodent Trapping		MJK & ALN	10-Apr-10										
13	Site	Plot	Adult	Species	grams	Comments								
14	deep well		1 y	woodrat	13									
15	riosalado		2 y	PERO	24.5									
16	riosalado		3 y	Clegap	91									
17														
18														
19														
20														

- Inconsistent data format
- Column names
- Order of columns
- Different spelling, capitalization,
- Spaces in site names
- Code used for sites names but spelled out for others
- Text and numbers in same column

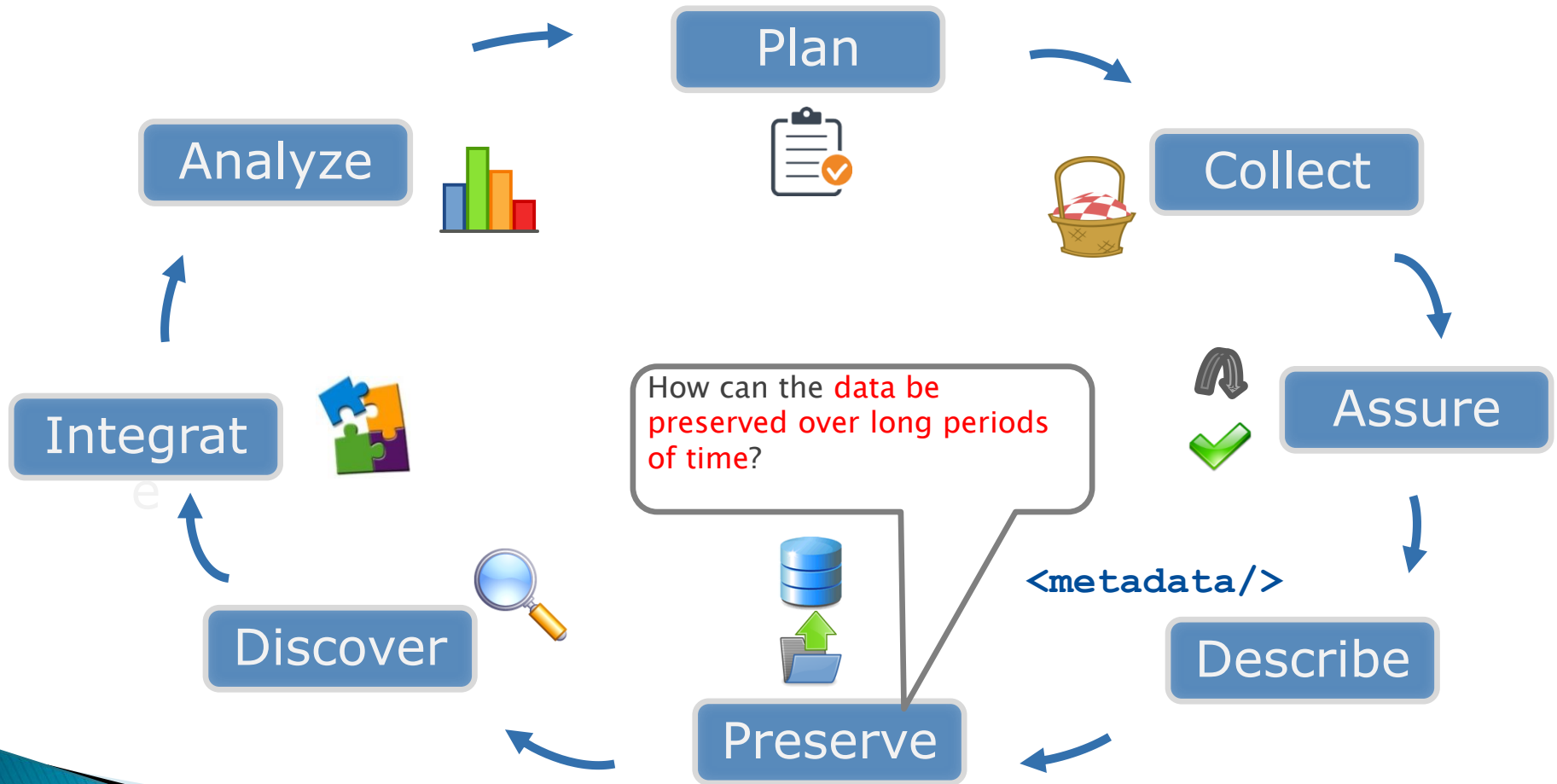
Scientific Data Lifecycle



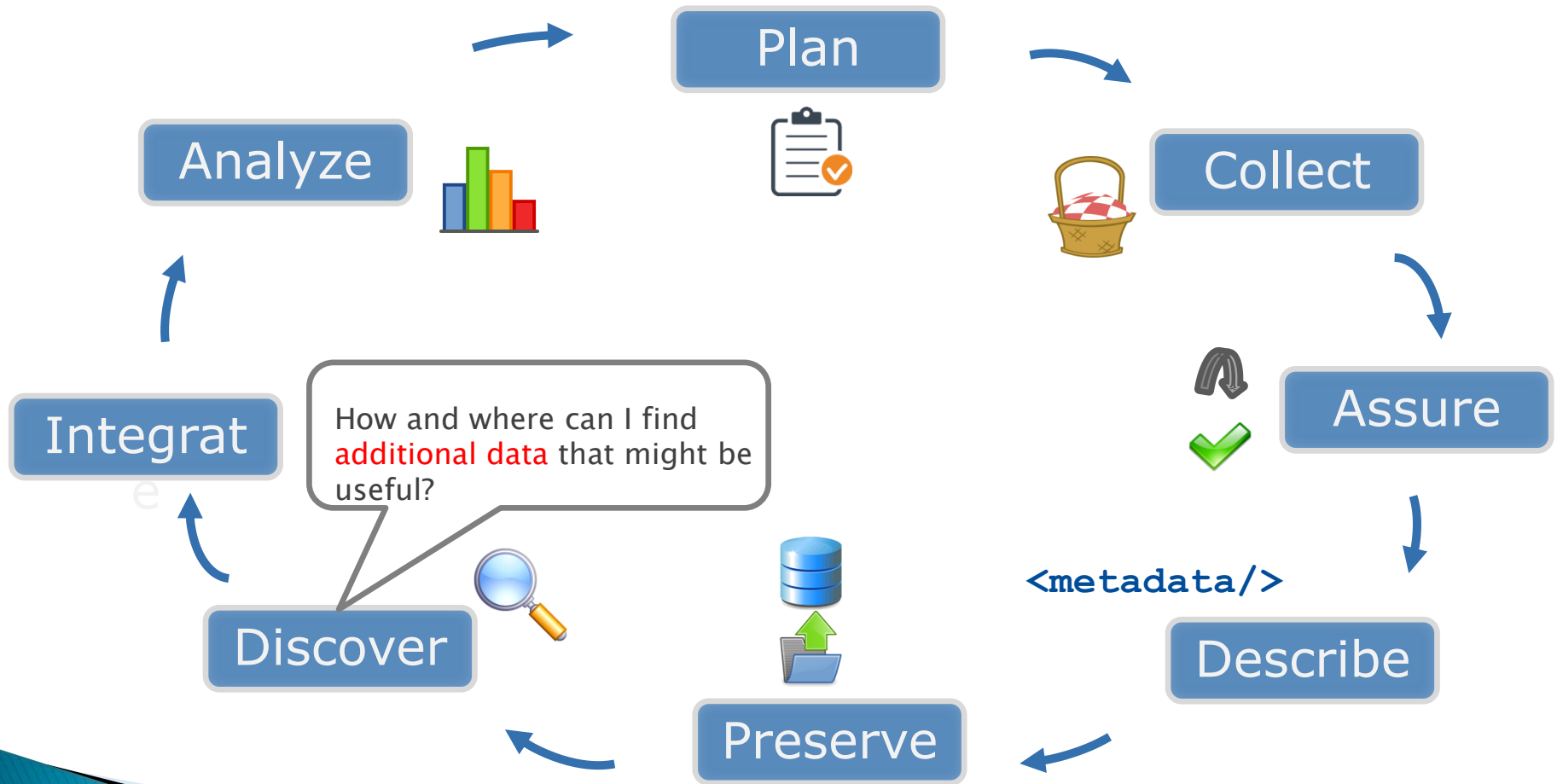
Scientific Data Lifecycle – Describe

- Metadata (**individual file, needs to be created additionally to primary data**) describing information: **What? Who? Where? When? How? (Used for data search)**
- **Standards, technical context:** names of datasets, format, tools, software, methods
- If primary data has considered **naming conventions, consistency of values, etc.**, metadata can be created automatically.

