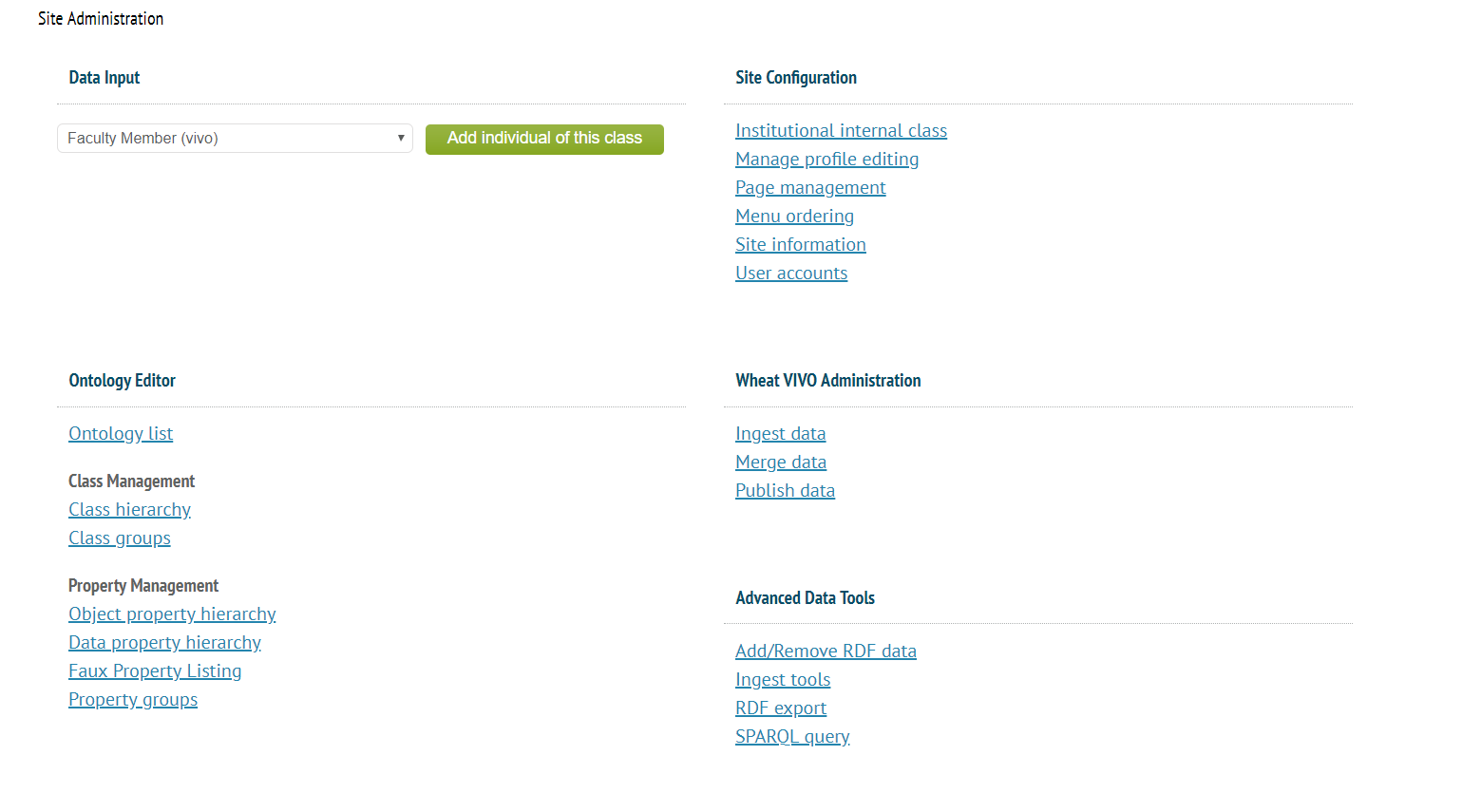
# WheatVIVO Administator’s Guide

Last updated 02 November 2018

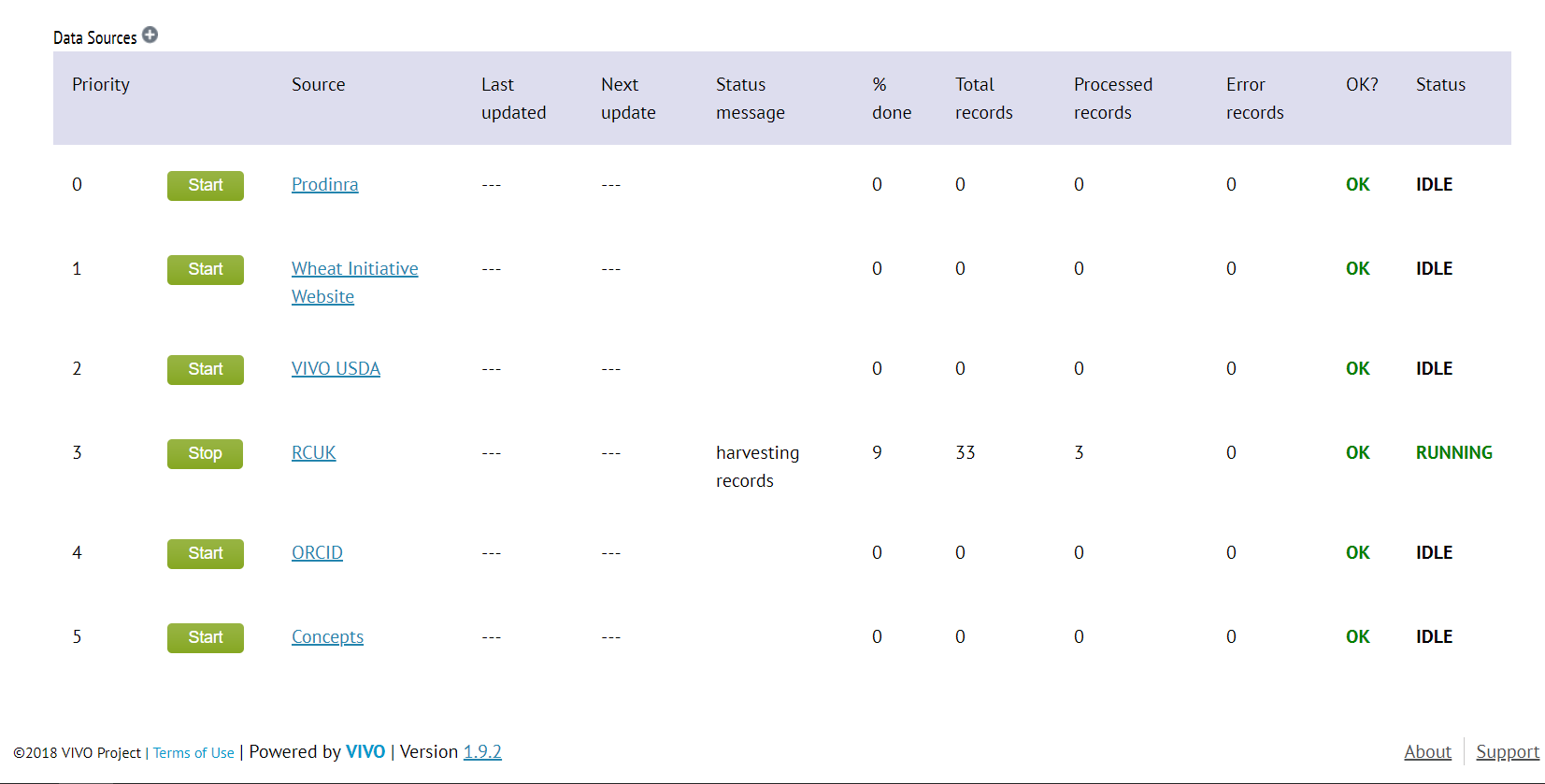


The Site Admin page of WheatVIVO includes a **Wheat VIVO Administration** menu with links for ingesting, merging (reconciling) data and publishing data to the public WheatVIVO portal.

