



RDA Global Adoption week

15 - 19 June 2020





Objective

Originally planned for Plenary 15, the RDA Adoption week aims to **demonstrate the wide variety of RDA adoptable and adopted solutions to data sharing challenges** that people in the field encounter in their daily jobs.

Purpose of the week:

- Learn about RDA Outputs
- Converse with speakers from all around the world who have created and implemented them
- Determine how best to integrate those data sharing solutions into your own projects



Tuesday 16th June

14:00

Data Description

Documentation is an essential component of research data management & allows researchers to make sense of data in the future.

- **[RDA/TDWG Attribution Metadata Working Group: Final Recommendations](#)** - Anne Thessen (Oregon State University)
- **[Research Data Repository Interoperability WG Final Recommendations](#)** - Thomas Jejkal (KIT)
- **[Wheat Data Interoperability Recommendations](#)**. A Model Of Data Integration Related To Wheat Genetic Resources And Resistance To Fusarium Head Blight - Dimitar Vassilev

Recommendations & Outputs Catalogue short-link: <https://bit.ly/2YlXxnZ>





Poll

Go to sli.do

Enter Code: #W343



Our Speakers

Anne Thessen

Assistant Prof. (Senior Research), Semantic Technology

Oregon State University

Co-chair RDA / TDWG Metadata Standards for attribution of physical and digital collections
stewardship

Thomas Jejkal

Data Manager, Karlsruhe Institute of Technology (KIT), Germany

Steinbuch Centre for Computing (SCC)

Co-chair RDA Research Data Repository Interoperability WG

Dimitar Vassilev

Bioinformatics group leader

Faculty of Mathematics and Informatics

Sofia University "St Kliment Ohridski"



Recommendations and outputs catalogue

- RDA Outputs are classified as **RDA Recommendations** (*official, endorsed results of RDA Groups*), **Supporting Outputs** (*useful solutions from our RDA Working and Interest Groups*) or **other Outputs**
- They can be searched according to their status, **Data Life Cycle topics** or scientific domain

1 Data Management
Develop best practices and tools for non-static, machine-readable data management plans which can evolve throughout the research data lifecycle, as well as be machine-readable by collaborators and stored with the data.

2 Data Collection
Capture scientific evidence that allows analysis to lead to the formulation of convincing and credible answers to questions of researchers.

3 Data Description
Documentation is an essential component of research data management and allows researchers to make sense of data in the future.

4 Identity, Store and Preserve
Collections and streams of digital objects are growing at an incredibly rapid pace. We need to understand on what these objects are, how we need to be documenting and storing them, and how they should all link and talk to one another before we get overtaken by a forest of data.

5 Disseminate, Link and Find
An increasing number of publishers and journals are implementing policies that require or recommend published articles to be accompanied by the underlying research data.

6 Policy, Legal Compliance and Capacity
Currently, all scientific communities use their own set of policies, if any. A generic set of policies that can be revised and adapted by users to build up their own data collection does not exist.

rd-alliance.org/recommendations-and-outputs/catalogue



Tell your adoption story

- **Are you an adopter?** RDA is actively seeking new adoption stories to inspire the further uptake of RDA outputs.
- **Submit your story here:**
<https://www.rd-alliance.org/tell-your-rda-adoption-story>

RDA ADOPTION STORIES



Adopters of RDA outputs share their experiences and lessons learned to inspire further uptake of RDA outputs

🔗 Read the current adoption stories

📄 Submit your story through the webform

[rd-alliance.org/tell-your-rda-adoption-story](https://www.rd-alliance.org/tell-your-rda-adoption-story)





CODATA CfP Data Science Journal

- **RDA special collection themes:**
 - Results produced by an IG or WG;
 - Description of an Adoption Case outlining how a specific recommendation or output has been implemented;
 - Other types of work related to RDA activities.
- RDA Europe 4.0 still has funds available for the publication of articles in DSJ
- Open to all interested applicants regardless of their geographical provenance.
- **Deadline 17 July**

Submit your article for the Data Science Journal Special Collection on RDA

RDA CODATA Data Science Journal special collection solicits high quality papers describing the latest results of RDA WG and IG that have recently published outputs and associated use cases.

Publication fees will be covered by the RDA Europe 4.0 project

Publication fees of the first selected 30 articles will be covered by the RDA Europe 4.0 project thanks to specific funding available until 17 July 2020 on a first com first served basis.