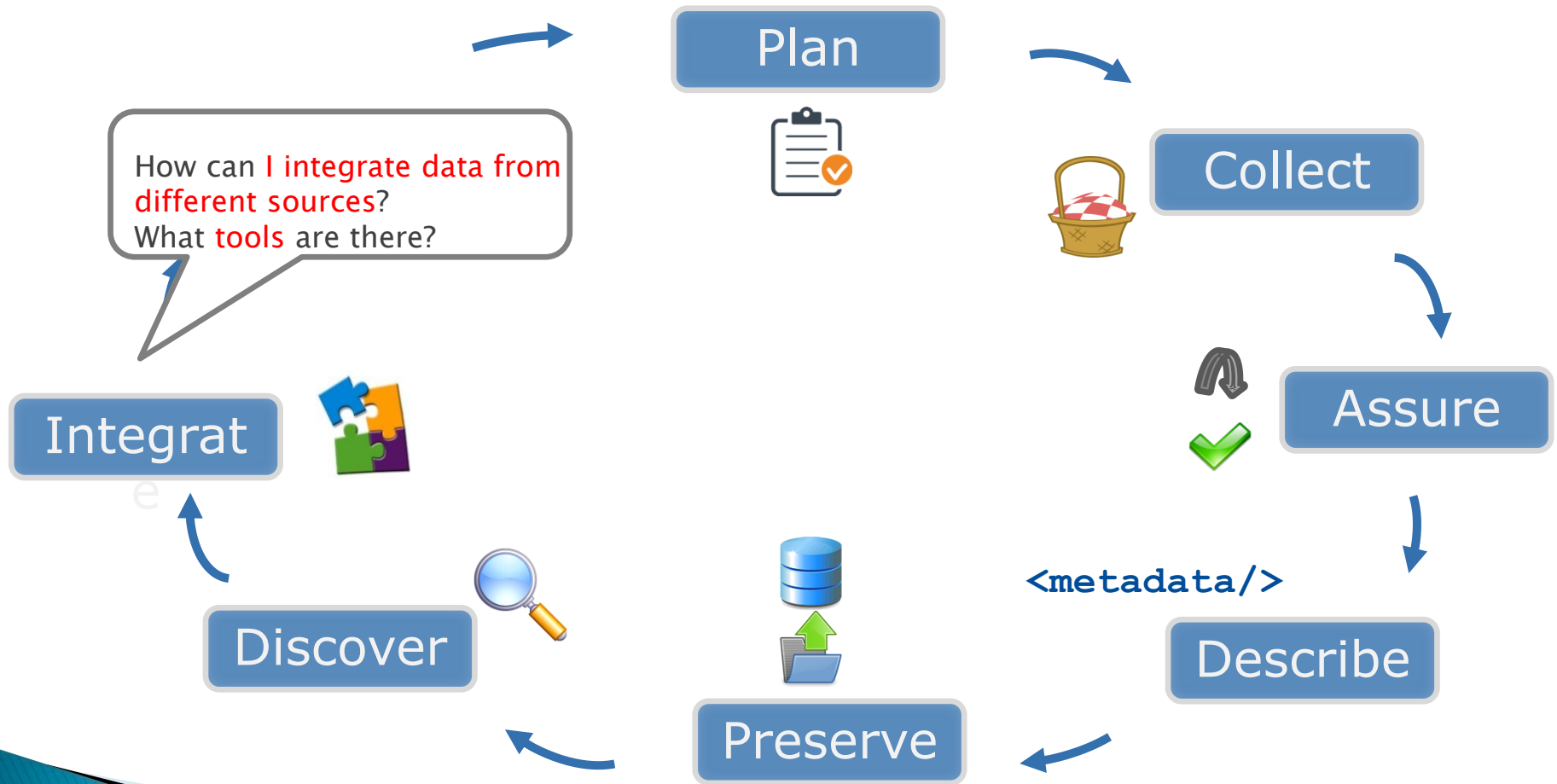
Scientific Data Lifecycle



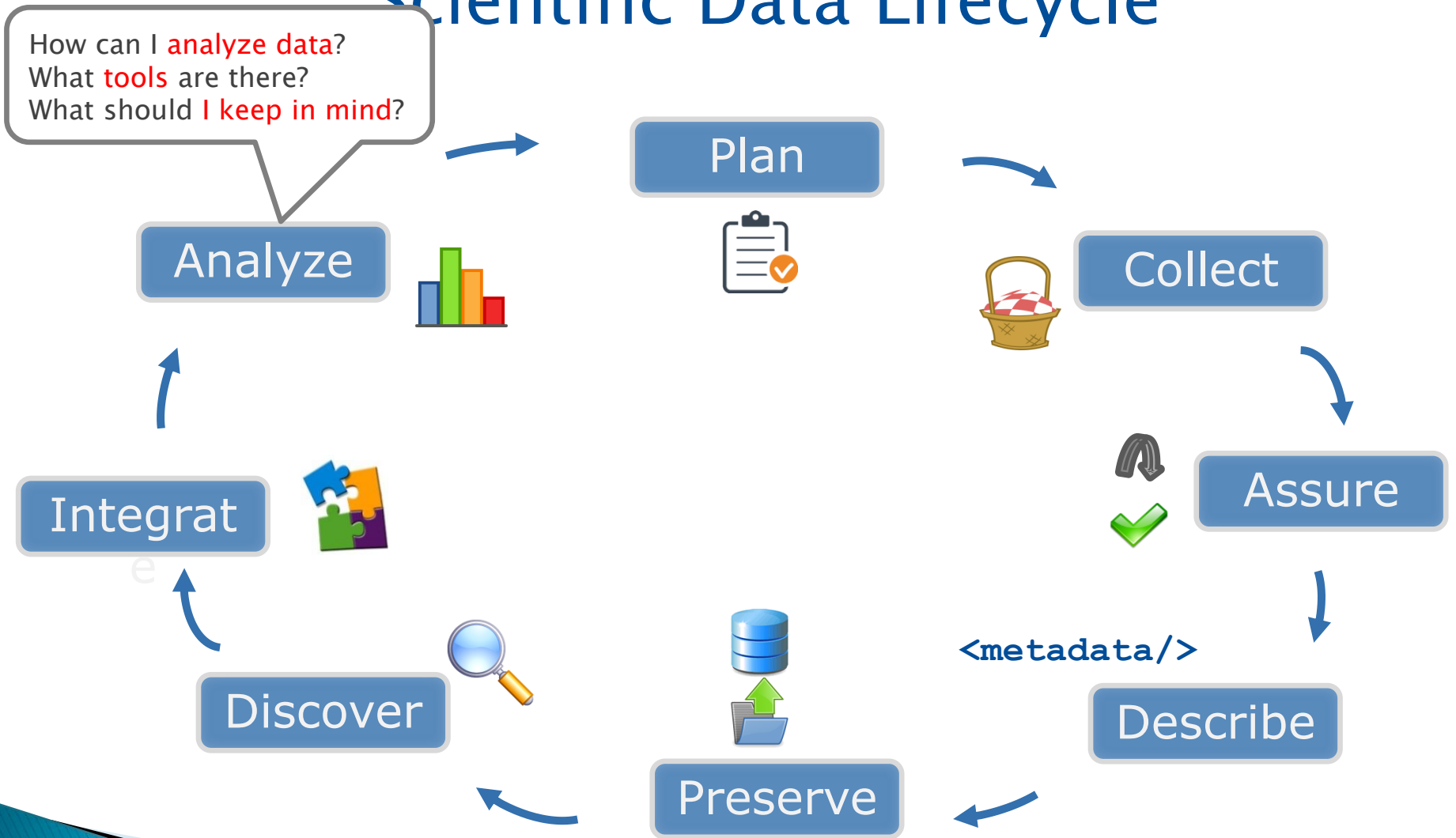
Scientific Data Lifecycle



Scientific Data Lifecycle



Scientific Data Lifecycle



Scientific Data Lifecycle

- **Statistical analysis** through visualization
- Overlay, modelling
- **Analytical technics**: machine learning, statistical tools, text mining, regression analysis
- Tools:





Current Researches



ASBIRG

- ▶ ASBERG stands for Ain Shams Biodiversity Research Group.
- ▶ ASBERG is a fruitful result of the **Biodialog project**.
- ▶ ASBERG is a research group of both **specialized Informatics researchers and Biodiversity scientists**.
- ▶ It aims at **applying informatics techniques** to manage, disseminate, analyze, share, and publish biodiversity data in local context.

