

Frictionless Data for Wheat

SIMON TYRRELL
RESEARCH SOFTWARE ENGINEER

XINGDONG BIAN
RESEARCH SOFTWARE ENGINEER

ROB DAVEY
GROUP LEADER



Decoding Living Systems

Designing Future Wheat (DFW)

The [BBSRC](#) funded Designing Future Wheat (DFW) Institute Strategic Programme, spans over 25 groups of scientists across 8 research institutes and universities and aims to develop new wheat germplasm containing the next generation of key traits.

It is anticipated that the world will need to produce 60% more wheat by 2050 to meet global demand. Since it takes between 15 and 20 years for current research to improve wheat varieties grown in farmers' fields, it is imperative that we act now to address problems facing us in the future

Taken from <https://designingfuturewheat.org.uk/>

DFW Data

DFW produces lots of scientific data

- Field Trial experiments
- Datasets
- Sequences

Target audiences

We have different groups of users

- Breeders
- Academics
- Data Scientists
- Industry

Challenge

To make the data accessible and usable for everyone

Grassroots Infrastructure

A suite that wraps up industry-standard software tools along with our own custom open-source ones

- Consistent JSON-based API
 - Language and platform agnostic
- Can be federated with other Grassroots instances
- Sharing data and services in a [FAIR](#) way

FAIR data principles - Findable

The first step in (re)using data is to find them.

- Data are described with rich metadata
- Metadata and data should be easy to find for both humans and computers.
- Machine-readable metadata are essential for automatic discovery of datasets and services

Taken from <https://www.go-fair.org/fair-principles/>

FAIR data principles - Accessible

Once the user finds the required data, they need to know how can they be accessed, possibly including authentication and authorization.

- (Meta)data are retrievable by their identifier using a standardized communications protocol
- Metadata are accessible, even when the data are no longer available

Taken from <https://www.go-fair.org/fair-principles/>

FAIR data principles - Interoperable

The data usually need to be integrated with other data.

- Able to be easily integrated with applications or workflows for analysis, storage, and processing.
- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

Taken from <https://www.go-fair.org/fair-principles/>

FAIR data principles - Reusable

The ultimate goal of FAIR is to optimize the reuse of data.

- Metadata and data should be well-described so that they can be replicated and/or combined in different settings
- Metadata and data are associated with detailed provenance

Taken from <https://www.go-fair.org/fair-principles/>

Grassroots Infrastructure

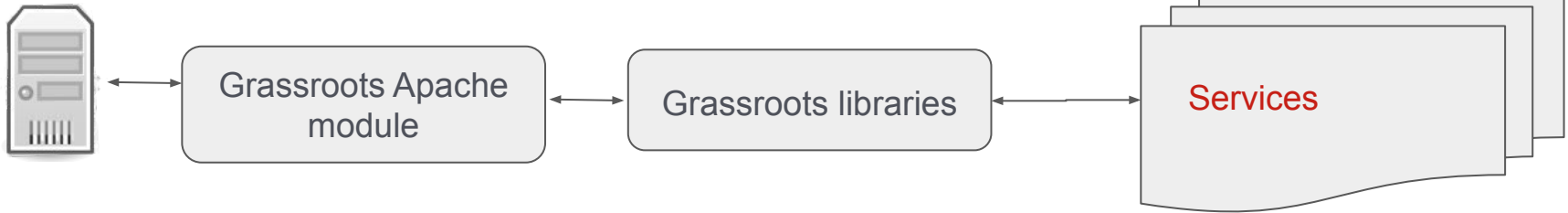
Apache httpd
Web Server



- Grassroots Apache module acts as a bridge between Apache and Grassroots
- A set of cross-platform libraries that can be used by Grassroots components including
 - Networking code to access web resources
 - Server and Service management tools
 - API to & from our web services and their parameters