Don't miss out, submit your paper now!
datascience.codata.org/about/submissions

A QR code located at the bottom right of the poster, linking to the submission page.

The RDA logo is positioned at the bottom center of the poster.



WWW.RD-ALLIANCE.ORG/
[@RESDATALL](https://twitter.com/RESDATALL)



RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn - www.linkedin.com/in/ResearchDataAlliance

Slideshare - <http://www.slideshare.net/ResearchDataAlliance>

RDA Europe

Email - info@europe.rd-alliance.org

Twitter - [@RDA_Europe](https://twitter.com/RDA_Europe)

RDA US

Twitter - [@RDA_US](https://twitter.com/RDA_US)

RDA/TDWG Attribution Metadata Standards for Collections

Anne E Thessen, David Shorthouse, Deb Paul, Mike Conlon,
Matt Woodburn, Dimitris Koureas, Sarah Ramdeen

annethessen@gmail.com

@diatomsRcool

Problem

- Curating and maintaining research collections is very important, but current incentives do not support this work
- Creating a metadata standard for attributing this work will help people and institutions get credit - thus changing the incentive structure

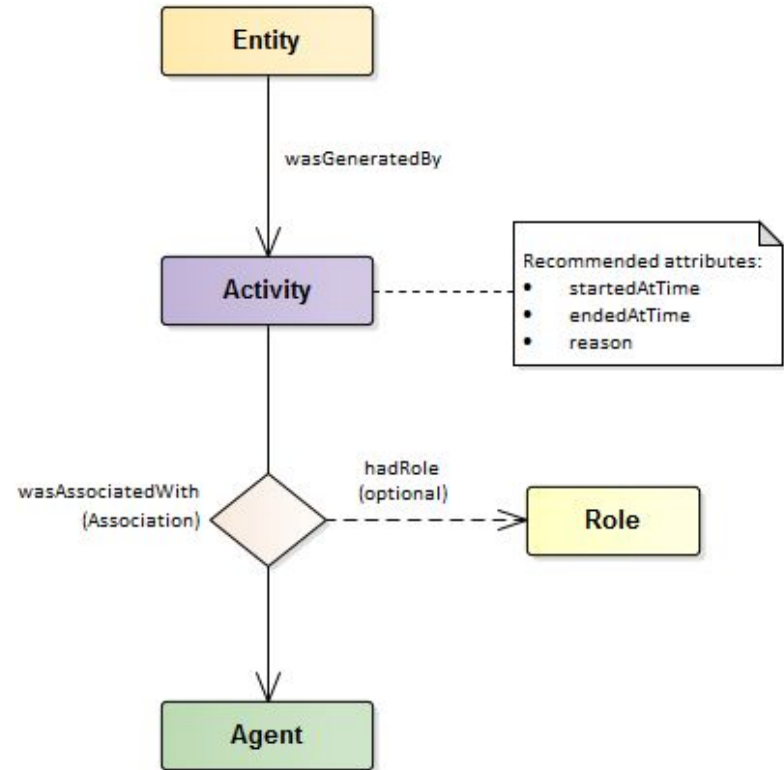


RDA/TDWG Process

- RDA: recommendations submitted to Council and to Data Science Journal
- TDWG: IG is using RDA recommendations to justify a Darwin Core extension
 - <https://github.com/tdwg/attribution>
 - Identifiers for people IG
- Recommendations are only the beginning!

Recommendations

- The key elements of the model for attribution are:
 - Entity wasGeneratedBy Activity
 - Activity wasAssociatedWith Agent
 - Association hadRole Role
- with some additional attributes assigned to the Activity class:
 - Activity has attribute StartDateTime
 - Activity has attribute EndDateTime
 - Activity has attribute Reason (added as comment)
- Provide meta-model to integrate data across disciplines



Adopters

- Tier 1: providers of attribution metadata

- Arctos
- iDigBio
- Bloodhound
- TaxonWorks



- Tier 2: consumers and presenters of attribution metadata

- ORCID
- ImpactStory
- Altmetric



Adopters


- This recommendation is a way to exchange attribution metadata between Tier 1 and Tier 2 adopters
- Pilot projects with ORCID, BloodHound, Data-Futures



ORCID Pilot Projects

David Peter
Shorthouse

ORCID iD

 <https://orcid.org/0000-0001-7618-5230>

Works (15 of 15)

Sort

Natural history specimens collected and/or identified and deposited.