	Informatics	Biodiversity	Total
Faculty Members	4	3	7
Postgraduate Students	4	1	5
Undergraduate Students	10	2	12
Total	18	6	24



Informatics Sub-Group



Dr. Ibrahim Moawad



Dr. Rania ElGohary



Dr. Mohammed
Hamdy



Dr. Ahmed Hassan



T.A. Dina Ali



T.A. Ghada
Farouk



T.A. Esraa A.
Hamed



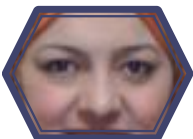
T.A. Mariam
Hesham



Asmaa
Mohammed



Alaa Abd_El baky



Heba Ebrahim



Aya khairy



Thanaa Maher



Ola Farag



Mohamm
ed Khaled



Moataz
samy



Mahmoud
Mohamed Sayed



Biodiversity Sub-Group



Prof. Hesham El-Kassas



Prof. Ahmed Fahmy Abo Doma



Prof. Assoc. Youssef Abdallah



Hussein Omar



Nourhan Atef



Mostafa Mahmud

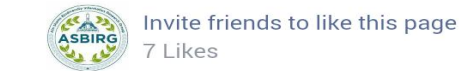


Facebook Page

AinShams Biodiversity Informatics Research Group.

objectives of Facebook are:

- To market the group*
- To publish ASBIRG activities*
- To collaborate with others*
- To share in developing knowledge-based society*
- To prepare for a website for ASBIRG*
- Others....*



AinShams Biodiversity Informatics Research Group (ASBIRG)

See All >





ASBIRG Current Researches

- ▶ Studying soil nematodes in fields irrigated with mixed agricultural drainage water.
- ▶ Studying wheat genetics and field data to discriminate wheat species by applying different knowledge discovery algorithms.
- ▶ Enhancing Scientific Data Management using Semantic Data Mining Approach.
- ▶ From low quality spreadsheets to high quality structured Database.
- ▶ A Personalized Recommender System for Biomedical Ontologies.