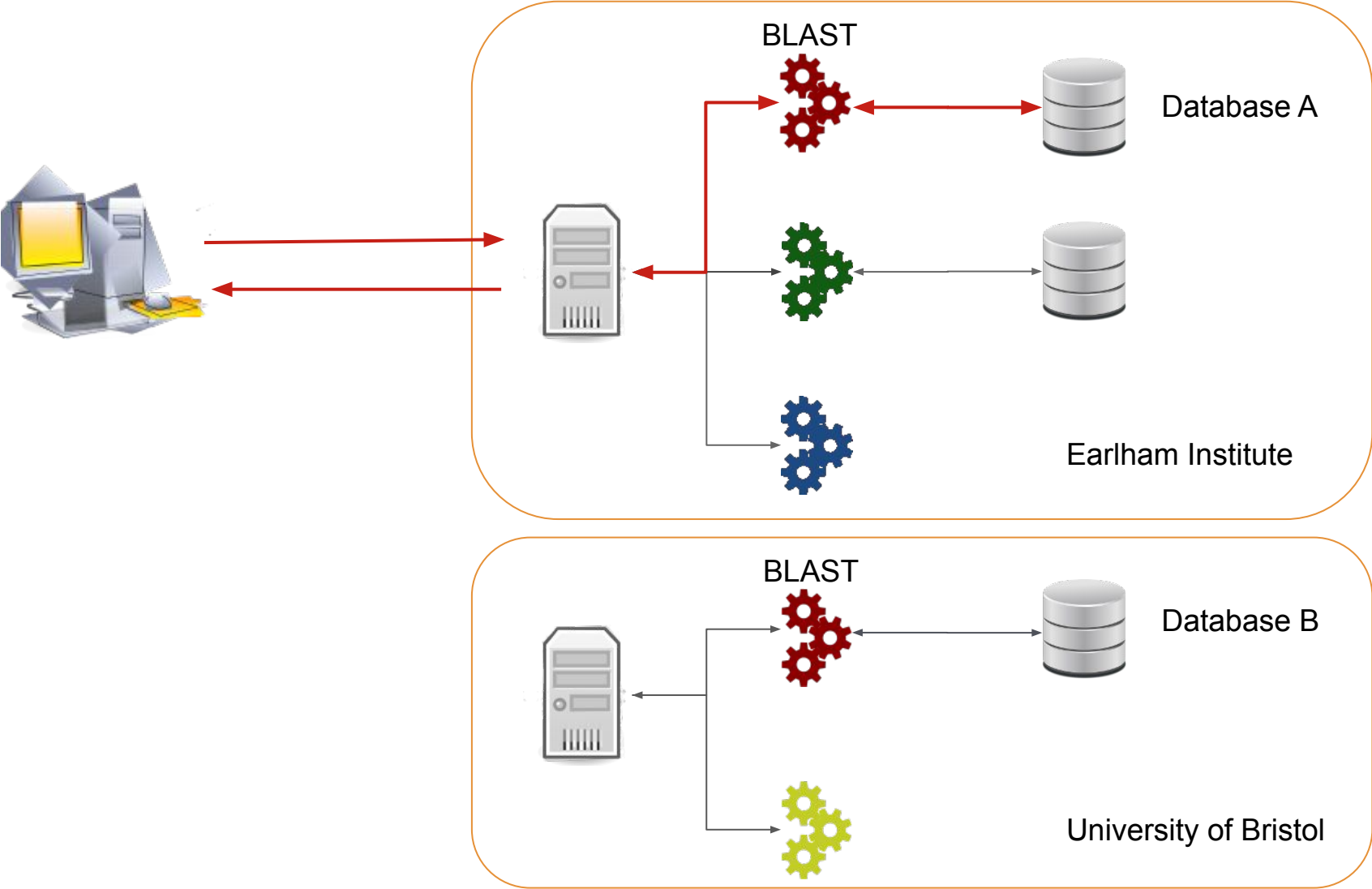
Grassroots Infrastructure - Services

Apache httpd
Web Server

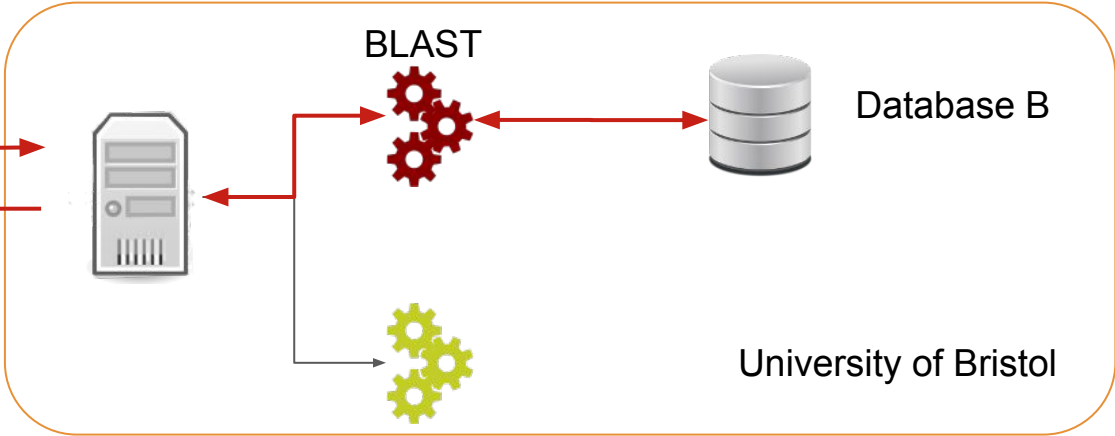
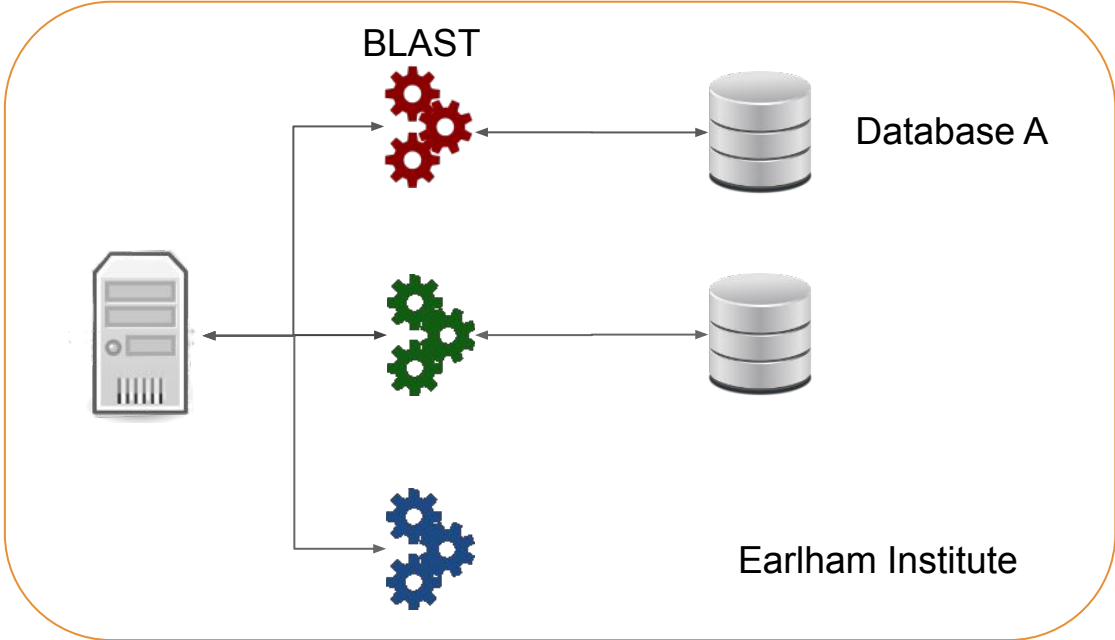


- Components that perform scientific analysis
 - Adapting existing programs
 - Writing our own bespoke tools
- Tools that conform to the Grassroots Services API, which is a well-defined set of standards to access tools and data *e.g.*
 - BLAST
 - - Find areas of similarity between biological sequences
 - Field Trials
 - Unified Search

