# Bioschemas

## Enabling a FAIR ecosystem for Life Science websites

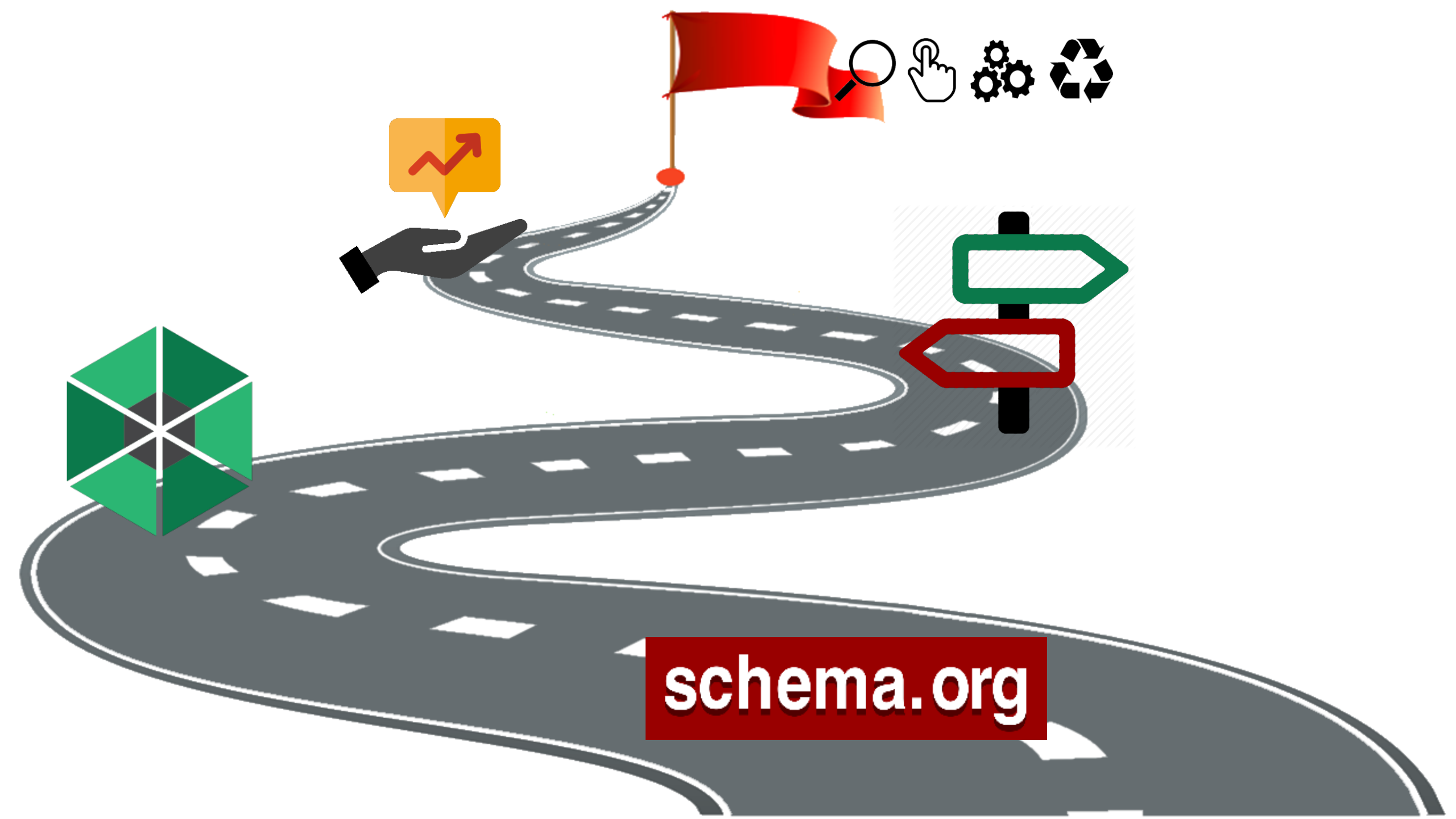
Leyla Jael Garcia Castro (ljgarcia@ebi.ac.uk)