## Ingest data

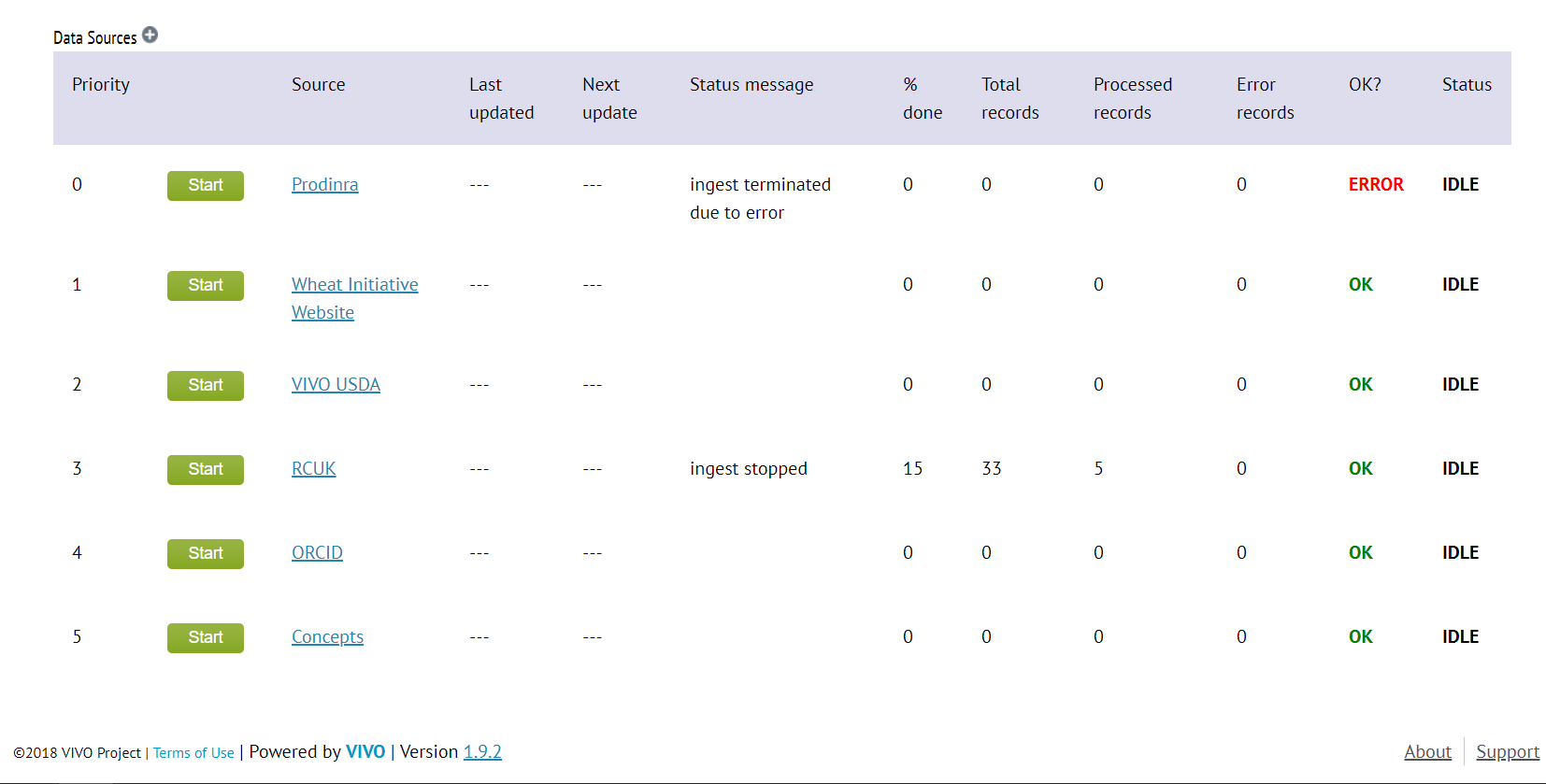


The Ingest Data screen lists the available sources and the status of the connector for each source. The Start/Stop buttons allow connectors to be run individually from this screen:

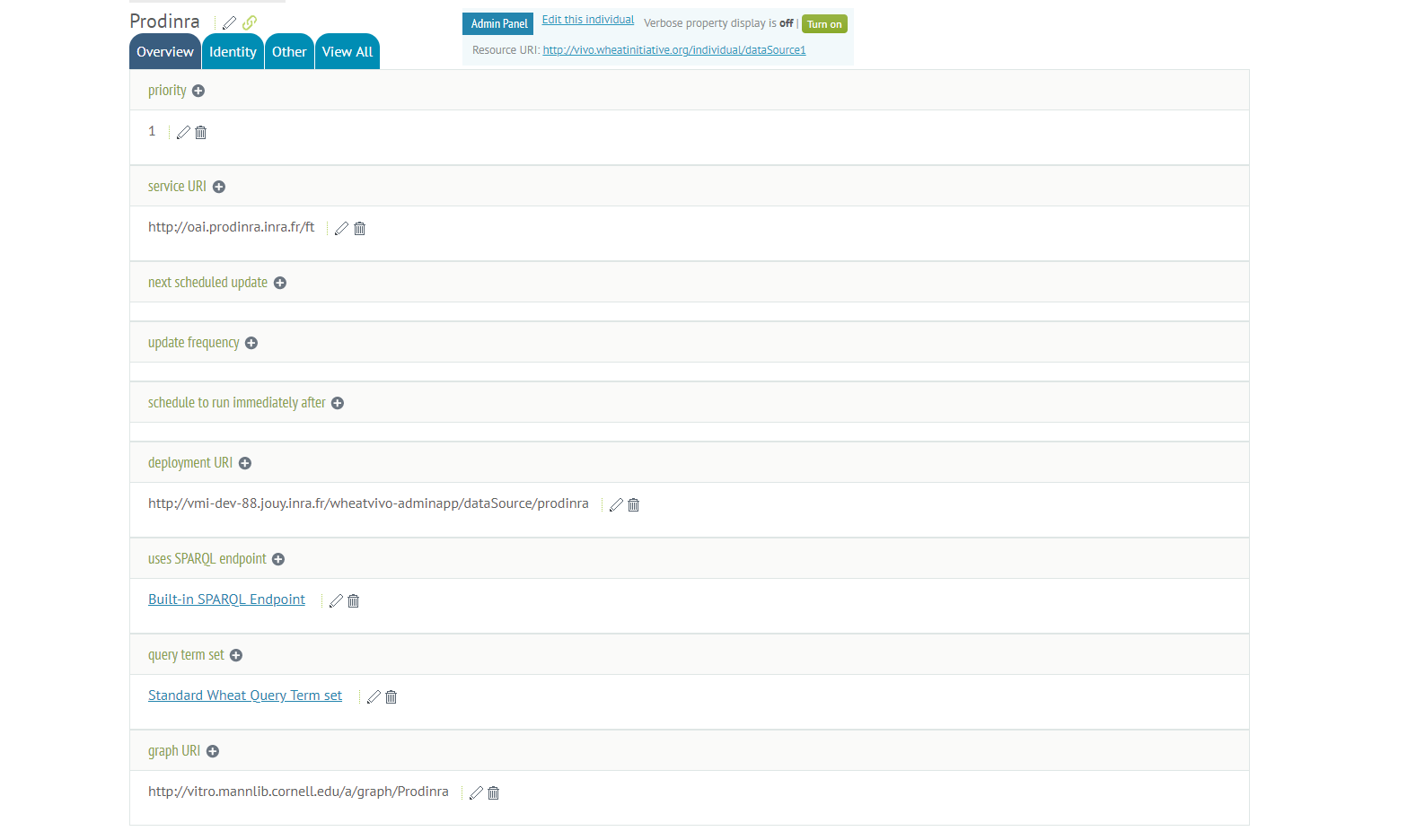


As individual records are retrieved and converted, the „Processed records” count is incremented. Most connectors provide a count of the total records to be processed: in these cases, the „% done” column will also be updated as the ingest progresses. In the event that individual records are unable to be retrieved or converted properly but the ingest is otherwise able to continue, the „Error records” column will be incremented. Clicking on the count will show more detailed error messages.

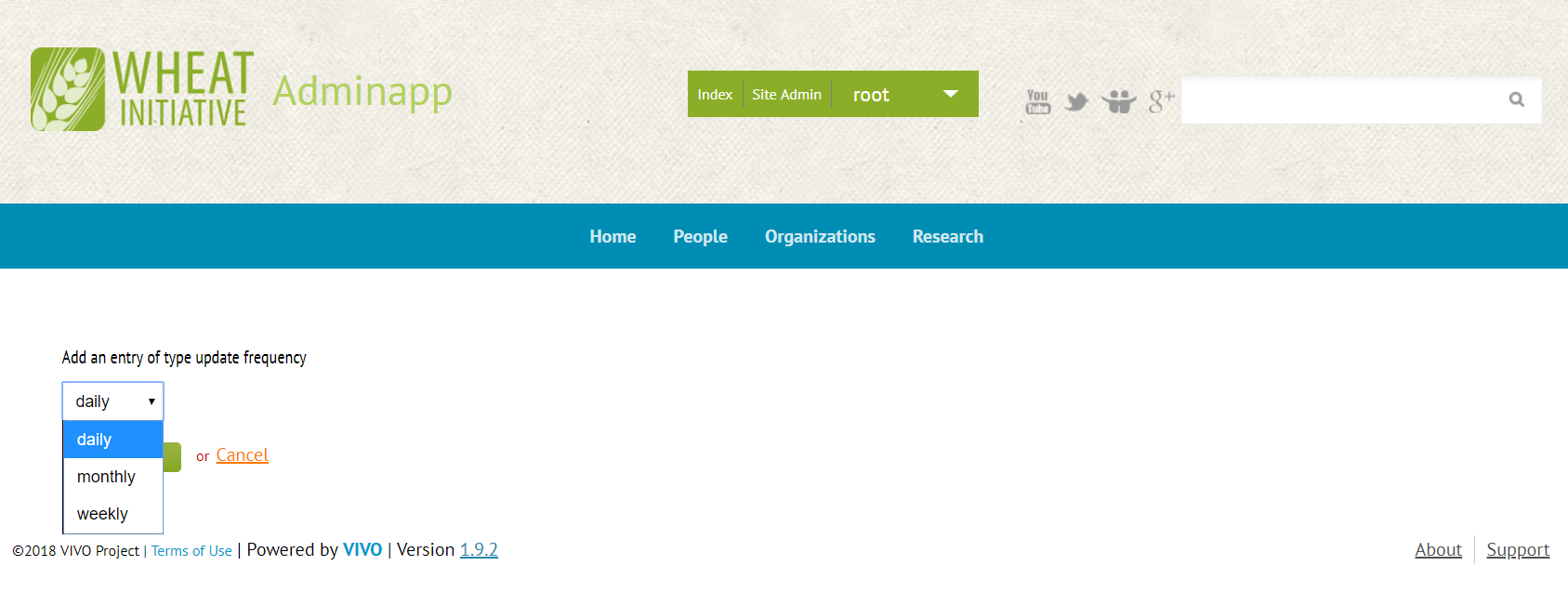
If a fatal error prevents an ingest from completing (such as when a source is unavailable), the OK? column will display a red ERROR alert and the status message will be updated to show the immediate cause of the error. The file wheatvivo-adminapp.all.log in Tomcat’s logs directory may include additional details.



Clicking on the name of a data source will reveal additional properties that can be edited:



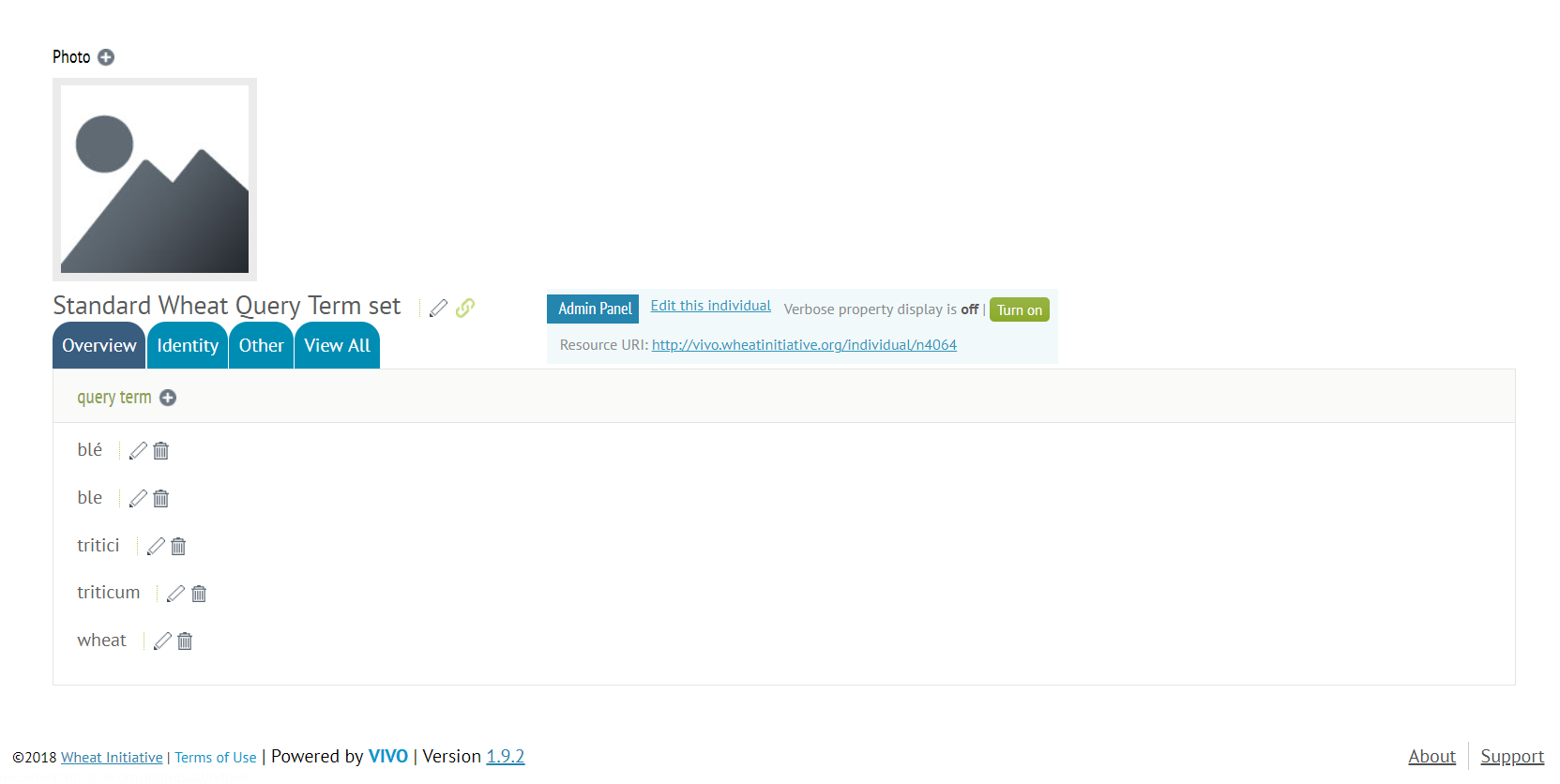
These properties allow a connector to be set to run automatically at a certain time and (optionally) repeating on a daily, weekly or monthly interval:



In addition, the „schedule to run immediately after” property allows a connector to immediately follow another in a serial chain. This may simplify scheduling and reduce server load by avoiding inadvertently scheduling connectors to run at the same time.

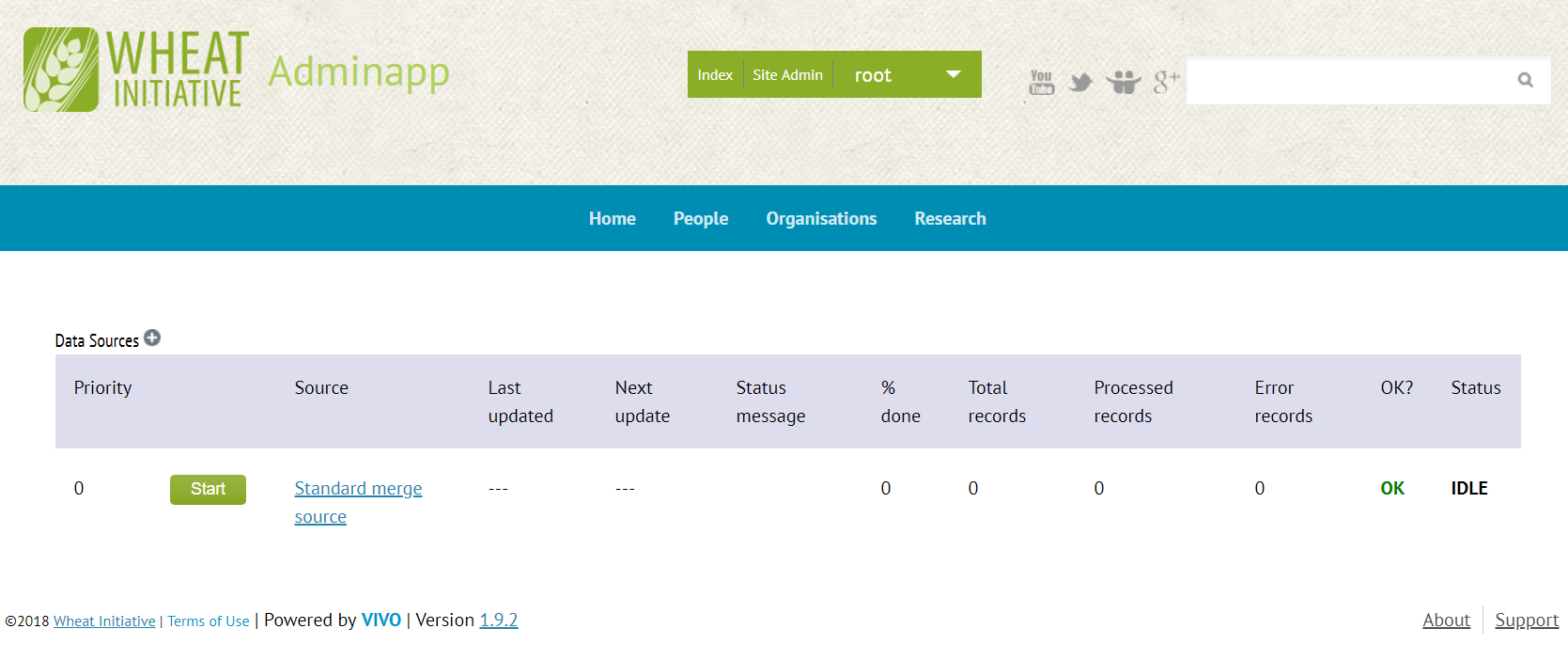
### Query Term Set

The search terms used by the connector to find wheat-relevant data can be set by editing the „query term set” for the connector. Individual connectors may use their own query term sets, or share a common set:

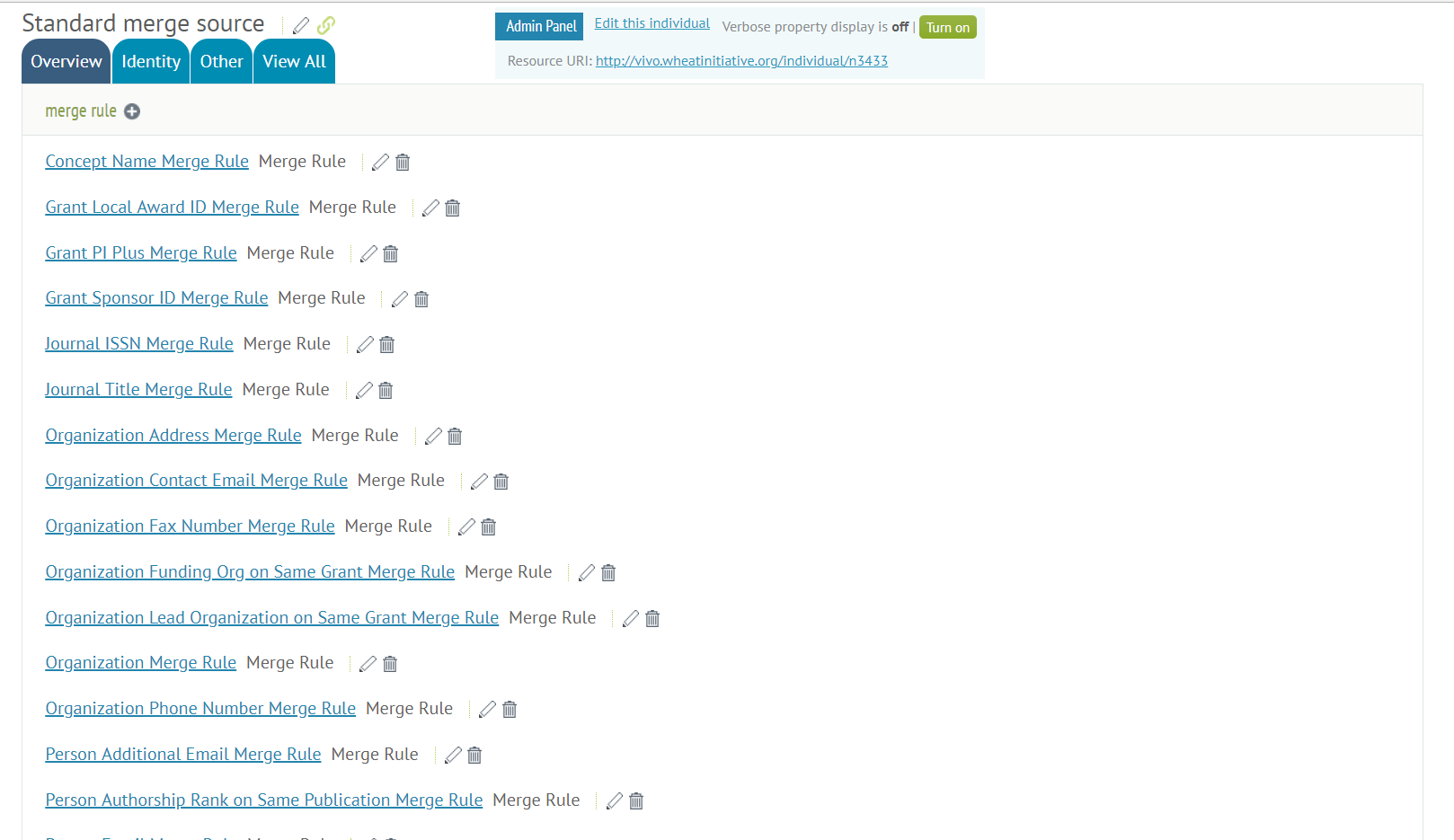


## Merge Data

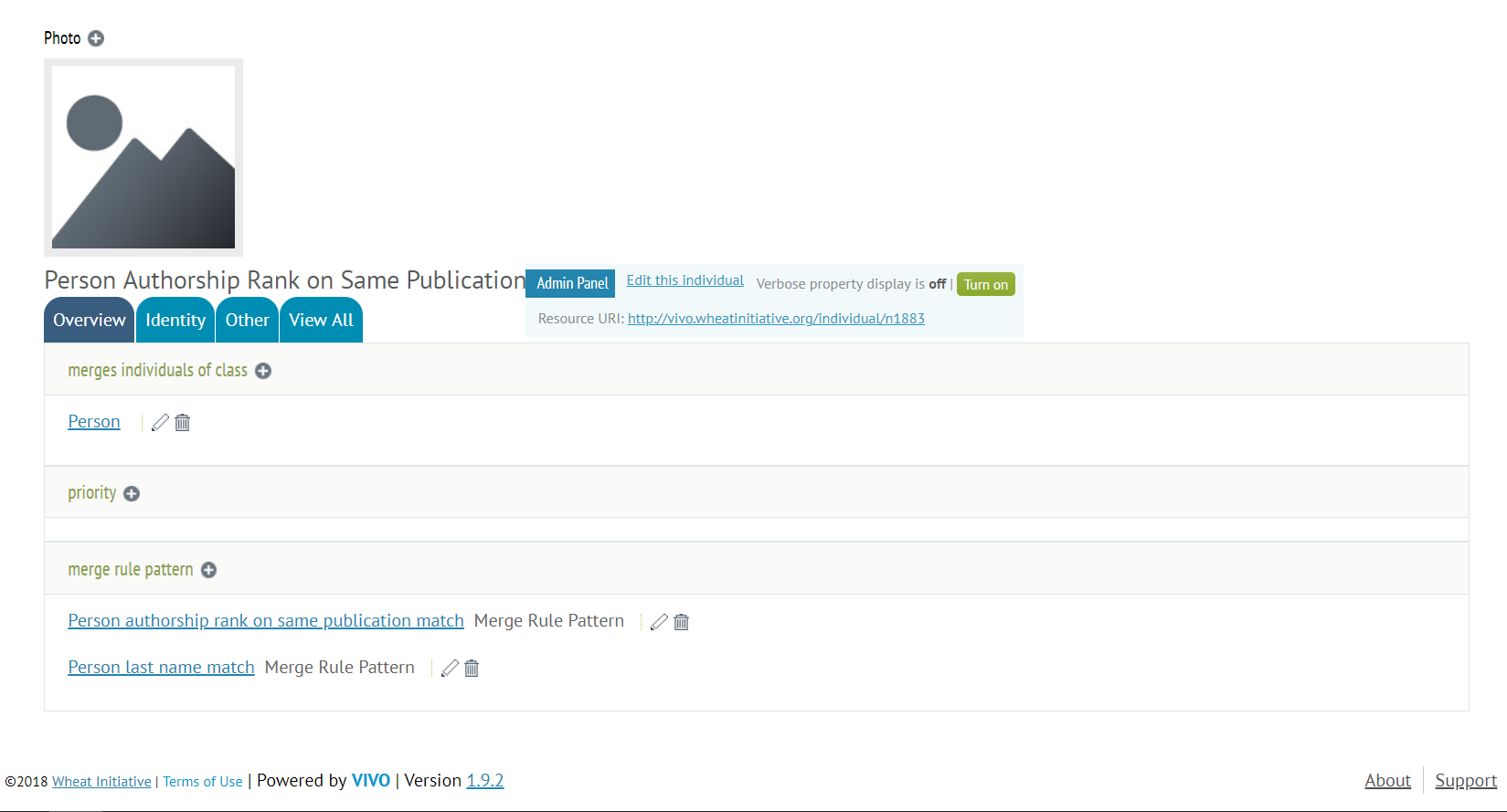
The merge data (reconciliation) process automatically adds „sameAs” assertions to link together entities harvested from different sources that represent the same real-world objects. Merging is done according to sets of rules that can be configured from the interface:



The merge process is set up just like the data connectors and has similar controls. Instead of retrieving data from remote sources, however, it queries the data already gathered in the admin app and adds sameAs triples to the same application.