Standard Web Service Interaction



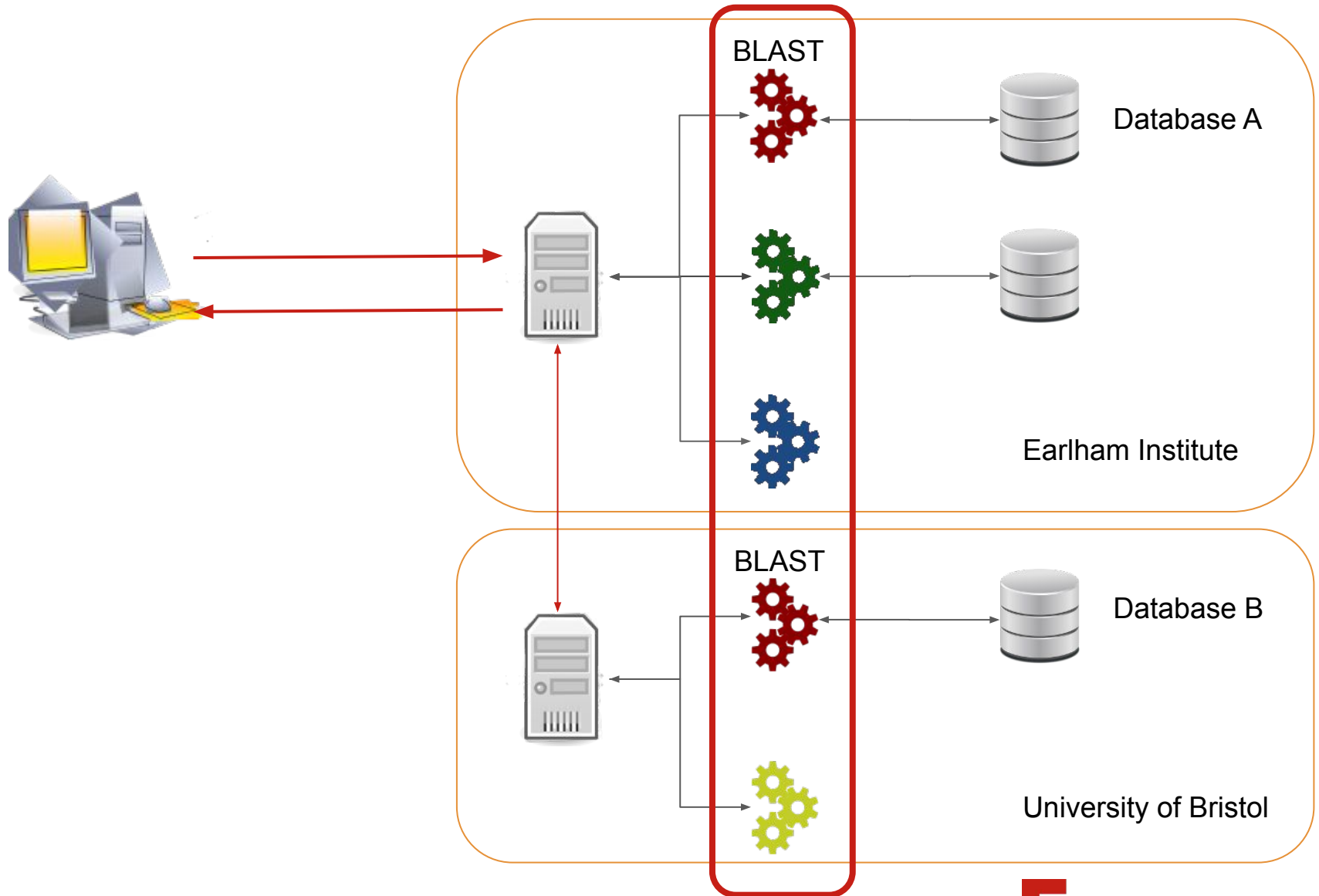
Standard Web Service Interaction



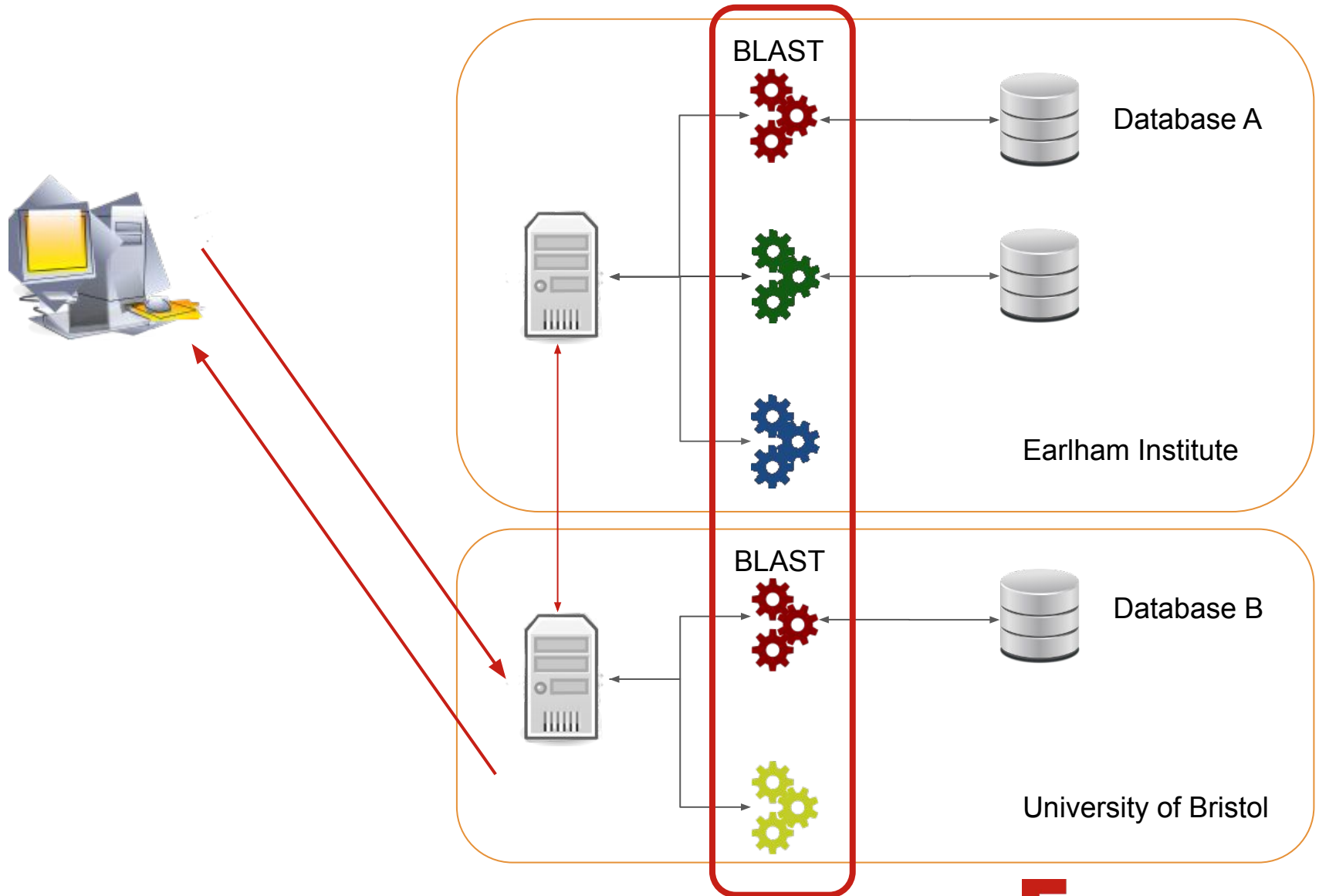
Issues

- Manually having to access each Service individually
- Collation of results
- Human error
- Not running each service with the same parameters
- Mistakes when putting the results together
- Time consuming