Knowledge and semantic web coordinator

Elixir HUB

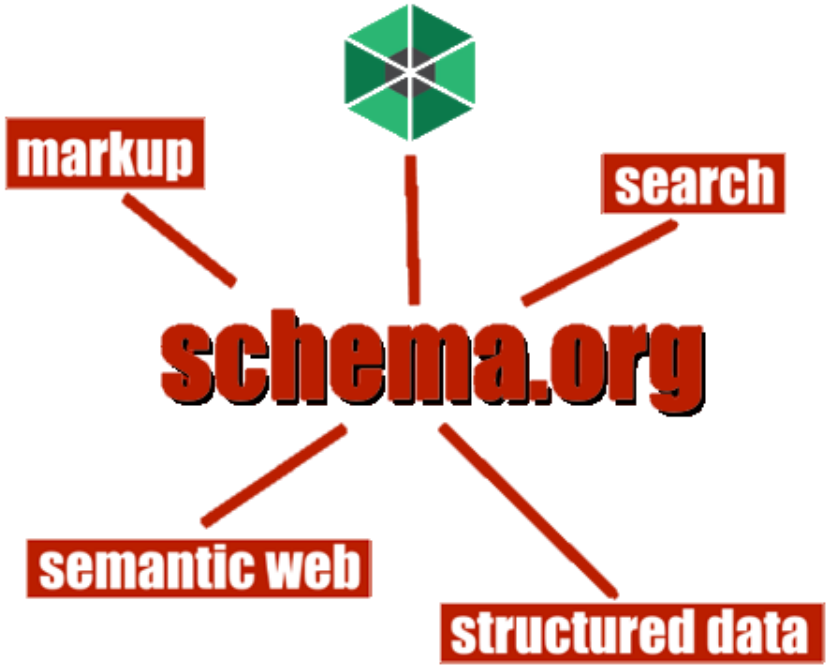
NETTAB

Genoa, 22nd to 24th of October 2018


schema.org

# schema.org

- Collaborative, community activity with a mission to create, maintain, and promote **schemas** for **structured data** on the Internet, on **web pages**, in email messages, and beyond



# Structured data → descriptors

**What we are talking about**

**Types**

Core vocabulary

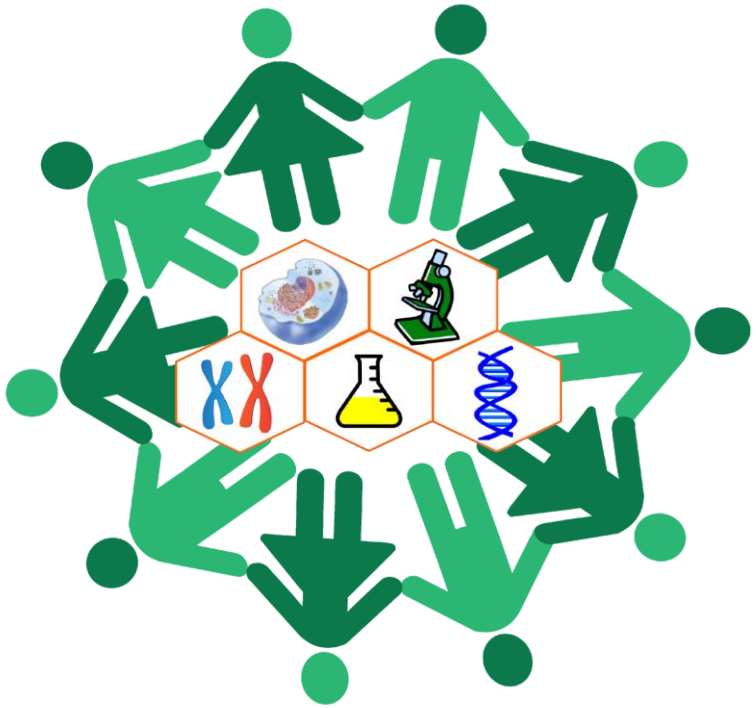
- Thing
  - Action
    - AchieveAction
      - LoseAction
      - TieAction
      - WinAction
    - AssessAction
      - ChooseAction
        - VoteAction
      - IgnoreAction
      - ReactAction
        - AgreeAction
        - DisagreeAction
        - DislikeAction
        - EndorseAction
        - LikeAction
        - WantAction
      - ReviewAction
    - ConsumeAction
      - DrinkAction
      - EatAction
      - InstallAction
      - ListenAction
      - ReadAction
      - UseAction
        - WearAction
      - ViewAction
      - WatchAction
    - ControlAction
      - ActivateAction
      - DeactivateAction
      - ResumeAction