Zenodo

2019-05-05 | data-set

DOI: 10.5281/zenodo.2667809

Source: DataCite

★ Preferred source

Natural history specimens collected and/or identified and deposited.

Zenodo

2019-05-05 | data-set

DOI: 10.5281/zenodo.2667631

Source: DataCite

★ Preferred source

May 5, 2019

Dataset Open Access

Natural history specimens collected and/or identified and deposited.

Shorthouse, David Peter

Natural history specimen data collected and/or identified by Shorthouse, David Peter, <https://orcid.org/0000-0001-7618-5230>. Claims were made on Bloodhound, <https://bloodhound-tracker.net> using specimen data from the Global Biodiversity Information Facility, <https://gbif.org>.

Preview

action	gbifID	occurrenceID	dateIdentified	decimalL
recorded	1414822067	http://bins.boldsystems.org/index.php/Public_RecordView?processid=COLNF1882-15		
recorded	1414821985	http://bins.boldsystems.org/index.php/Public_RecordView?processid=COLNF1660-15		
recorded	1414822099	http://bins.boldsystems.org/index.php/Public_RecordView?processid=COLNF2036-15		
identified	769279710	urn:catalog:UASM:UASM329573	2010	56.839

Files (8.4 kB)

11

views

9

downloads

See more details...

Indexed in

OpenAIRE

Publication date:

May 5, 2019

DOI:

DOI: [10.5281/zenodo.2667809](https://doi.org/10.5281/zenodo.2667809)

Keyword(s):

[specimen](#) [natural history](#) [taxonomy](#)

License (for files):

[Creative Commons Zero v1.0 Universal](#)

Versions

Version 2 10.5281/zenodo.2667809 May 5, 2019

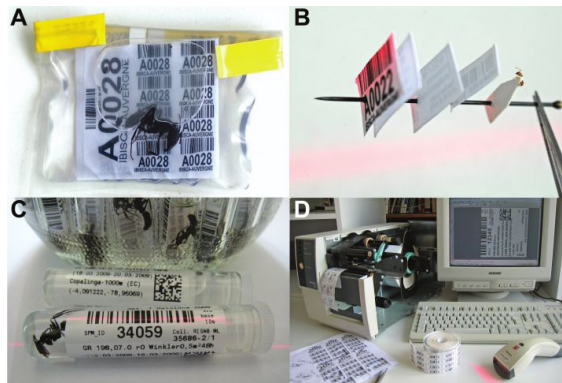
Version 1 10.5281/zenodo.2667632 May 5, 2019

Implementation Elephants

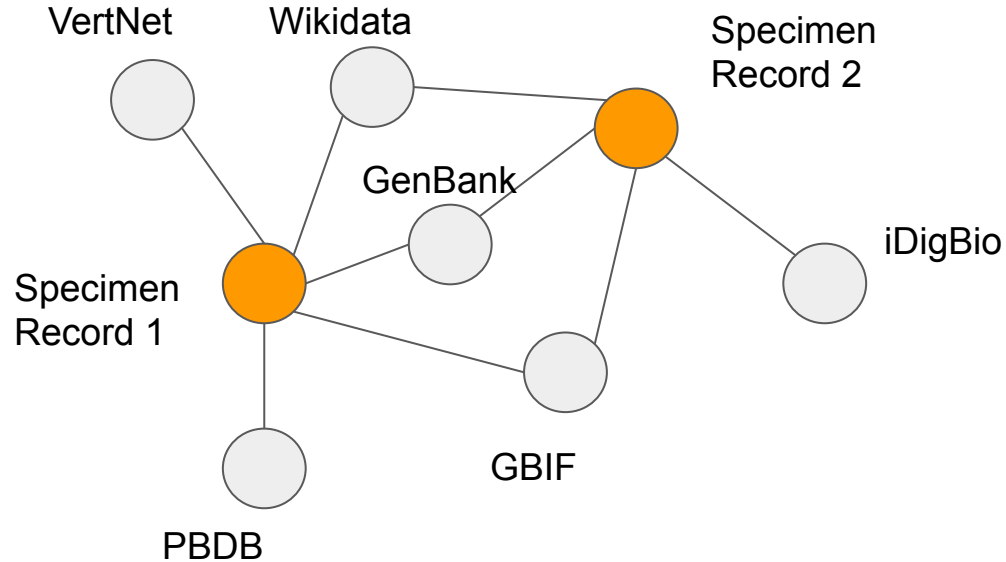
- Specimen identifiers are notoriously poor
 - Compare identifier graph
- 1000+ specimens on an ORCID profile
 - Pilot project: Archive lists and archive collections