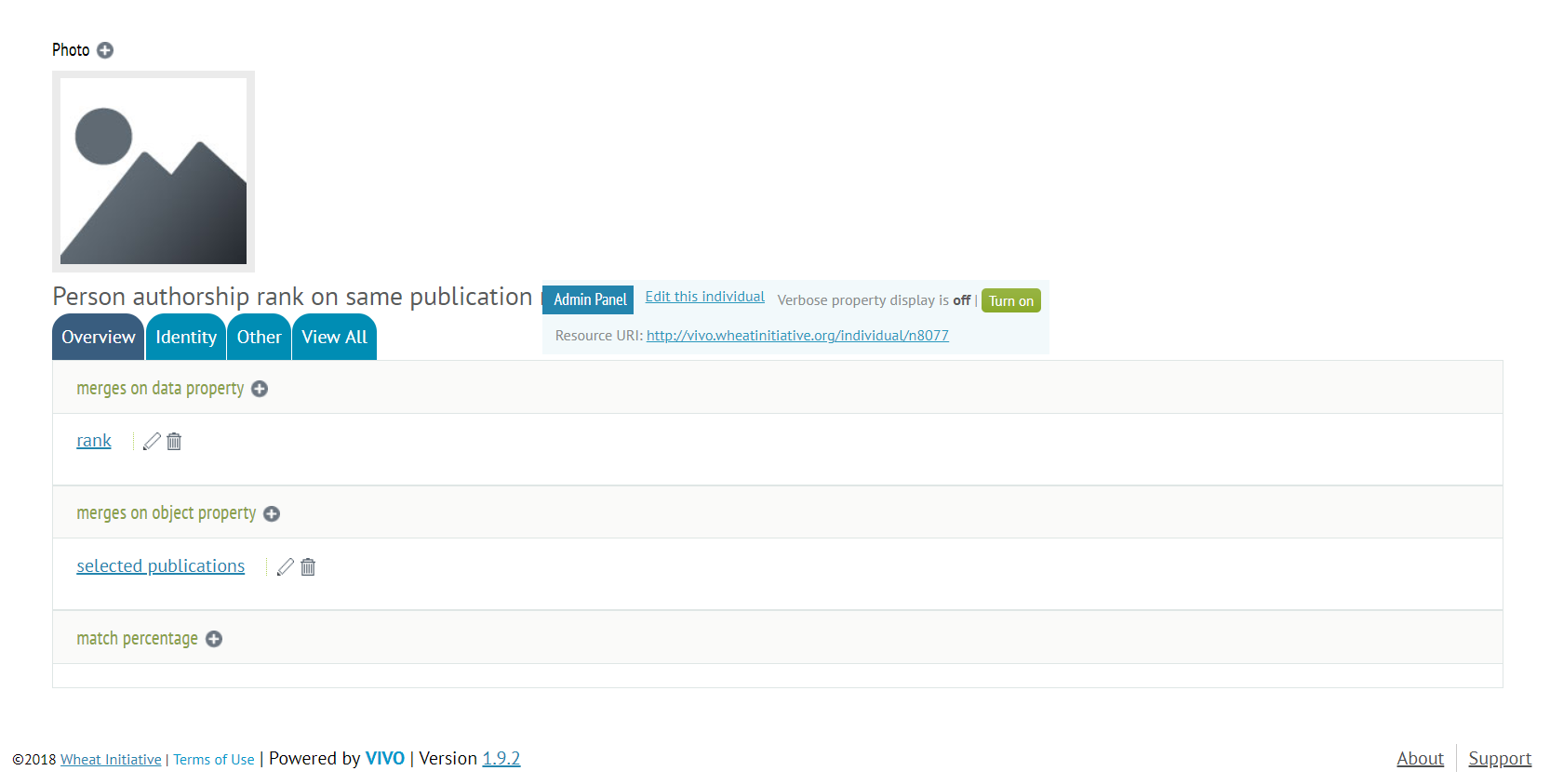


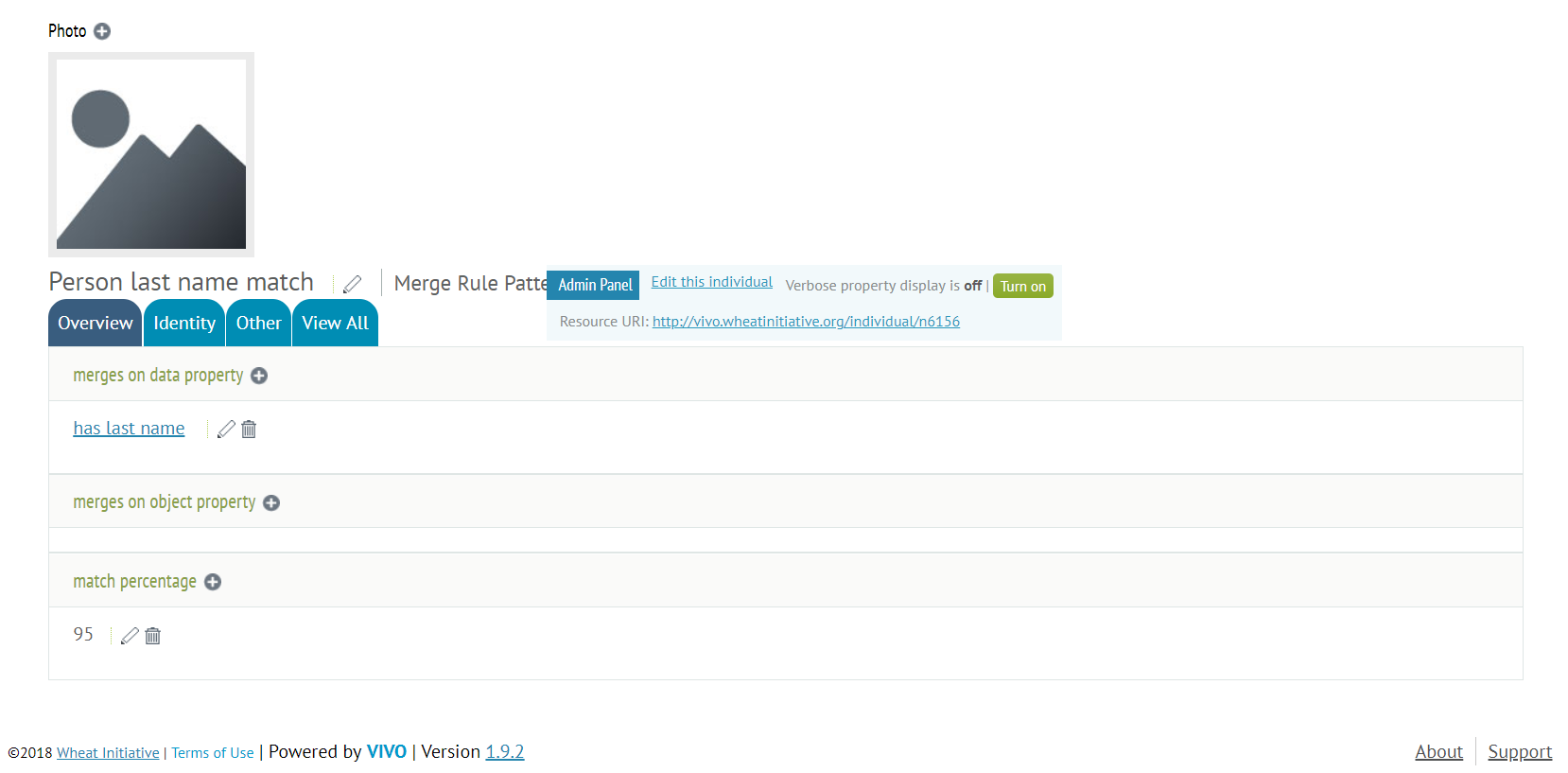
The standard merge source includes a default set of merge rules. Additional rules can be added by clicking the + sign next to „merge rule.”