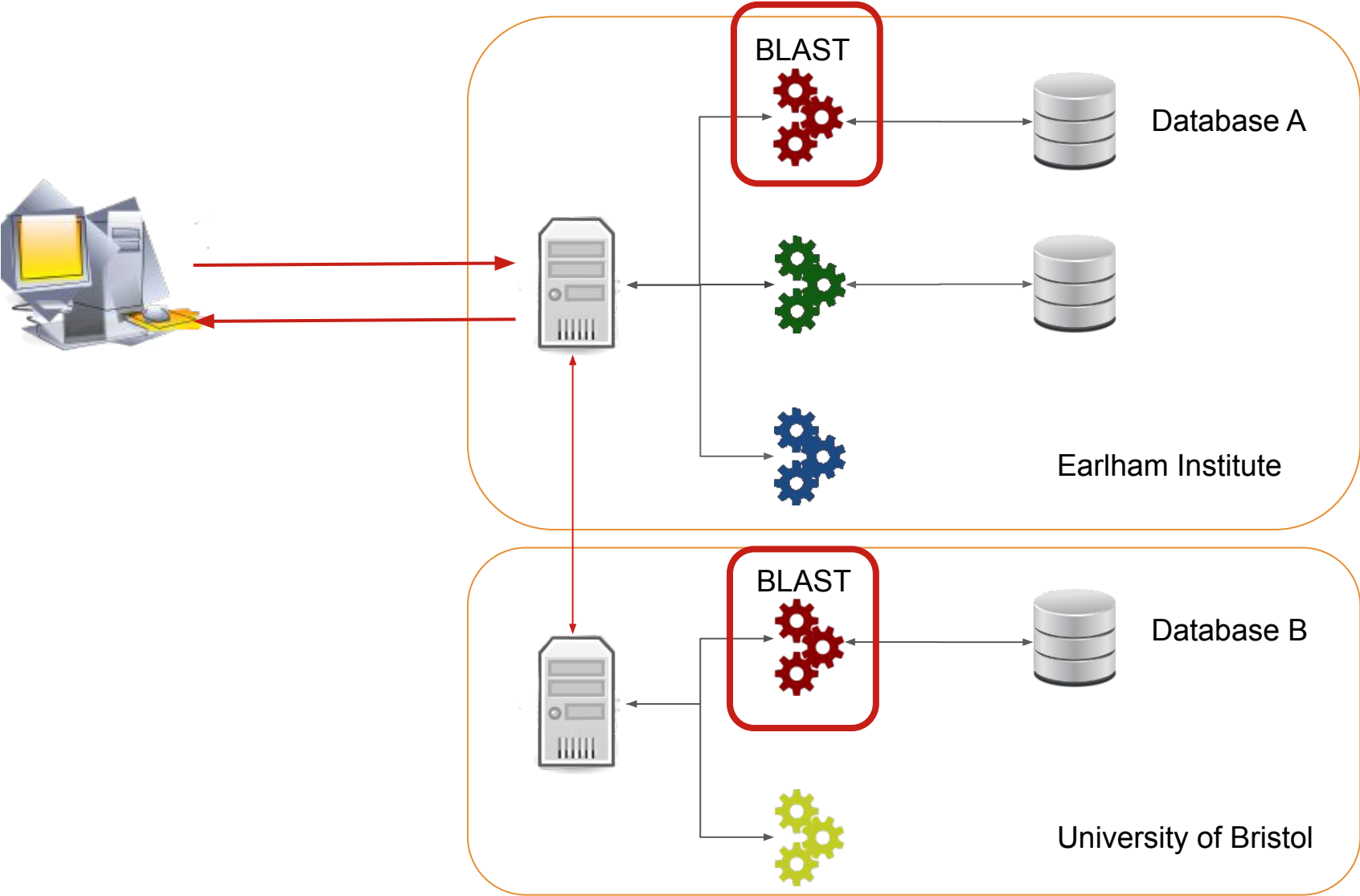
Federating Services



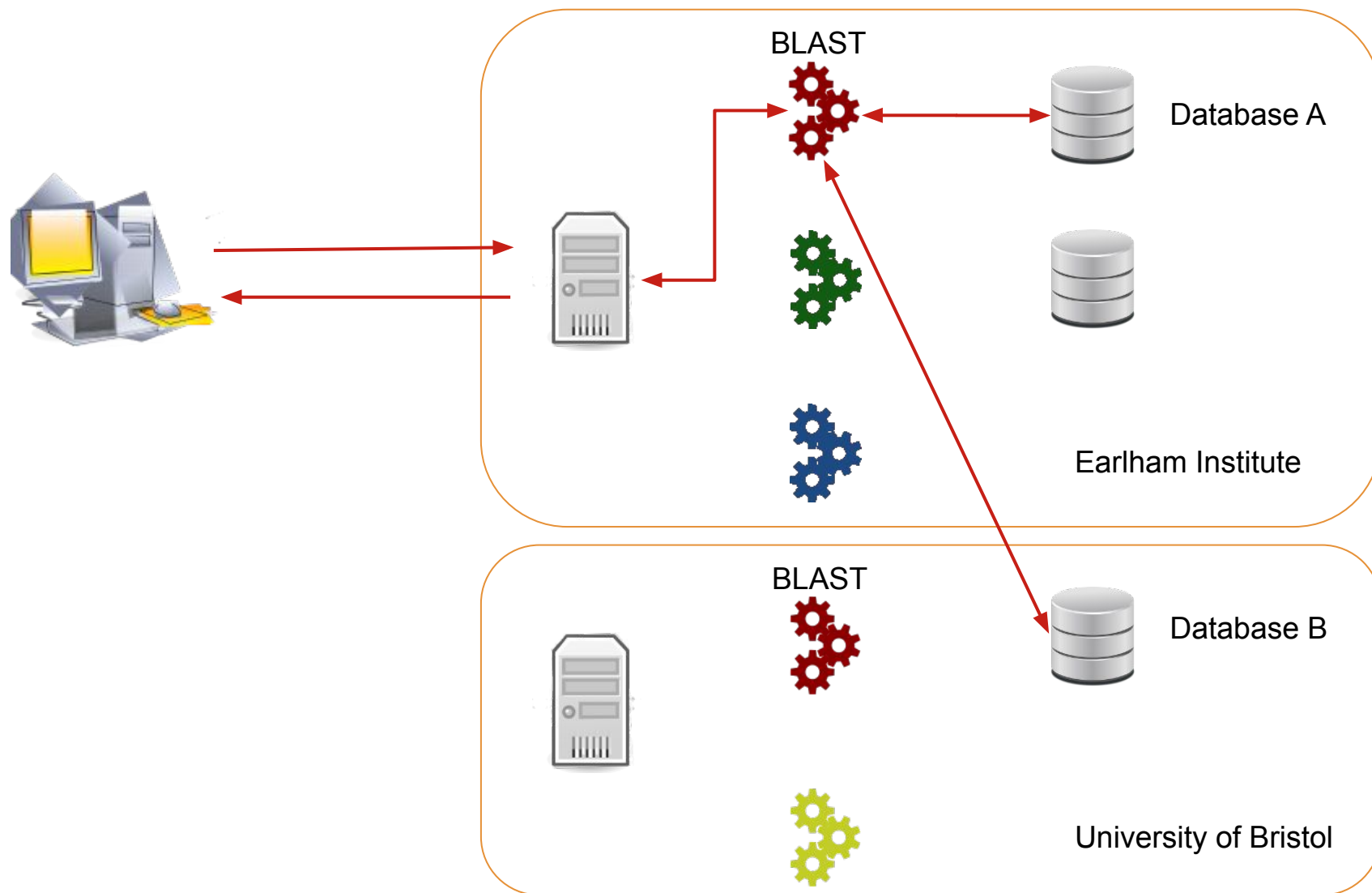
Different Server, Same List of Services



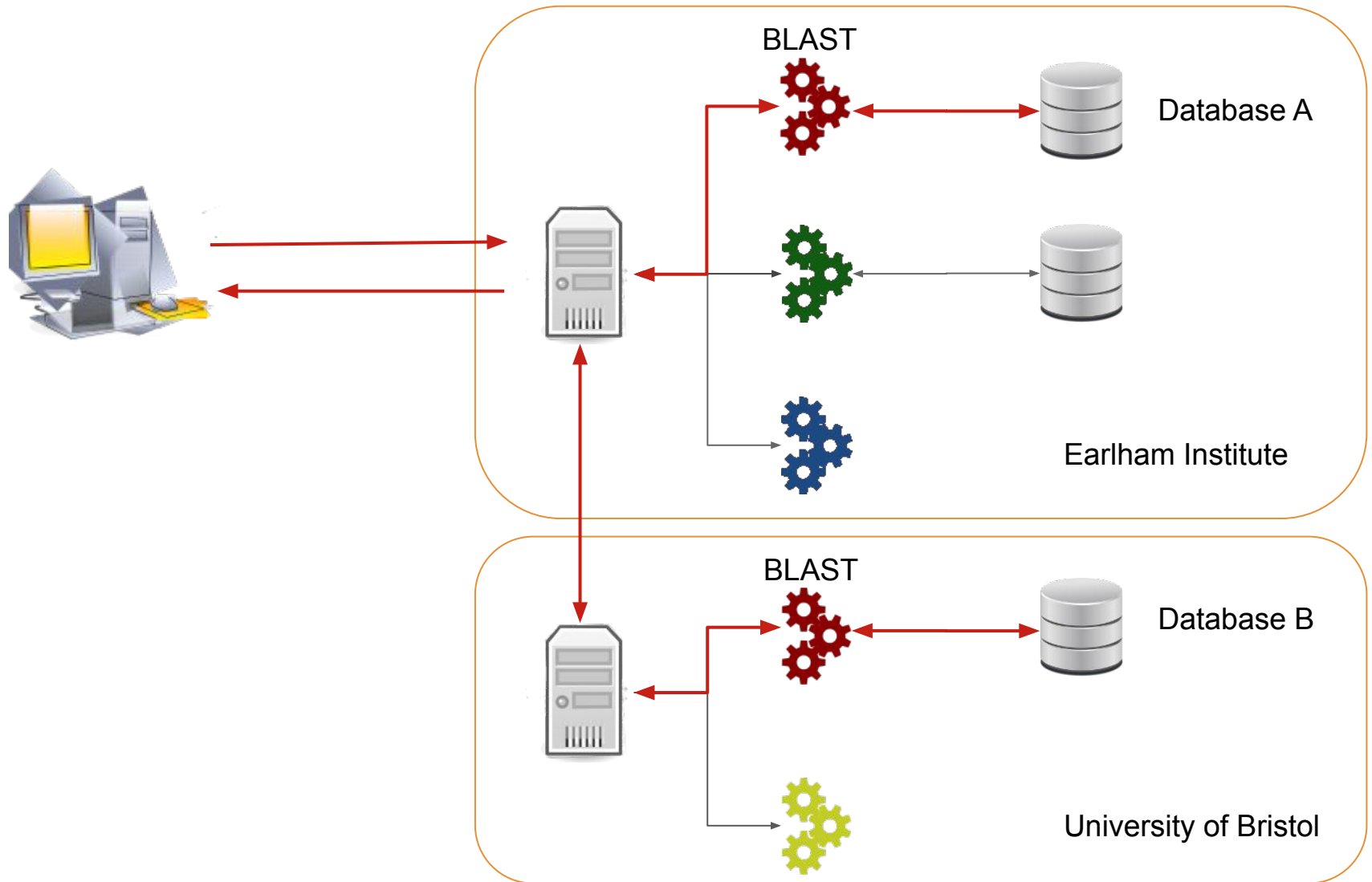
Same Services...



... Get Amalgamated



Under the Hood



DFW Data Portal

Available at

https://opendata.earlham.ac.uk/wheat/under_license/toronto/

- A repository for all data generated within DFW
- Based upon the [Toronto data agreement](#)
 - Prepublication data sharing
 - This agreement does not expire by time but only upon publication of the first global analysis by the data producers and contributors.

iRODS

An open source data management software used by research, commercial, and governmental organizations worldwide.

- Data virtualization
- Data discovery
 - Search on standard file and directory attributes
 - Metadata for each object
- Workflow automation
- Secure collaboration
- Federation

<https://irods.org/>

mod_eirods_dav

Our open source Apache module, forked from <https://github.com/UtrechtUniversity/davrods> to access iRODS repositories using standard web technologies.

- Themeable listings similar to mod_autoindex
- Metadata display and editing
- Show public or authenticated user data
- Full REST Web Service API

<https://github.com/billyfish/eirods-dav>

DFW Data Portal - Metadata

Metadata based upon the [Minimum Information About a Plant Phenotyping Experiment \(MIAPPE\)](#) standard

The screenshot displays the 'Grassroots Data' portal interface. A modal window titled 'Properties for BW_01001/' is open, showing the following metadata fields:

Accession: BW_01001
City: Behshahr
Country: Iran
GRU entry number: TOWWC002
insert size: 500
Instrumentation: Illumina HiSeq
Laboratory techniques: Whole Genome Sequencing
Latitude: 36.6953
Longitude: 53.5365
Project Code: BB/P016855/1
project uuid: 9791ca43-ddf0-4bfe-8fc0-50ff69cd6229
Species: Aegilops tauschii
State: Mazandaran
Taxonomy ID: 37682

The background shows a table with the following entries:

Name
BW_01001/
BW_01002/
BW_01003/
BW_01004/

The page also includes the Earham Institute logo in the top right and a timestamp '2018-02-21 16:13' at the bottom center.

DFW Data Portal - Project

The screenshot shows a web browser window with the URL https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez. The page title is "Grassroots Data Repository" and the main heading is "THE TRANSCRIPTIONAL LANDSCAPE OF HEXAPLOID WHEAT ACROSS TISSUES, CULTIVARS, AND STRESS CONDITIONS". The authors listed are Ricardo Ramirez-Gonzalez, Philippa Borrill, and Cristobal Uauy. The text describes a genome-wide analysis of homoeolog expression patterns in hexaploid bread wheat. The project code is BB/P016855/1. The data is available under the [Toronto Agreement](#). The location path is: Home > under_license > toronto > Ramirez-Gonzalez_et_al_2018-06025-Transcriptome-Landscape. Below the text is a table listing data packages:

Name	Size	Date	Properties
datapackage.json	203KB	2020-09-21 11:34	
data/		2020-08-03 05:47	
expvip/		2020-08-03 05:44	
scripts/		2020-08-03 05:47	
synthetic/		2020-08-03 05:47	

Brought to you by [mod_eirods_dav](#)

Projects have

- Titles
- Authors
- Descriptions
- License details
- Data

All indexed and searchable using our [Lucene-based](#) text search engine

DFW Data Portal - Frictionless Data

The screenshot shows a web browser window with the URL https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez. The page title is "Grassroots Data Repository" and the main heading is "THE TRANSCRIPTIONAL LANDSCAPE OF HEXAPLOID WHEAT ACROSS TISSUES, CULTIVARS, AND STRESS CONDITIONS" by Ricardo Ramirez-Gonzalez, Philippa Borrill, and Cristobal Uauy. The article text discusses the coordinated expression of highly related homoeologous genes in polyploid species. Below the article, there is a "Location" breadcrumb: Home > under_license > toronto > Ramirez-Gonzalez_etal_2018-06025-Transcriptome-Landscape. A table lists data packages with columns for Name, Size, Date, and Properties. The first row, "datapackage.json", is highlighted with a red box. The footer of the page says "Brought to you by mod_eirods_dav".