DATA
FUTURES
PARTNERSHIP

Parathemisto (P.) japonica Bovallius		Cat. No.	100988
Stack	Shelf	(Size vial)	Alt. D XXX 25 specimens
LOCALITY	52° 03' N., 154° 20' E. from stomach of		
	<u>Oncorhynchus nerka</u>		
Station	382	DEPTH	When collected Aug. 10, 1955
Collected by	W. Aaron		
Received	Nov. 1956 from "Univ. of Wash.		
Remarks	GC		
Identified by	T. E. Bowman	Acc. No.	212615
U. S. NATIONAL MUSEUM GPO 16-41434-1 DIVISION OF MARINE INVERTEBRATES			



Specimen Identifiers: Partial Solution



Terminology Extensions

- ORCID is adding a “physical object” work type
- Contributor Role Ontology
 - Collector Role
 - Curator Role
- VIVO
 - Collecting Process
 - Curating Process
 - Georeferencing Process
 - Measuring Process
 - Taxonomic Identification Process
 - Data Collection Process
- What terms would you like to see? [Tell us!](#)



VIVO

connect • share • discover

Questions?

- Anne Thessen - annethessen@gmail.com
- @diatomsRcool



Research Data Repository Interoperability WG

Thomas Jejkal, David Wilcox

06/16/20

Exchange of Digital Content between Repository Platforms

-
- Exchange of digital content between repository platforms is challenging
 - Different data models, metadata, semantics, interfaces
 - Current state of the art (e.g. Dublin Core, OAI-PMH, SWORD) well adopted but limited
 - Upcoming approaches (e.g. Linked Data Platform, Resource Sync, OAI-ORE) supported by a comparable small subset of platforms
 - Some varied use cases: migration, replication, cross-domain sharing
 - Migration between same platform, but different version
 - Migration between different platforms
 - Scientists sharing research data within or cross-domain
 - Collected in primer document: <https://doi.org/10.15497/RDA00020>

Highlights of the Recommendation

-
- Consensus on BagIt-based approach
 - Recommend metadata location and naming scheme
 - Recommend to include basic metadata in Datacite 4.0 schema
 - Recommend to use machine readable profile information according to BagIt Profiles Specification

Bag Layout and Content (1)

BagIt-specific

```
.  
|___bagit.txt  
|___bag-info.txt  
|___tagmanifest-sha256.txt  
|___data  
|___manifest-sha256.txt  
|___metadata  
| |___datacite.xml  
|___fetch.txt
```

Bag Layout

```
BagIt-Version: 0.97  
Tag-File-Character-Encoding: UTF-8
```

bagit.txt

```
Bagging-Date: 2017-11-29  
Contact-Phone: +49 721 608-24042  
Source-Organization: Research Data Alliance  
Contact-Name: Thomas Jejkal  
Contact-Email: thomas.jejkal@kit.edu  
External-Identifier: 10.15497/RDA00020  
External-Description: Packaging containing the Research Data Repository Interoperability WG primer document.  
Bag-Size: 135 KB  
Payload-Oxum: 131643.1  
Source-Identifier: RDA File Repository  
BagIt-Profile-Identifier: https://raw.githubusercontent.com/RDAResearchDataRepositoryInteropWG/bagit-profiles/master/generic/0.1/profile.json
```

bag-info.txt

Bag Layout and Content (2)

BagIt Profile

```
{  
  "BagIt-Profile-Info": {  
    "BagIt-Profile-Identifier": "https://raw.githubusercontent.com/RDAResearchDataRepositoryInteropWG/bagit-profiles/master/generic/0.1/profile.json",  
    "Source-Organization": "rd-alliance.org",  
    "Contact-Name": "Research Data Repository Interoperability WG",  
    "Contact-Email": "rda-rdrinterop-wg@rda-groups.org",  
    "External-Description": "Base BagIt profile recommended by the RDA Research Data Repository Interoperability WG.",  
    "Version": "0.1"  
  },  
}
```

profile.json

Bag Layout and Content (2)