- What we can say about it
- Properties

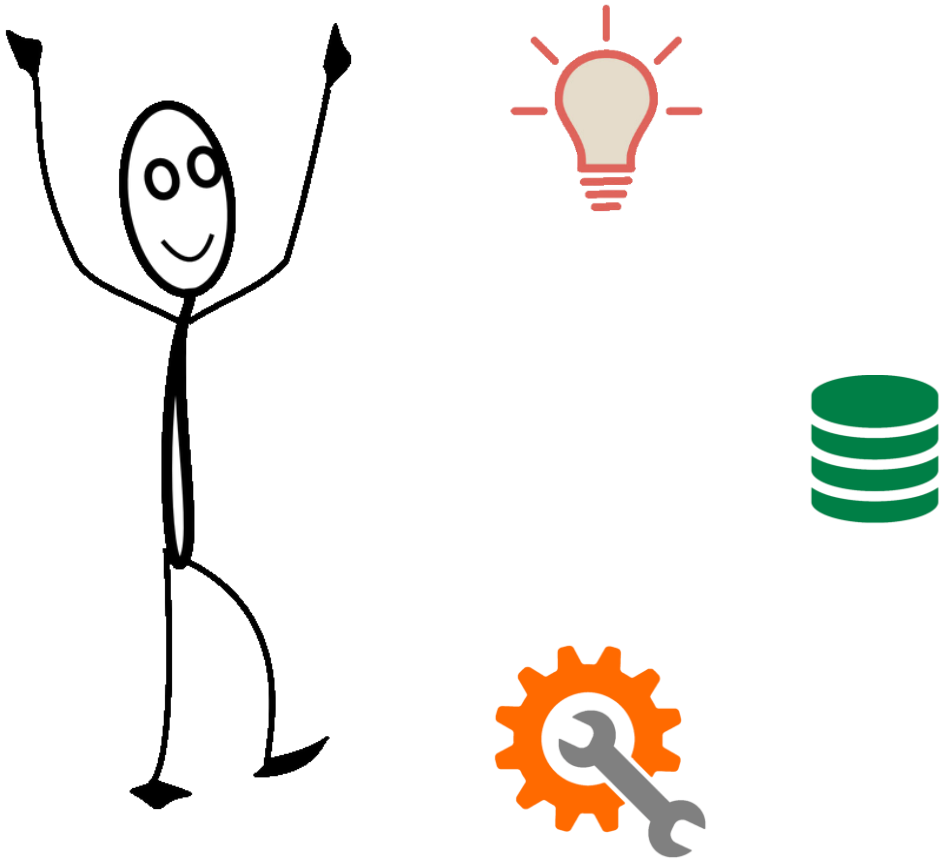
| Property | Expected Type | Description |
|---|---|---|
| **Properties from Person** | | |
| additionalName | Text | An additional name for a Person, can be used for a middle name. |
| address | PostalAddress or Text | Physical address of the item. |
| affiliation | Organization | An organization that this person is affiliated with. For example, a school/university, a club, or a team. |
| alumniOf | EducationalOrganization or Organization | An organization that the person is an alumni of. Inverse property: alumni. |
| award | Text | An award won by or for this item. Supersedes awards. |
| birthDate | Date | Date of birth. |
| birthPlace | Place | The place where the person was born. |
| brand | Brand or Organization | The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person. |
| children | Person | A child of the person. |
| colleague | Person or URL | A colleague of the person. Supersedes colleagues. |
| contactPoint | ContactPoint | A contact point for a person or organization. Supersedes contactPoints. |
| deathDate | Date | Date of death. |
| deathPlace | Place | The place where the person died. |
| duns | Text | The Dun & Bradstreet DUNS number for identifying an organization or business person. |
| email | Text | Email address. |
| familyName | Text | Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property. |
| faxNumber | Text | The fax number. |
| follows | Person | The most generic uni-directional social relation. |
| | Organization or | A person or organization that supports (sponsors) something |