Grassroots Data Repository





















THE TRANSCRIPTIONAL LANDSCAPE OF HEXAPLOID WHEAT ACROSS TISSUES, CULTIVARS, AND STRESS CONDITIONS

Ricardo Ramirez-Gonzalez, Philippa Borrill, Cristobal Uauy

The coordinated expression of highly related homoeologous genes in polyploid species underlies the phenotypes of many of the world's major crops. However, the balance of homoeolog expression across diverse tissues, stress conditions, and cultivars remains poorly understood. Here we combine extensive gene expression datasets with the fully annotated genome sequence to produce a comprehensive, genome-wide analysis of homoeolog expression patterns in hexaploid bread wheat. Bias in homoeolog expression varied between tissues, with ~30% of wheat homoeologs showing unbalanced expression. We found expression asymmetries along wheat chromosomes, with genes showing the largest inter-tissue, inter-cultivar, and coding sequence variation most often located in the high-recombination distal ends of chromosomes. These transcriptionally dynamic genes potentially represent the first steps towards neo/sub-functionalization of wheat homoeologs. Co-expression networks revealed extensive coordination of homoeologs throughout development and, alongside a detailed expression atlas, provide a framework to target candidate genes underpinning agronomic traits in polyploid wheat. Project Code: BB/P016855/1

This data is made available under the [Toronto Agreement](#)

Location: [Home](#) > [under_license](#) > [toronto](#) > [Ramirez-Gonzalez_etal_2018-06025-Transcriptome-Landscape](#)

Name	Size	Date	Properties
 datapackage.json	203KB	2020-09-21 11:34	  
 data/		2020-08-03 05:47	  
 expvip/		2020-08-03 05:44	  
 scripts/		2020-08-03 05:47	  
 synthetic/		2020-08-03 05:47	  

Brought to you by [mod_eirods_dav](#)

Open Knowledge Tool Fund
to expose our DFW Data
Portal datasets and
publications as Frictionless
Data Packages

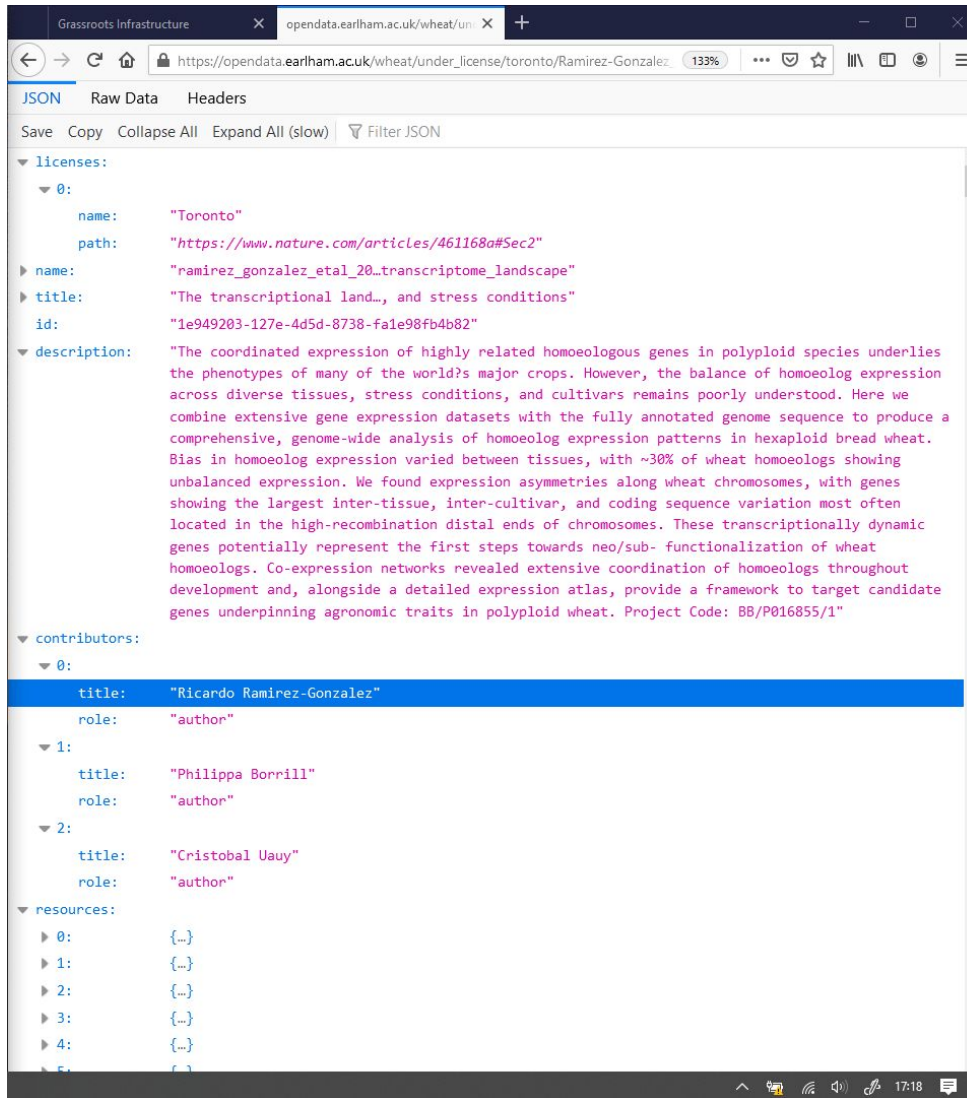
Frictionless Data

A Frictionless Data Package is a simple container format used to describe and package a collection of data.

- Can package any kind of data.
- Simple
- Extensible
- Metadata that is human-editable and machine-usable
- Reuse of existing standard formats for data
- Language, technology and infrastructure agnostic

<https://frictionlessdata.io/data-package/>

DFW Data Portal - Frictionless Data Package



The screenshot shows a web browser window displaying a JSON data package. The browser's address bar shows the URL: https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez. The JSON content is as follows:

```
{
  "licenses": [
    {
      "name": "Toronto",
      "path": "https://www.nature.com/articles/461168a#Sec2"
    },
    {
      "name": "ramirez_gonzalez_et_al_20...transcriptome_landscape",
      "title": "The transcriptional land., and stress conditions",
      "id": "1e949203-127e-4d5d-8738-fa1e98fb4b82"
    }
  ],
  "description": "The coordinated expression of highly related homoeologous genes in polyploid species underlies the phenotypes of many of the world's major crops. However, the balance of homoeolog expression across diverse tissues, stress conditions, and cultivars remains poorly understood. Here we combine extensive gene expression datasets with the fully annotated genome sequence to produce a comprehensive, genome-wide analysis of homoeolog expression patterns in hexaploid bread wheat. Bias in homoeolog expression varied between tissues, with ~30% of wheat homoeologs showing unbalanced expression. We found expression asymmetries along wheat chromosomes, with genes showing the largest inter-tissue, inter-cultivar, and coding sequence variation most often located in the high-recombination distal ends of chromosomes. These transcriptionally dynamic genes potentially represent the first steps towards neo/sub-functionalization of wheat homoeologs. Co-expression networks revealed extensive coordination of homoeologs throughout development and, alongside a detailed expression atlas, provide a framework to target candidate genes underpinning agronomic traits in polyploid wheat. Project Code: BB/P016855/1"
  },
  "contributors": [
    {
      "title": "Ricardo Ramirez-Gonzalez",
      "role": "author"
    },
    {
      "title": "Philippa Borrill",
      "role": "author"
    },
    {
      "title": "Cristobal Uauy",
      "role": "author"
    }
  ],
  "resources": [
    { ... },
    { ... },
    { ... },
    { ... },
    { ... },
    { ... }
  ]
}
```

Each Project within the Data Portal has a Frictionless Data Package currently containing

- License
- Name
- Description
- Authors
- Title
- Id

mod_eirods_dav - Frictionless Data additions

Dynamic creation of Frictionless Data Packages by generating values from the iRODS metadata values

- Can be cached and written back to iRODS
- Uses standard Apache configuration directives to decide which directories have Frictionless Data Packages generated
- The iRODS metadata keys to use are completely configurable
- Multiple iRODS metadata values can be combined
- Tabular Data Resources are also supported
 - Column headings and variable types are completely configurable