BagIt Profile

```

{
  "BagIt-Files-Hashes": {
    "Bag-Info": {
      "Bagging-Date": {
        "required": true
      },
      "Source-Organization": {
        "required": false
      },
      "Contact-Name": {
        "required": false
      },
      "Contact-Phone": {
        "required": false
      },
      "Contact-Email": {
        "required": true
      },
      "External-Identifier": {
        "required": false
      },
      "External-Description": {
        "required": true
      },
      "Bag-Size": {
        "required": true
      },
      "Payload-Oxum": {
        "required": true
      },
      "Source-Identifier": {
        "required": false
      }
    },
    "Bagging-Date": {
      "required": true
    },
    "Source-Organization": {
      "required": false
    },
    "Contact-Name": {
      "required": false
    },
    "Contact-Phone": {
      "required": false
    },
    "Contact-Email": {
      "required": true
    },
    "External-Identifier": {
      "required": false
    },
    "External-Description": {
      "required": true
    },
    "Bag-Size": {
      "required": true
    },
    "Payload-Oxum": {
      "required": true
    },
    "Source-Identifier": {
      "required": false
    }
  },
  "Bagging-Date": {
    "required": true
  },
  "Source-Organization": {
    "required": false
  },
  "Contact-Name": {
    "required": false
  },
  "Contact-Phone": {
    "required": false
  },
  "Contact-Email": {
    "required": true
  },
  "External-Identifier": {
    "required": false
  },
  "External-Description": {
    "required": true
  },
  "Bag-Size": {
    "required": true
  },
  "Payload-Oxum": {
    "required": true
  },
  "Source-Identifier": {
    "required": false
  }
}

```

profile.json

Bag Layout and Content (2)

BagIt Profile

```

{
  "BagIt-Files-Required": [
    "Manifests-Required": [
      "sha256"
    ],
    "Allow-Fetch.txt": true,
    "Serialization": "optional",
    "Accept-Serialization": [
      "application/zip",
      "application/tar",
      "application/tar+gzip"
    ],
    "Accept-BagIt-Version": [
      "0.97"
    ],
    "Tag-Manifests-Required": [
      "sha256"
    ],
    "Tag-Files-Required": [
      "metadata/datacite.xml"
    ]
  ],
  "Bag-Info": {
    "Bagging-Date": {
      "required": true
    },
    "Source-Organization": {
      "required": false
    },
    "Contact-Name": {
      "required": false
    },
    "Contact-Phone": {
      "required": false
    },
    "Contact-Email": {
      "required": true
    },
    "External-Identifier": {
      "required": false
    },
    "External-Description": {
      "required": true
    },
    "Bag-Size": {
      "required": true
    },
    "Payload-Oxum": {
      "required": true
    },
    "Source-Identifier": {
      "required": false
    }
  },
  "Source-Organization": "RDAResearchDataRepository Interoperability WG",
  "Contact-Name": "RDAResearchDataRepository Interoperability WG",
  "Contact-Email": "rdar@rda-groups.org",
  "External-Description": "This profile recommended by the Interoperability WG.",
  "Accept-BagIt-Version": "0.97",
  "Tag-Manifests-Required": [
    "sha256"
  ],
  "Tag-Files-Required": [
    "metadata/datacite.xml"
  ]
}

```

profile.json

Bag Layout and Content (3)

DataCite Metadata

```
<resource xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.15497/RDA00025</identifier>
  <creators>
    <creator>
      <creatorName>Jejkal, Thomas</creatorName>
      <nameIdentifier>0000-0003-2804-688X</nameIdentifier>
      <nameIdentifierScheme>ORCID</nameIdentifierScheme>
    </creator>
    <creator>
      <creatorName>Wilcox, David</creatorName>
      <nameIdentifier>0000-0001-5411-9208</nameIdentifier>
      <nameIdentifierScheme>ORCID</nameIdentifierScheme>
    </creator>
    <creator>
      <creatorName>Members of the RDA Research Data Repository Interoperability WG</creatorName>
    </creator>
  </creators>
  <titles>
    <title>Research Data Repository Interoperability Primer</title>
  </titles>
  <publisher>Research Data Alliance</publisher>
  <publicationYear>2017</publicationYear>
  <subjects>
    <subject>Repository Platform</subject>
    <subject>Interoperability</subject>
    <subject>Standards</subject>
    <subject>API</subject>
    <subject>Tools</subject>
  </subjects>
  <language>eng</language>
  <resourceType resourceTypeGeneral="Text">RDA Working Group Deliverable</resourceType>
  <version>1</version>
  <descriptions>
    <description descriptionType="Abstract">Huge amounts of research data stored in a multitude of research data repository platforms can often only be used by a comparably small audience. On the one hand, this is caused by differences in semantics, underlying data models and metadata schemas, whose complexity and number prevents scientists from taking advantage of them. On the other hand, a lack of interoperability between research data repository platforms causes research data not to be used to their full potential. The goal of the RDA Research Data Repository Interoperability WG (RDRIWG) is to achieve consensus on an adoptable approach to facilitating research data repository interoperability for a defined set of initial use cases.
  </description>
  </descriptions>
</resource>
```