# Bioschemas

- Community initiative built on top of **schema.org**
- Aim
  - Improve data discoverability and interoperability in Life Sciences
- How
  - Adding Life Science types to schema.org
  - Providing usage guidelines, examples and tools

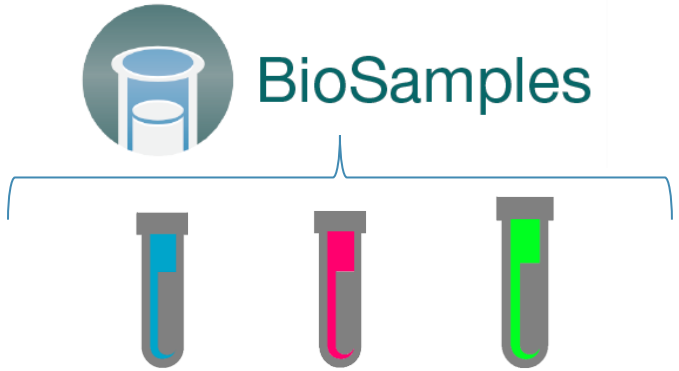


# Motivation



- Researcher looking for help to
  - sequence a cactus → people, expertise
  - find similar sequences → datasets
  - compare and visualize → tools

# Use cases




- Generic
  - Search engines
  - Google dataset specialized search
- Life sciences
  - Data depositions can harvest mark up from small providers → small providers can benefit back from data depositions
  - Tess (training and events registry) → Automatic population → trainers add mark up
  - Common/rare terms → Resource index

# Guess the person

Name: Albert

Last name: Einstein

Birth date: 14 March 1879

Birth place: Ulm, Kingdom of Württemberg, German Empire

Height: 1.72 cm

Alumni of: University of Zurich

Weight: unknown

Death date: 18 April 1955

Work location: 1 Einstein Drive Princeton, New Jersey 08540 USA

Know about: Physics, relativity, mass energy equivalence

Honorific prefix: Dr. Prof.

Knows language: German, English

Award: Nobel prize in physics

Siblings: Maria

isicv4: 7210

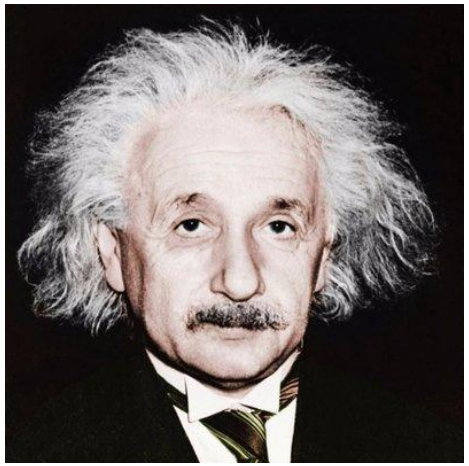
# Guess the person

Name

Last name

Birth date: 14 March 1879

Birth place: Ulm, Kingdom of Württemberg, German Empire



Height: 1.72 cm

Alumni of: University of Zurich

Weight: unknown

Death date: 18 April 1955

Work location: 1 Einstein Drive Princeton, New Jersey 08540 USA

Know about: Physics, relativity, mass energy equivalence

email


Honorific prefix: Dr. Prof.