Each rule adds sameAs triples between resources of a given ontology class. This class is specified in the „merges individuals of class” property. The rule above merges resources of type foaf:Person.

The resources are inferred to be sameAs one another according to patterns added via the „merge rule pattern” property. In the case above, the rule finds persons who have the same author rank number on the same publication and adds a safety check by requiring that the persons’ last name value must match. Each rule may have any number of patterns. Only those resources who satisfy **all** of the merge rule patterns will be inferred to be sameAs one another.



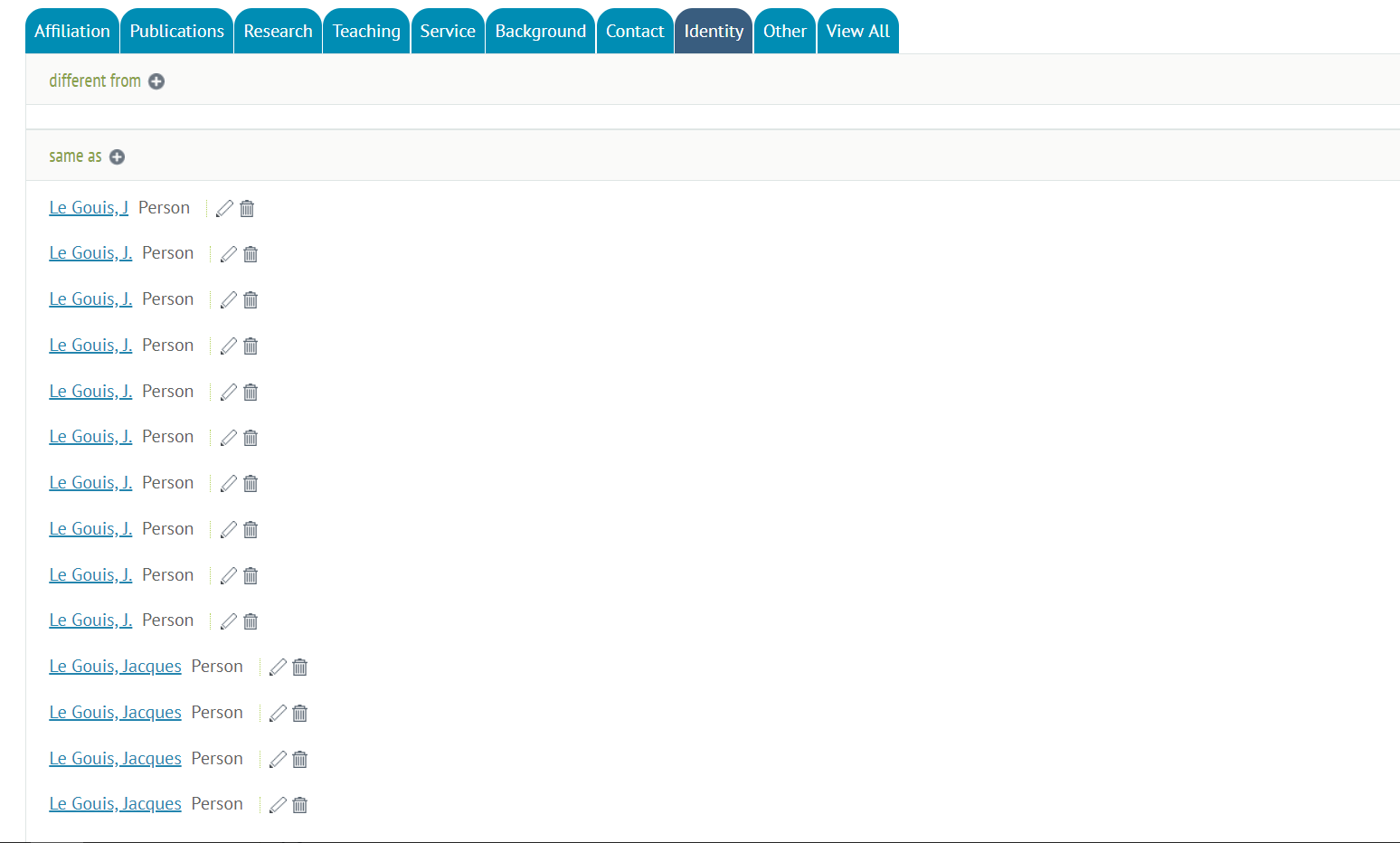


When matching on data properties with string values (e.g. last name), the „match percentage” (match score on a scale from 1-100) property sets the threshold for accepting inexact matches. Any value less that 100 will allow for example, case-insensitive matches. Leaving the score at 100 (or not specifying it at all) will make for faster rule execution in cases where truly exact matches are desired.