metadata/datacite.xml

Bag Layout and Content (4)

Content-specific

<https://www.rd-alliance.org/system/files/ResearchDataRepositoryInteroperabilityPrimerfinal.pdf> 131643 data/primer_final.pdf

fetch.txt

70cc12fb6207ff00b64d280450b7c82916d7326c39ab0385754ec25480ea63a8 data/primer_final.pdf

manifest-sha256.txt

4082975711093941170f02e0d4ca777b4af78c8c39a19fb06ef920c831753590 metadata/datacite.xml

tagmanifest-sha256.txt

Impact of the Recommendation

-
- Comparably easy adoptable, packaging and exchange format
 - Adoption can be done by community
 - Alignment of existing packaging solutions rather easy
 - Support for both state-of-the-art and legacy platforms
 - Provides users with a common tool to e.g. transport, exchange or backup their content
 - Basis for further agreements of package content to improve interpretability of package content

Endorsements/Adopters

-
- Enables exchange of digital content between repository platforms
 - First prototypes implemented for Fedora Commons, ICAT, Dariah Repository and KIT Data Manager
 - Further adoptions available for Dataverse and 'Qualitative Data Repository'
 - Generic profile containing all recommended elements available
 - Tools and demonstrator available:
<https://github.com/RDAResearchDataRepositoryInteropWG>
 - Final recommendations: <https://doi.org/10.15497/RDA00025>

A Model Of Data Integration Related To Wheat and Fusarium spp Resources

Iliyan Mihaylov, Stefan Tsonev, Dimitar Vassilev
Sofia University "St. Kliment Ohridski"

RDA Practice Webinar "Adoption Week Data Description" June 9, 2020

Fungi of genus *Fusarium* are opportunistic plant pathogens



Macroconidia of *Fusarium graminearum*

Source: https://nl.wikipedia.org/wiki/Fusarium_graminearum



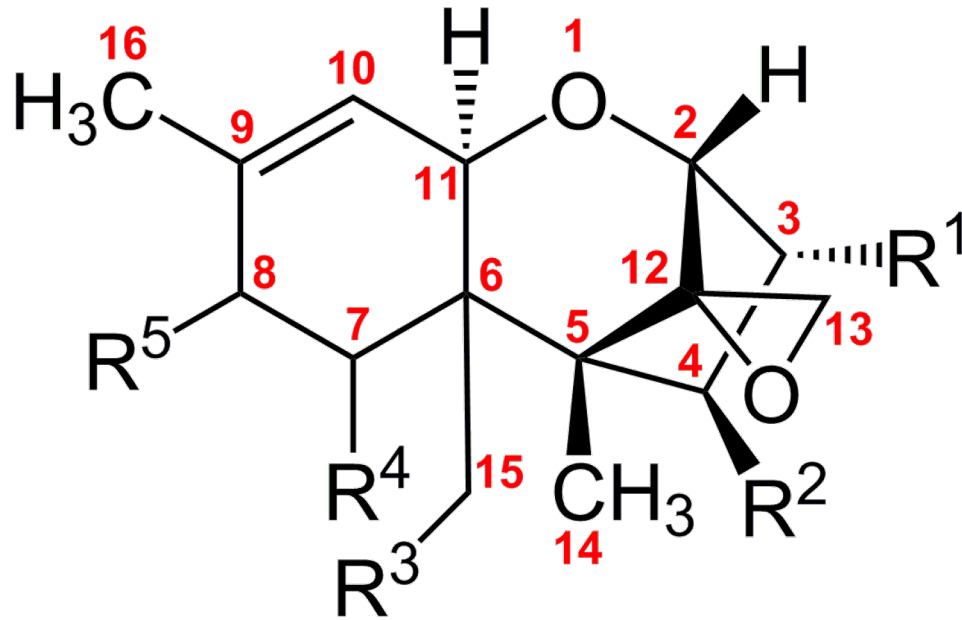
Diseased wheat ear – Fusarium Head Blight