Knows language: German, English

Award: Nobel prize in physics

orcid

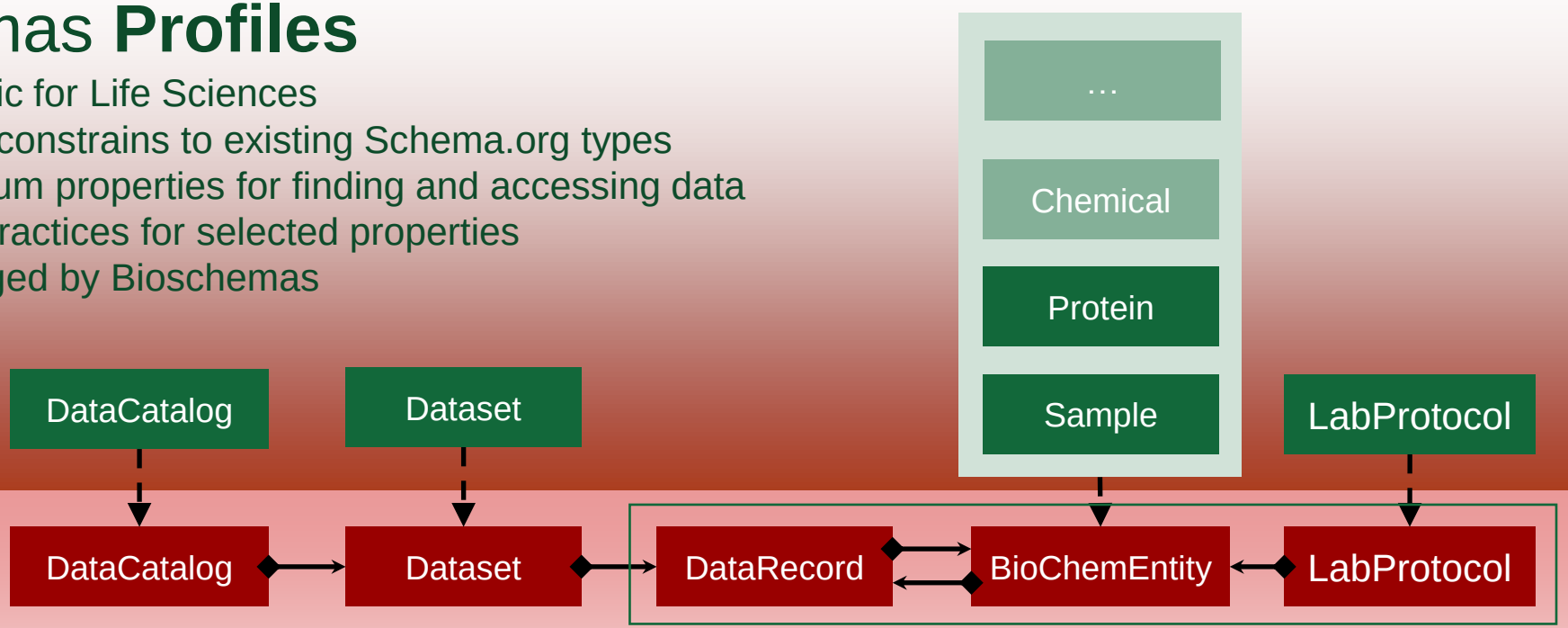
Siblings: Maria

isicv4: 7210

# Bioschemas vs schema.org

## Bioschemas **Profiles**

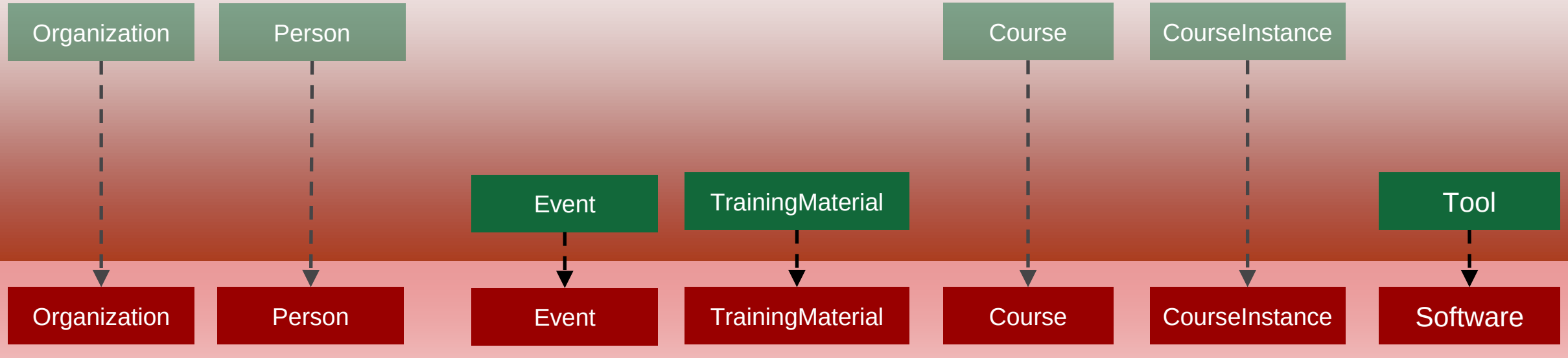
- Specific for Life Sciences
- Apply constrains to existing Schema.org types
- Minimum properties for finding and accessing data
- Best practices for selected properties
- Managed by Bioschemas




## Schema.org **Types**

- Generic data model
- Generous list of properties to describe data types
- Managed by Schema.org

# Beyond databases → digital resources



# Why Bioschemas?

- Focus on key properties prioritized as Minimum, Recommended and Optional based on community agreements and common practices

- Additional recommendations regarding properties cardinality

- Customization on schema.org types to better supports needs on the life sciences community

- Terms reused from well-known ontologies thus avoiding reinventing the wheel

# Why structured data with Bioschemas?


# How structured data contributes to FAIRability

**Findability**

Finding proteins, samples, phenotypes and so exposed via web pages

**Interoperability**

Gathering data from different life sciences resources following a common format, linking to each other