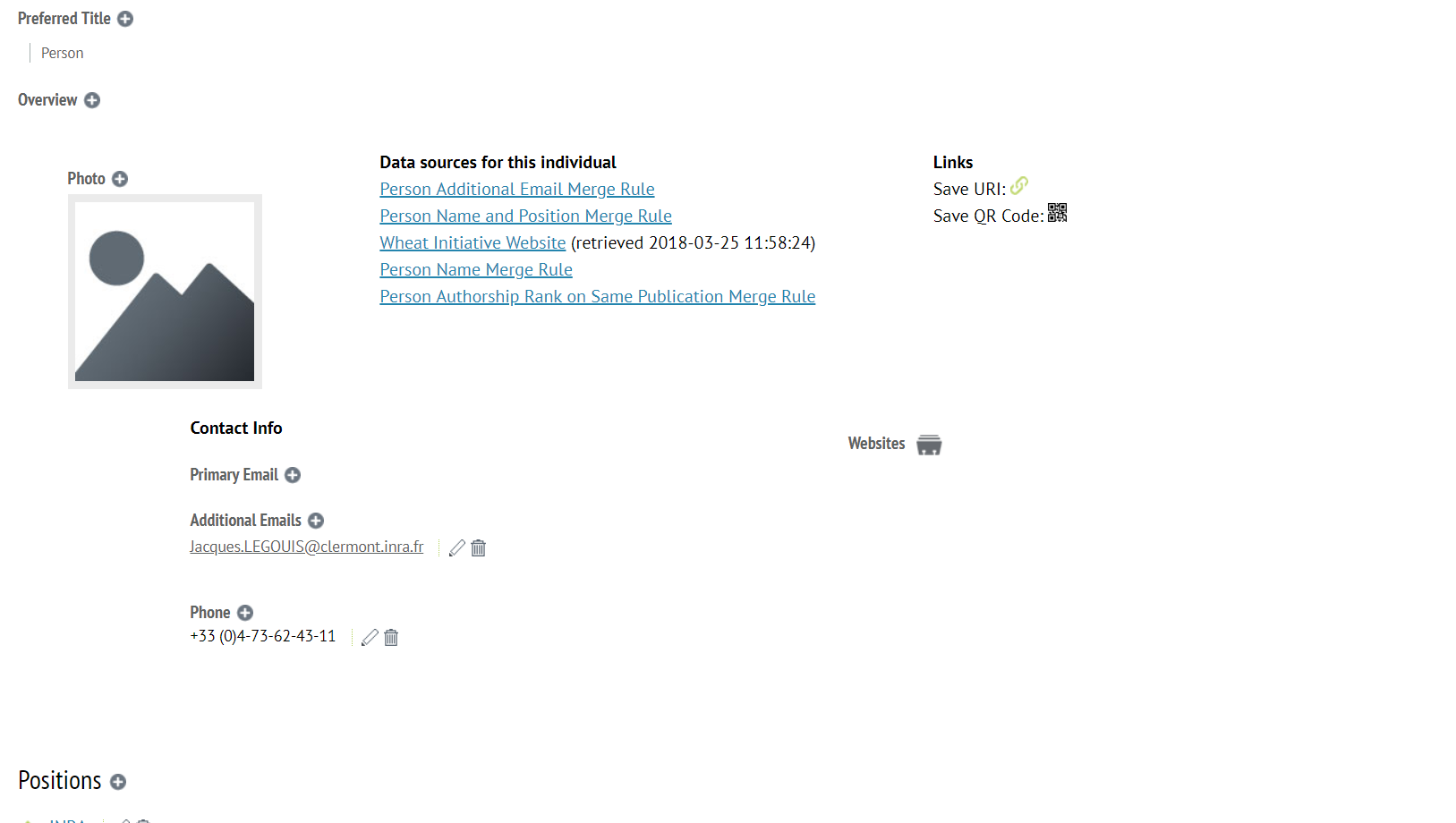
There is also currently a magic match percentage of 19. Using a setting of 19 will match string according to their initials: E.g. „J. A.” and „John Anderson” will be considered a match under a score of 19.

### Invalidating Proposed Merges

After the merge data source has run, the inferred sameAs links will appear in the admin app interface:



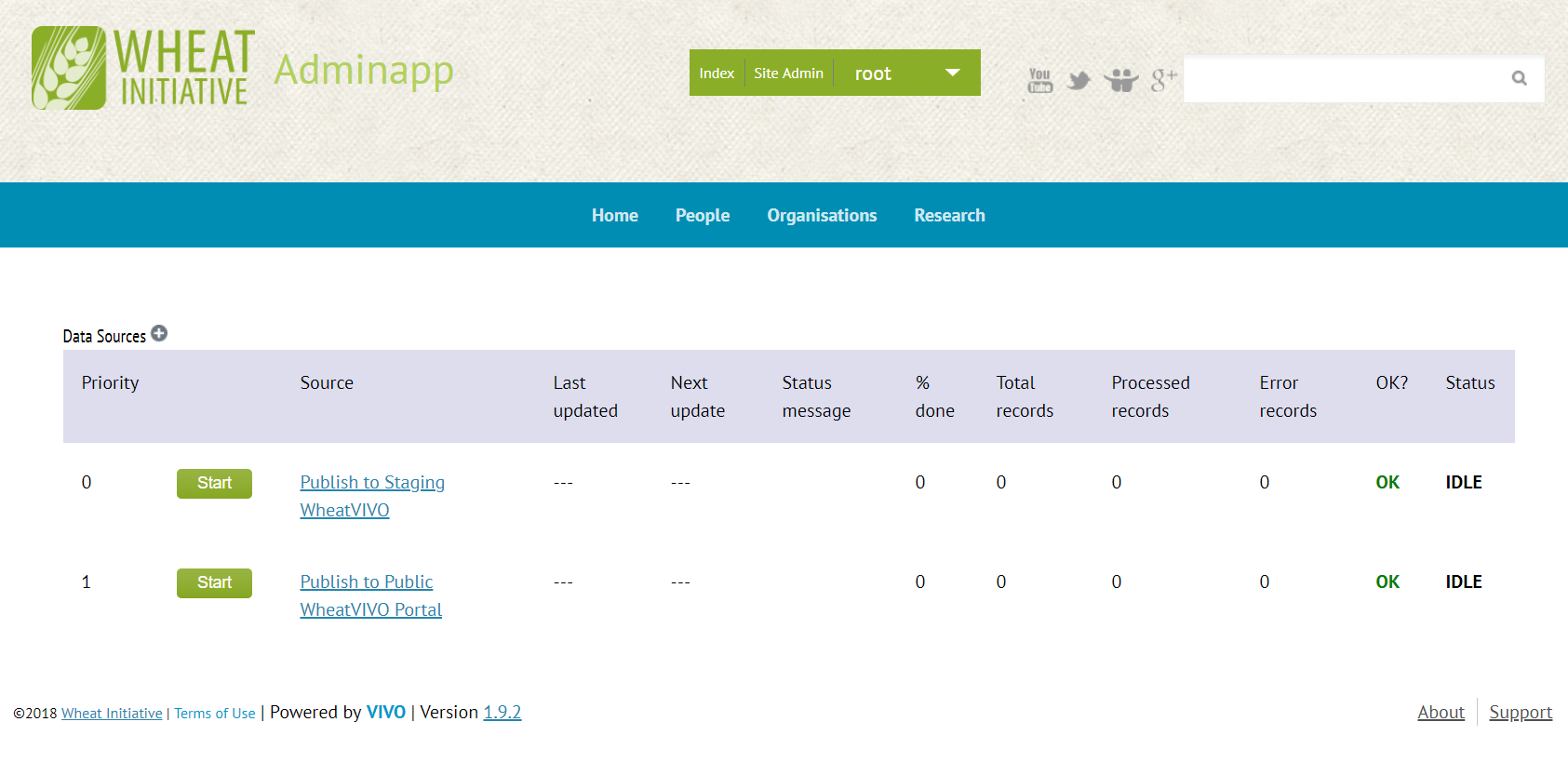
If a sameAs triple was added erroneously, there are two methods for addressing the problem. First, a provenance display will show the rules responsible for inferring sameAs triples for a given individual:



These merge rules may be revisited and adjusted. If the rules are generally sound, but certain troublesome inferences simply need to be invalidated, „different from” links may be manually added (just above „same as” in the interface) to show that the merge source should not attempt to merge the two given individuals. After adding different from assertions, re-run the merge source and the offending inferences will be removed.

## Publish Data

The publish step is run after the merge and takes the resulting sameAs triples to combine resources into a single representation of a real-world entity with a single URI. The URI is chosen according to the priority property of the original data source. Similarly, properties for which only one value is desired (e.g. labels) will have their values selected according to source priority.

 The publisher copies the combined resources to either a staging/testing or public WheatVIVO server. It may be configured and scheduled similarly to the other data sources.