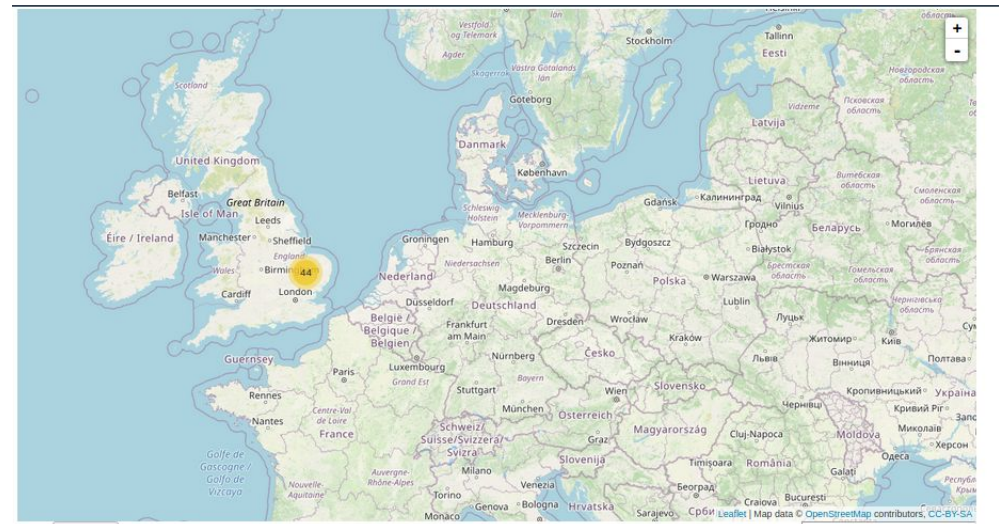
Field Trials

Experiments where different crops are planted in plots within a field, differing treatments applied and then traits are measured.

- Standardised template for submitting the genotype (the genetic material of the crop) and the phenotype (the characteristics that you want to measure) data
- To facilitate publishing of data compliant with FAIR sharing principles

Field Trials - Findable

The experimental data can be accessed using a map-based view and a searchable table of the data...



Show 10 entries

Name	Team	Description	Sowing Date	Harvest Date	Plots	Address	Popup Info	Links
Drilling Date Robigus x Claire Experiment	Simon Griffiths						Study Info	• Study
DFWTKNIL Set 2	Simon Griffiths		2017-10-11	2018-08-04	View plots	Coppers Bawburgh GB NR9 3QL	Study Info	• Study • Field Trial
DFW Field Phenotyping Facility - Wheat	Rothamsted Research					Great Field 1/2 Phenotyping Area St Albans UK AL5 2GT	Study Info	• Study • Field Trial

Field Trials - Findable

... or via a text-based search web page

SEARCH FIELD TRIALS

A service to search field trial data

For more information and help, go to the [user documentation](#)

Simple options Advanced options

Search

toolkit

Type

Any

Page

0

Page size

10

Submit

Show 10 entries

Search:

Rank	Type	Title	Info	Link
1	Study	DFW Toolkit lines	Broad Mead UK MK43 0XF	View Study
2	Field Trial	DFW WP3 - DFW Academic Toolkit Trials	DFW WP3	View Field Trial
3	Field Trial	DFW WP3 - DFW Breeders Toolkit Trials	DFW WP3	View Field Trial
4	Field Trial	Andrew Riche - DFW Academic Toolkit RRes	Andrew Riche	View Field Trial
5	Study	DFW Academic Toolkit Trial H2019	Black Horse St Albans United Kingdom AL3 7PX	View Study
6	Study	DFW Toolkit lines 2nd year	Black Horse St Albans United Kingdom AL3 7PX	View Study
7	Study	DFW Academic Toolkit RRes Harvest 2020	Meadow, Rothamsted Experimental Farm Redbourn	View Study

Field Trials - Accessible

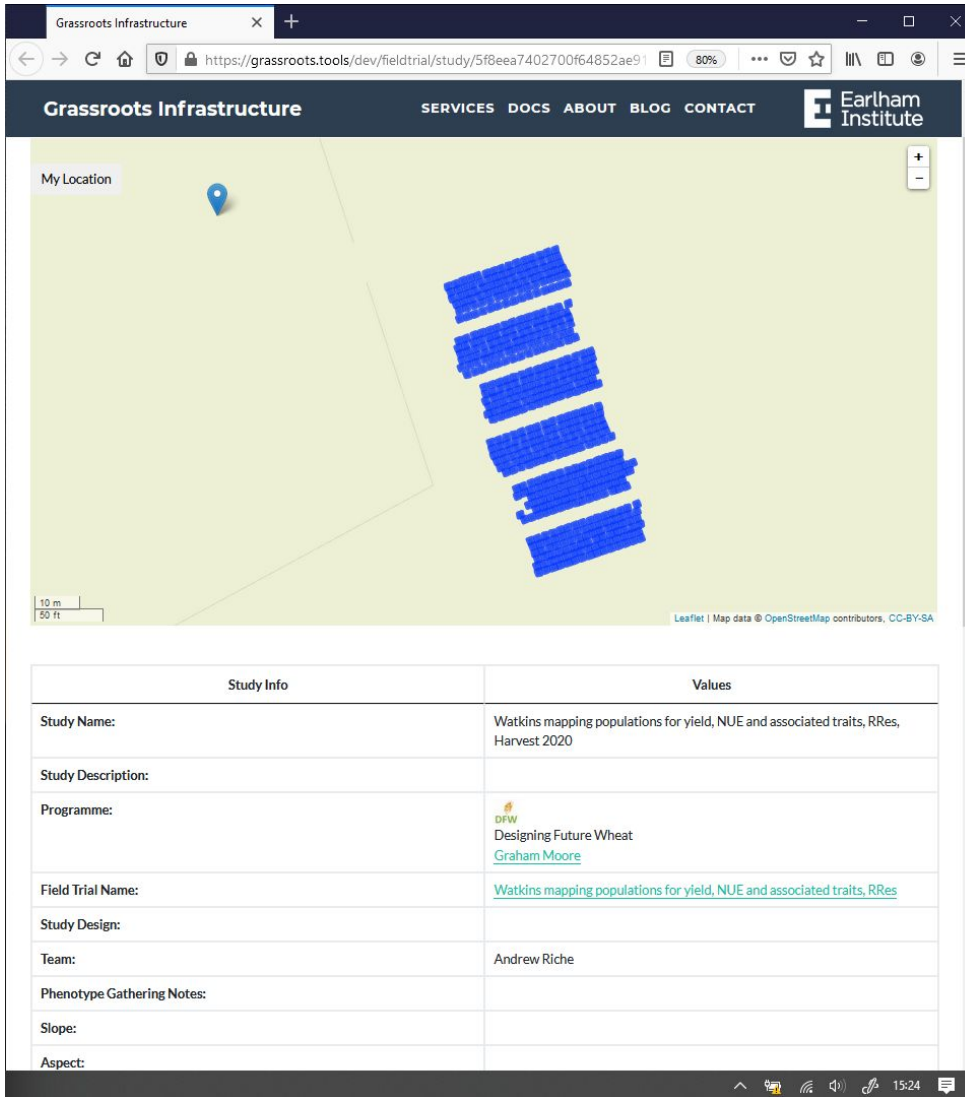
- All data is openly available
- All Field Trials, Studies, *etc.* have a unique identifier and are accessible through standard web technologies

Field Trials - Interoperability


The Field Trials data and metadata is exposed using both [BrAPI](#) which is a community-driven standardized RESTful web service API specification to enable interoperability among plant breeding databases.

```
▼ metadata:
  ▼ pagination:
    currentPage: 1
    pageSize: 44
    totalCount: 44
    totalPages: 1
    datafiles: []
    status: []
  ▼ result:
    ▼ data:
      ▼ 0:
        studyName: "1st vs 3rd wheat take-all resistance trial"
        studyDbId: "5dd8009ade68e75a927a8274"
        locationName: "Stackyard RES"
        locationDbId: "5d67a6f124ce205d7f6bbc53"
        ▼ additionalInfo:
          study_design: "Randomised block design"
          ▼ phenotype_gathering_notes: "Sponsors to take plant samples. Farm to record yields."
          ▼ trialName: "DFW - Designing Future Wheat - Work package 2 (WP2) - Added value and resilience"
          trialDbId: "5d5ac41c24ce20420b23322a"
      ▼ 1:
        studyName: "2017 DFW Paragon x Watkins Mapping Populations 6th Year"
        studyDbId: "5ef1d9de02700f433d408463"
        locationName: "Meadow, Rothamsted Experimental Farm"
        locationDbId: "5ef1dbb702700f447d624323"
        commonCropName: "wheat"
        startDate: "2016-10-19"
        endDate: "2017-08-15"
        active: "false"
        ▼ additionalInfo:
          study_design: "Split plot randomised & blocked"
          ▼ so:description: "7 PxW Mapping populations grown at 2 N levels plus 2 Robigus x Watkins mapping populations"
          ...
```

Field Trials - Plots Geolocations



The screenshot displays the Grassroots Infrastructure web application interface. At the top, there is a navigation bar with the logo and name 'Earlham Institute' and links for 'SERVICES', 'DOCS', 'ABOUT', 'BLOG', and 'CONTACT'. Below the navigation bar is a map showing a field trial layout with several blue rectangular plots. A blue location pin is placed on the map, labeled 'My Location'. A scale bar indicates 10 meters or 50 feet. The map is powered by Leaflet and OpenStreetMap data.