**Accessibility**

Enabling data extraction over web pages

**Reusability**

Making reuse rules explicit and providing provenance information

Findable

Accessible

Interoperable

Reusable

- Unique identifiers
- Descriptive metadata
- Indexed and available

**F**indable

**A**ccessible

- HTTP
- Authorization via login or IP

**I**nteroperable

- Schema.org →JSON-LD
- Links to other resources

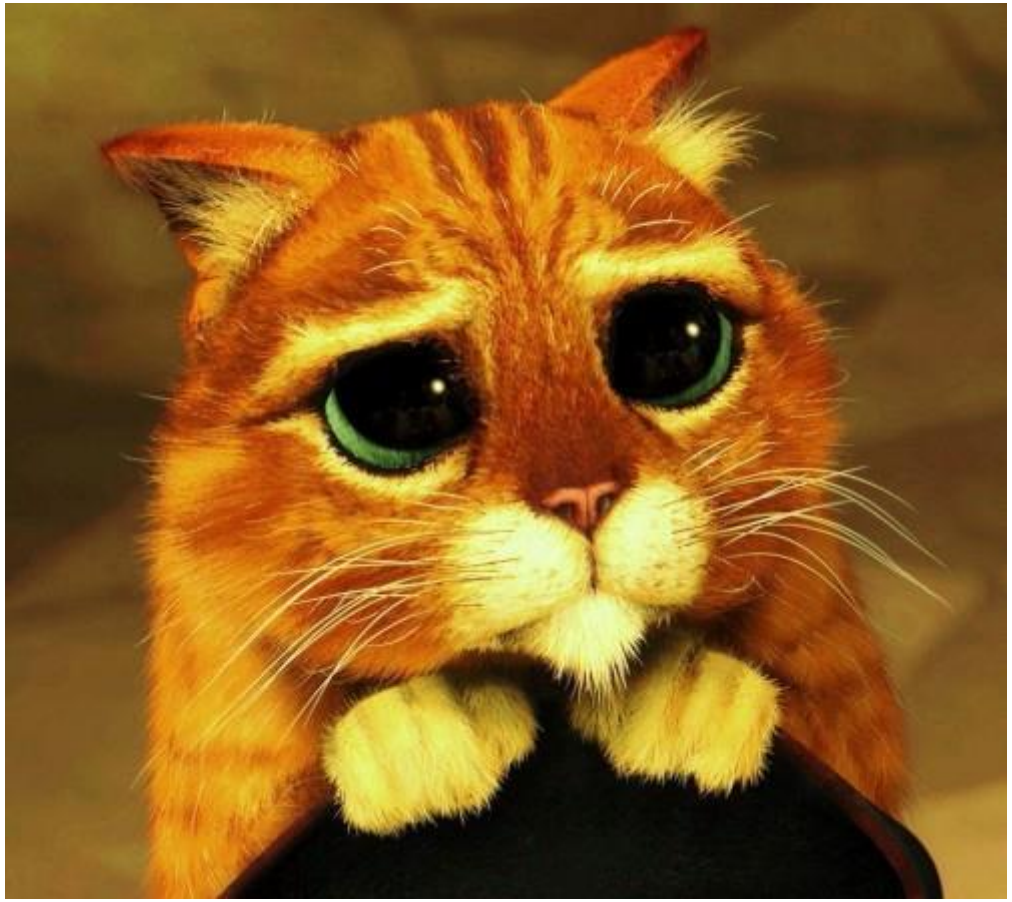
**R**eusable

- License
- Provenance

# Be part of the game!


Bioschemas

# Beyond Life Sciences



Bioschemas is for Life Sciences

BUT

The same idea can be applied to other sciences

Research schema

# Acknowledgements


- Stephen Anyango
- Ricardo Arcila
- Haydee Artaza
- Terri Atwood
- Phil Barker
- Dominique Batista
- Michael Baudis
- Niall Beard
- Cath Brooksbank
- Tony Burdett
- Guillermo Calderon Mantilla
- Ethy Cannon
- Denise Carvalho-Silva
- Luca Cherubin
- Justin Clark-Casey
- Martin Cook
- Manuel Corpas
- Michael R Crusoe
- Pavel Dallakian
- Luc Deltombe
- Victoria Dominguez
- Michel Dumontier
- Stephen Ficklin
- Robert Finn
- Alexander Garcia
- Leyla Garcia
- Carole Goble
- Alejandra Gonzalez-Beltran
- Alasdair Gray