Source: <https://ohioline.osu.edu/factsheet/plpath-cer-06>

The genus *Fusarium* consists of a great number of saprophytic fungi, which under appropriate conditions (high humidity and temperatures during the plant flowering phase) cause pathological conditions in important crops leading to huge losses.

In wheat the representatives of the genus cause crown and root rotting or Fusarium Head Blight (FHB)

Fusarium mycotoxins cause additional losses contaminating the grains and making them unusable for feed or food



General structure of trichothecenes – *Fusarium* mycotoxins

Economically important mycotoxins:

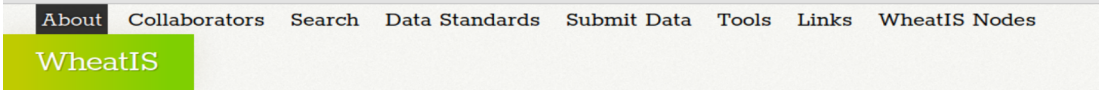
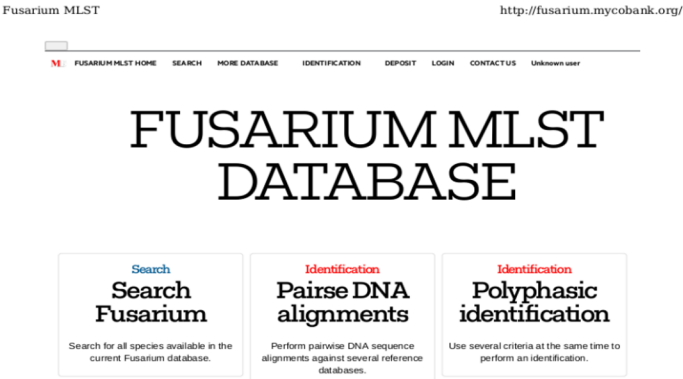
	R ¹	R ²
• Deoxynivalenol	OH	H
• 3-Acetyldeoxynivalenol	OAc	H
• Nivalenol	OH	OH

Strains producing 3-Acetyldeoxynivalenol are more aggressive, causing a higher disease severity and greater losses due to toxin contamination. In addition this chemotype has an ecological advantage in the context of rising global temperatures.

The study of *Fusarium* fungi is challenging because of the complex taxonomic relations in the genus arising from:

- ◆ The presence of species complexes whose representatives are difficult to distinguish using morphological characteristics;
- ◆ Usual molecular approaches for species distinction are not applicable due to the highly conservative sequence of the internal transcribed spacer of the ribosomal gene cluster;
- ◆ Chemotypes (type of toxins produced) occurring within the species pose another problem, as they tend to exhibit different aggressiveness.