Study Info	Values
Study Name:	Watkins mapping populations for yield, NUE and associated traits, RRes, Harvest 2020
Study Description:	
Programme:	 DFW Designing Future Wheat Graham Moore
Field Trial Name:	Watkins mapping populations for yield, NUE and associated traits, RRes
Study Design:	
Team:	Andrew Riche
Phenotype Gathering Notes:	
Slope:	
Aspect:	

The geolocations of each plot within a study, coupled with automatic location updating, allows the scientists to walk around a study and see which plot they are within in real time.

Field Trial – Reusable data

Plot data is standardized using ontological terms for each plot

PLOT DETAILS

Row: 20
Column: 1
Length: 3.594m
Width: 1.8m
Study Design:
Sowing Date: 2019-10-30
Harvest Date: 2020-08-10
Treatment:
Comment: Slight height segregation



Replicate	Rack	Accession	Pedigree	Gene Bank	Links
1 (Current Plot)	1	DFW SEL 0208		Germplasm Resources Unit	
3 (Plot Row:3 - Col:23)	1	DFW SEL 0208		Germplasm Resources Unit	
2 (Plot Row:14 - Col:15)	1	DFW SEL 0208		Germplasm Resources Unit	

PHENOTYPES


Close

Field Trials - Plot Phenotypes

The screenshot shows the Grassroots Infrastructure web application. The main content area displays a 'PHENOTYPES' table with the following data:

Replicate	Rack	Date	Raw Value	Corrected Value	Trait	Measurement	Unit
2 (Current Plot)	1		0		Grain filling period	GFP pdt Computation	day
2 (Current Plot)	1		0		Anthesis thermal time	TTA Computation	°C day
2 (Current Plot)	1		0		Physiological maturity thermal time	TTM Computation	°C day
2 (Current Plot)	1		0.989712219		Plant height	PH Measurement	cm
2 (Current Plot)	1		6.762351485		Grain yield	GY Computation	t/ha

Below the phenotypes table is a 'Study Info' table:

Study Info	Values
Study Name:	Watkins mapping populations for yield, NUE and associated traits, RRes, Harvest 2020
Study Description:	
Programme:	 Designing Future Wheat Graham Moore
Field Trial Name:	Watkins mapping populations for yield, NUE and associated traits, RRes
Study Design:	
Team:	Andrew Riche
Phenotype Gathering Notes:	
Slope:	
Aspect:	

Phenotypes stored as

- Trait
 - What to Measure
- Method
 - How it was measured
- Unit
- Value(s)
- Date

All of these are well-defined terms from the [Crop Ontology](#)

Field Trials - Exposing Data

All of the data and metadata are available via Web Service APIs

- Grassroots
- Partial BrAPI support

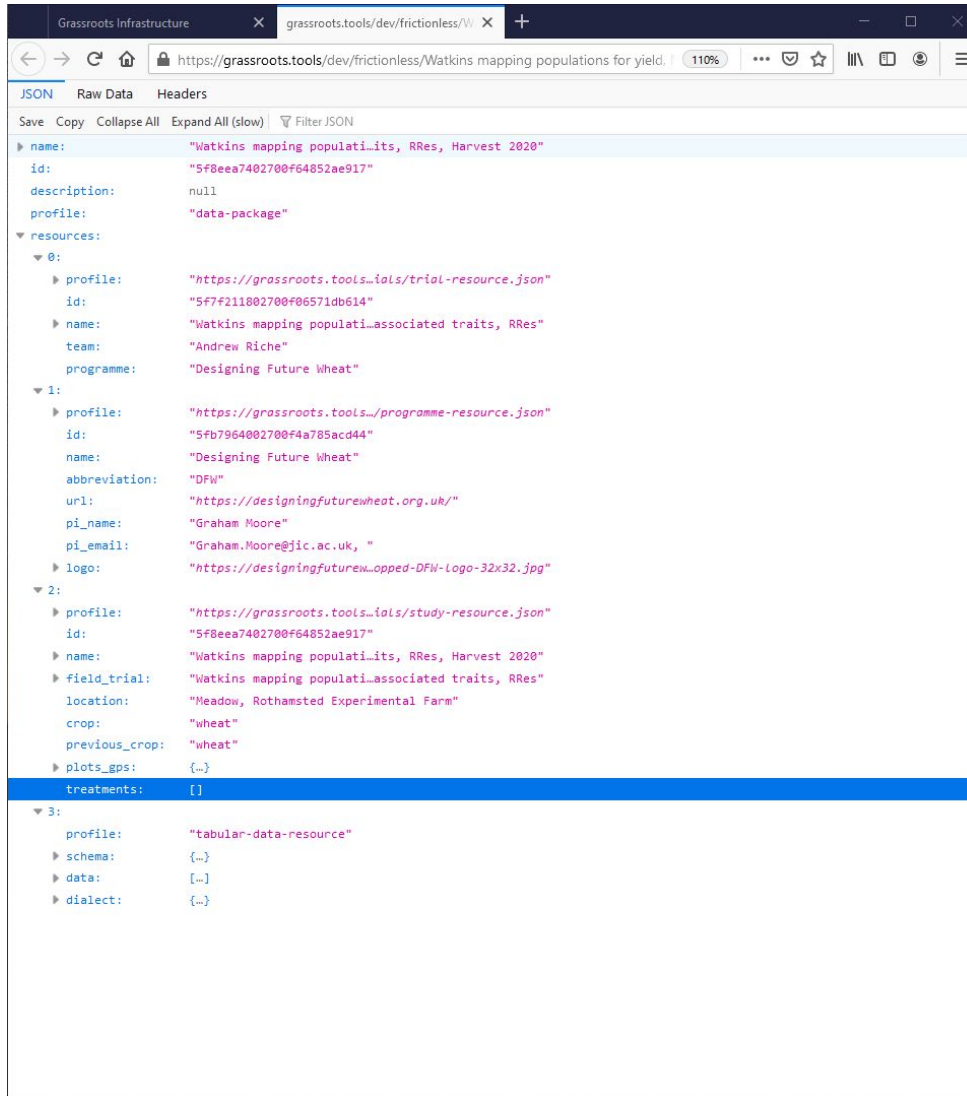
Field Trials - Exposing Data

Q: APIs work for people comfortable scripting and programming, but what about people who just want the basic data...

A: Frictionless Data!

- Grassroots Schemas published at <https://grassroots.tools/frictionless-data/>
- Other DFW work on Frictionless Data by Richard Ostler at Rothamsted Research

