Document to brainstorm and add anything useful that help us during the process to create the Bioschemas specification for data repositories

# Documents

* [Use case spreadsheet](https://docs.google.com/spreadsheets/d/1XzrZxFIuG3TS9RU8vACoUjAvaADLmI_FrIk7O3BEkxY/edit?usp=sharing)
* [Mapping spreadsheet](https://docs.google.com/spreadsheets/d/1H12h5VpVNJFzNs2RQJWjXkauCEn3qEsVFzKRoiHHffY/edit?usp=sharing)
  + [Schema.org data catalog](http://schema.org/DataCatalog)
* [Hennings presentation about data repositories](https://docs.google.com/presentation/d/1b4o1STsic69y3EIDe1ksiwqM4-0sfTTqSChdXpFT7Hk/edit?usp=sharing)
* Old data repositories specifications
  + [Examples sent to Nick and Sarala](https://docs.google.com/document/d/1Fwy4Vj2O-kvC3gSWE4FjKTxr982-fo7eOE5IqEWx4TE/edit?usp=sharing)
  + [Simplified version by John and Sarala](https://docs.google.com/document/d/19HjZymfI8UyPt7QAgMXqIaTHj04OrCoN6D1H7YK3JvM/edit?usp=sharing)
* [PrefixCommons data repositories mapping](https://docs.google.com/spreadsheets/d/1Ziou1jKqBF2lOQAY3CP2cwoI7YZYC0CDhKnmRw9aS6E/edit?usp=sharing)

# Todo and actions

* Sarala - Look for the right properties from schema.org mapping to identifiers properties from use case.
  + Try to reuse as much as possible existing prop. Eg. [PropertyValue](http://schema.org/PropertyValue)
  + We might need to propose new properties.
* ???? - Look at what schema.org properties to use to describe release information
* We need guideline for keywords (topics + taxon + database\_type)
* Consult with Schema.org potential new properties
  + Referencing the data repository tools and interfaces
    - Properties
      * O - access type (ftp, rest etc.) R - access type location (URL) R - access change date O - version R- Name
      * O - Tools:name,url
    - Comments
      * Could we use citation??
  + M - release date R - release version R - release type (enumeration: continuous, discontinuous)
  + O - Metric:name,value

# Process

1. Define use case
2. Metadata crosswalk and mapping to schema.org
   1. Metadata providers
   2. Metadata registries
   3. Standards defining metadata
3. Bioschemas specification
   1. Define minimum properties based on “finding” use cases
   2. Define cardinality and suggested controlled vocabularies
   3. Align with existing best practices
4. Test with crosswalk existing entries
5. Adoption by data repositories and registries
6. Applications

# Important things to remember

* Align existing recommendation to our spec
  + Use of identifiers (Force11 recommendations [compact identifiers], 21st century identifiers recommendation, ...)
  + How to reference standards (formats, e.g. use BioSharing identifiers) and publications (e.g. use DOI)
  + Indicators recommendation from the data platform
* DataCatalogue to show datasets
  + Datasets
    - Different categories. Eg Uniprot
      * Predicted vs curated
      * Species: Human, Mouse, …
      * Release: Release 1, release 2, release 3
    - Properties to show

|  |
| --- |
| {  "@context": "http://schema.org/",  "@type": "DataCatalog",  "**dataset**": [  {  "@type": "**Dataset**",  "name": "X"  "url": "..."  },  {  "@type": "**Dataset**",  "name": "Y"  "url": "..."  }  ]  } |

# Other references