- Matthew Green
- Jeffrey Grethe
- Henning Hermjakob
- Petr Holb
- Richard Holland
- Carlos Horro
- Jon Ison
- Christa Janko
- Andy Jenkinson
- Rafael C Jimenez
- Claire Johnson
- Simon Jupp
- Nick Juty
- Lee Larcombe
- Nicolas Le Novère
- Mikael Linden
- Audald Lloret
- Ankit Kumar Lohani
- Aurélien Luciani
- Federico López Gómez
- Ronald Margolis
- Maria Martin
- Michaela Th. Mayrhofer
- Kenneth McLeod
- Peter McQuilton
- Franck Michel
- Gos Micklem
- Saqib Mir
- Sarah Morgan
- Chris Mungall
- Heimo Müller
- Aleksandra Nenadic
- Michał Nowotka
- Helen Parkinson
- Roberto Preste
- Giuseppe Profiti
- Anders Riutta
- Philippe Rocca-Serra
- Gabriella Rustici
- Gustavo Salazar-Orejuela
- Susanna A Sansone
- Vicky Schneider
- Serena Scollen
- Kaisa Silander
- Morris Swertz
- Mohameth François Sy
- Chris Taylor
- Milo Thurston
- Dan Timmons
- John Van Horn
- Susheel Varma
- Sameer Velankar
- Premysl Velek
- Andra Waagmeester
- Liz Williams
- Sarala Wimalaratne
- Anil Wipat
- Olga Ximena Giraldo
- Gianluigi Zanetti
- Anita de Waard
- David van Enckevort
- Peter van Heusden

# Thank you

http://bioschemas.org/ | @bioschemas | https://github.com/bioschemas/