Field Trials - Study Frictionless Data Package

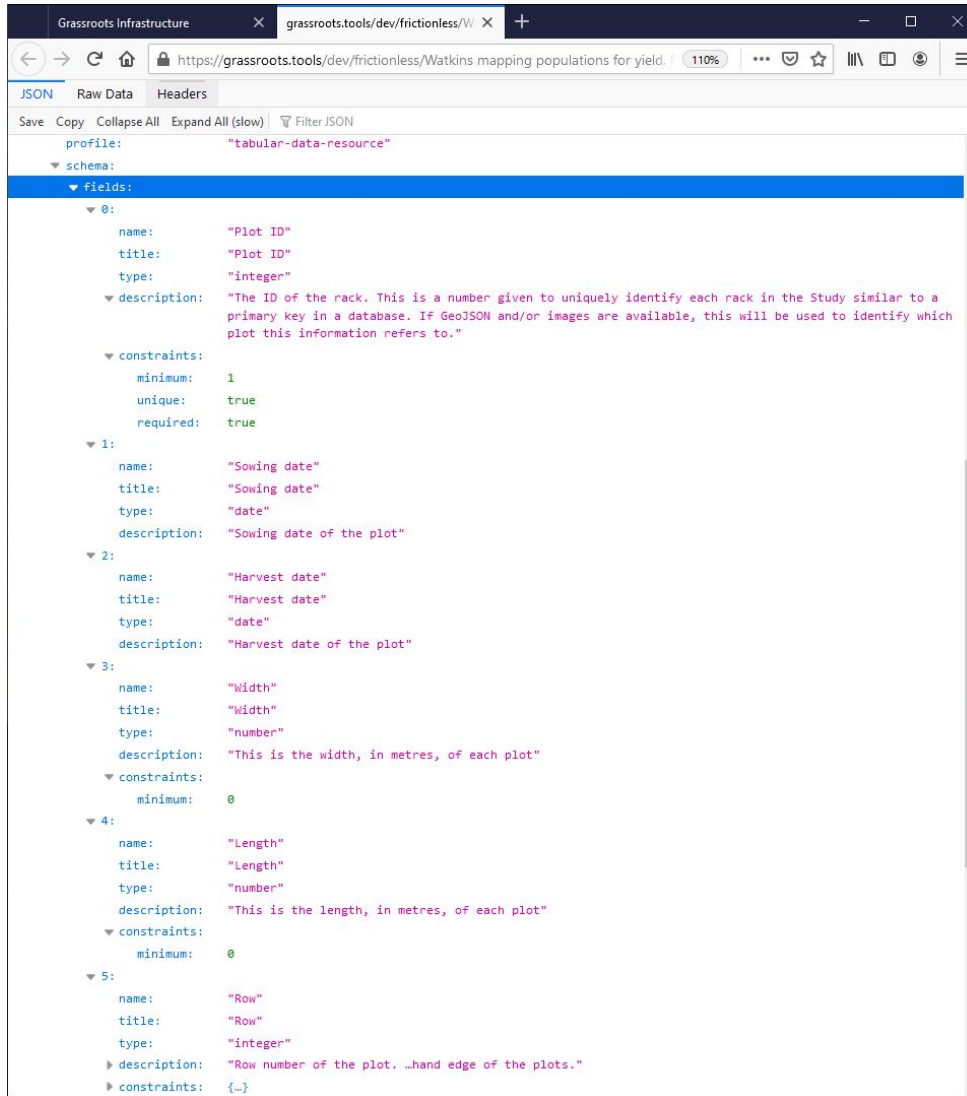


The screenshot shows a web browser window displaying a JSON data package. The URL is [https://grassroots.tools/dev/frictionless/Watkins mapping populations for yield](https://grassroots.tools/dev/frictionless/Watkins%20mapping%20populations%20for%20yield). The JSON structure is as follows:

```
{
  "name": "Watkins mapping populati...its, RRes, Harvest 2020",
  "id": "5f8eaa7402700f64852ae917",
  "description": null,
  "profile": "data-package",
  "resources": [
    {
      "profile": "https://grassroots.tools...ials/trial-resource.json",
      "id": "5f7f211802700f06571db614",
      "name": "Watkins mapping populati...associated traits, RRes",
      "team": "Andrew Riche",
      "programme": "Designing Future Wheat"
    },
    {
      "profile": "https://grassroots.tools.../programme-resource.json",
      "id": "5fb7964002700f4a785acd44",
      "name": "Designing Future Wheat",
      "abbreviation": "DFW",
      "url": "https://designingfuturewheat.org.uk/",
      "pi_name": "Graham Moore",
      "pi_email": "Graham.Moore@jic.ac.uk, ",
      "logo": "https://designingfuturew...apped-DFW-Logo-32x32.jpg"
    },
    {
      "profile": "https://grassroots.tools...ials/study-resource.json",
      "id": "5f8eaa7402700f64852ae917",
      "name": "Watkins mapping populati...its, RRes, Harvest 2020",
      "field_trial": "Watkins mapping populati...associated traits, RRes",
      "location": "Meadow, Rothamsted Experimental Farm",
      "crop": "wheat",
      "previous_crop": "wheat",
      "plots_gps": {...},
      "treatments": []
    },
    {
      "profile": "tabular-data-resource",
      "schema": {...},
      "data": [...],
      "dialect": {...}
    }
  ]
}
```

All of the details of each individual study are stored in a single Frictionless Data Package

Field Trials - Plots Schema Fields



```
profile: "tabular-data-resource"
schema:
  fields:
    0:
      name: "Plot ID"
      title: "Plot ID"
      type: "integer"
      description: "The ID of the rack. This is a number given to uniquely identify each rack in the Study similar to a primary key in a database. If GeoJSON and/or images are available, this will be used to identify which plot this information refers to."
      constraints:
        minimum: 1
        unique: true
        required: true
    1:
      name: "Sowing date"
      title: "Sowing date"
      type: "date"
      description: "Sowing date of the plot"
    2:
      name: "Harvest date"
      title: "Harvest date"
      type: "date"
      description: "Harvest date of the plot"
    3:
      name: "Width"
      title: "Width"
      type: "number"
      description: "This is the width, in metres, of each plot"
      constraints:
        minimum: 0
    4:
      name: "Length"
      title: "Length"
      type: "number"
      description: "This is the length, in metres, of each plot"
      constraints:
        minimum: 0
    5:
      name: "Row"
      title: "Row"
      type: "integer"
      description: "Row number of the plot. ...hand edge of the plots."
      constraints: { }
```

The data for the study's plots is tabular with dynamically generated schemas

Standard attributes

- Length
- Width
- Position
- *etc.*

Custom attributes

- Treatments *e.g.* fertilizers
- Phenotypes

Field Trials - Plots Tabular Data

```
JSON Raw Data Headers
Save Copy Collapse All Expand All (slow) Filter JSON
s:
  profile: "tabular-data-resource"
  schema:
    fields: [-]
    title: "Plots"
  data:
    0:
      Row: 1
      Column: 1
      Length: 1
      Width: 1
      Sowing date: "2019-11-21"
      Harvest date: "2020-08-10"
      Plot ID: 1
      Rack: 1
      Accession: "(Paragon x WATDE0021)_0043"
      Replicate: 1
      Mat_dto_day: "2020-08-02 (2020-08-02)"
      GFP_calc_day: "48.4"
      TTA_Calc_Cday: "1752.15"
      TTM_calc_Cday: "2588.85"
      PH_M_cm: "0.87671608"
      GY_Calc_tha: "6.365641345"
      HI_Calc_pct: "42.60693941"
      BM_Calc_tha: "14.94038631"
      Awms_E_0to9: "2020-07-06 (0)"
    1:
      Row: 1
      Column: 2
      Length: 1
      Width: 1
      Sowing date: "2019-11-21"
      Harvest date: "2020-08-10"
      Plot ID: 30
      Rack: 1
      Accession: "(Paragon x WATDE0021)_0014"
      Replicate: 1
      Ant_dto_day: "2020-06-17 (2020-06-17)"
      Mat_dto_day: "2020-07-24 (2020-07-24)"
      GFP_calc_day: "37"
      TTA_Calc_Cday: "1802.3"
      TTM_calc_Cday: "2418.25"
      PH_M_cm: "0.811513824"
      GY_Calc_tha: "6.060947109"
```

Field Trials - Exposing Data

So people can download a field trial as a Frictionless Data Package and Frictionless comes with a great API to extract the data

Can we make it simple for them to unpack the data?

Field Trials - Exposing Data

A tool to extract the resources within a Frictionless Data Package

- Downloads and uses the schemas specified by the profile value of each Data Resource
- Converts Data Resources
 - Markdown
 - HTML
- Converts Tabular Data Resources
 - CSV
- Cross Platform
- Not wheat-specific, works with any Data Package

Pre-release available at

<https://github.com/TGAC/grassroots-client-frictionless-data/>

Further Work

- Add more information to the generated Frictionless Data Packages
- Add Machine Learning to detect phenotypic values from media such as photos taken by drones
- Further collaboration with Richard Ostler refining a common Frictionless Data standard for both single-year and long-term field trial wheat studies

Acknowledgements

- Earlham Institute
 - Alice Minotto
 - Felix Shaw
 - Anthony Etuk
 - Nicola Soranzo
 - Martin Ayling
 - Catherine Hunter
 - Anil Thanki
 - Jon Wright
 - Luis Yanes
 - Anthony Hall
 - Laura Jayne Gardiner
 - Ben White
 - Joshua Ball
- John Innes Centre
 - Luzie Wingen
 - Simon Griffiths
- Rothamsted Research
 - Andrew Riche
 - Chris Rawlings
 - Richard Ostler
- University of Bristol
 - Paul Wilkinson
 - Mark Winfield
 - Gary Barker

And a BIG thank you to the csv,conf organizers!

SIMON TYRRELL

Research Software Engineer

simon.tyrrell@earlham.ac.uk



Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK
www.earlham.ac.uk



Decoding Living Systems