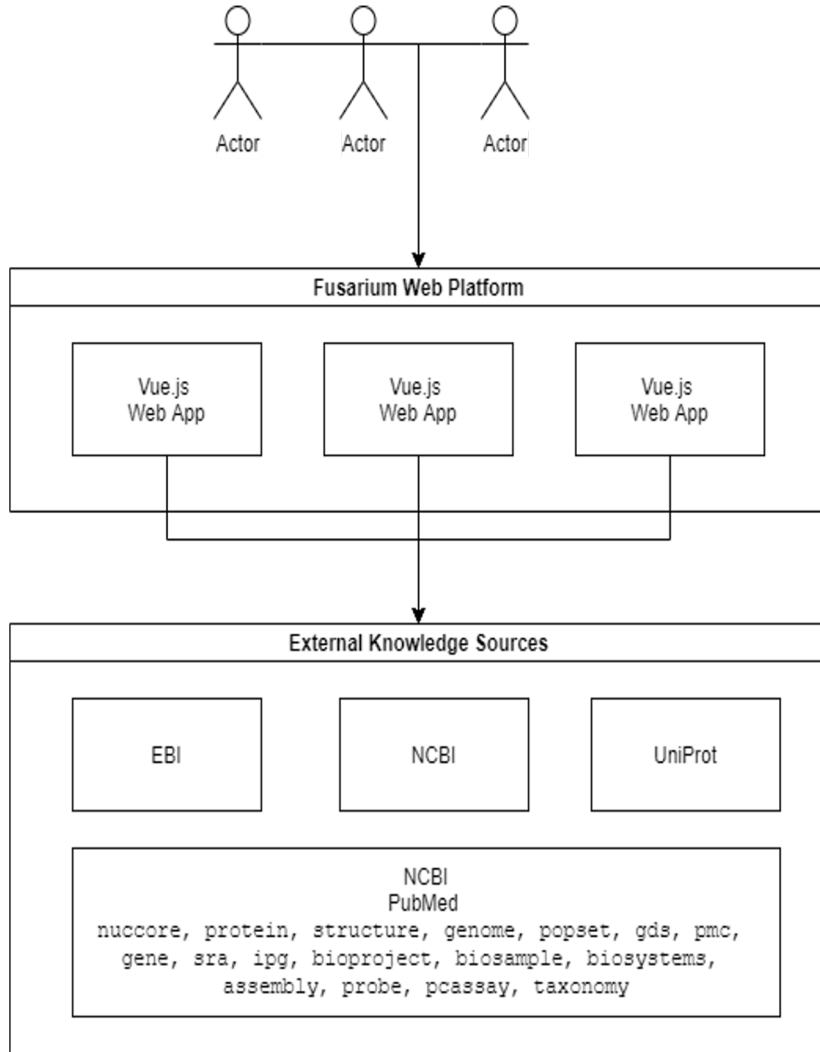
A growing body of research data exists, but it remains fragmented



One ring to rule them all



Concept



- One platform for all external sources related to fusarium
- RESTful integrated scenario - each existing platform which support RESTful standards can be integration in our scenario
- The realized platform is fully distributed based on modern solutions an environments as kubernetes, docker and all cloud providers supported images.
- Scenario for external sources integration without RESTful support can be applied with a development of RESTful API. Example

Fusarium Taxonomy

Search For Fusarium Taxonomy

Fusarium

SEARCH CLEAR

Name: Fusarium Mnemonic: 9HYPO Rank: genus Superregnum: E

Linked Databases

Nucleotide Protein Structure Genome Popset GEO Datasets PubMed Central Gene SRA Experiments Identical Protein Groups

Bio Project Bio Sample Bio Systems Assembly Probe PubChem BioAssay Taxonomy

Siblings Links

Aquanectria	Allantonectria	Microcera	Corallonectria
Mnemonic: 9HYPO	Mnemonic: 9HYPO	Mnemonic: 9HYPO	Mnemonic: 9HYPO
Rank: genus	Rank: genus	Rank: genus	Rank: genus
Superregnum: E	Superregnum: E	Superregnum: E	Superregnum: E
ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link

Children Links

Fusarium cortaderiae	Fusarium fusarioides	Fusarium mesoamericanum	Fusarium acaciae-mearnsii
Mnemonic: FUSCO	Mnemonic: 9HYPO	Mnemonic: FUSMI	Mnemonic: FUSAC
Rank: species	Rank: species	Rank: species	Rank: species
Superregnum: E	Superregnum: E	Superregnum: E	Superregnum: E
ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link	ParentLink: UniProt Link NCBI Link

Items per page: 5 1-8 of 124

Usage

- Autocomplete search is based on external knowledge sources.
- Links for all reference related databases from NCBI (*nucleotide, protein, structure, genome, popset, gds, pmc, gene, sra, ipg, bioproject, biosample, biosystems, assembly, probe, pccassay, taxonomy*)
- Separate module for each child and sibling related fusarium data.
- Separate module for each fusarium data with links to all databases where we found it.
- Support huge amount of data via async requests and visualization

Integrating Mycotoxin information

Information about mycotoxins - compounds, biosynthetic pathways, proteins and genes will be retrieved from NCBI (PubChem) and MetaCyc

Additional information about geographical distribution of strains producing particular toxins will be retrieved from the European *Fusarium* database - LuxMCC

Conclusion

We develop a findable, accessible, interoperable and reusable semantic model of a platform for integrating data from different resources, concerning representatives of genus *Fusarium* and their interaction with the wheat.

Platform contributes the discovery and presenting of new knowledge concerning the role and importance of *Fusarium* spp. in cereals and in particular in wheat