

GDSR - FAIR

FAIR Data Standards in the GDSR

H.Mason, N.Lasierra, D.Clement, I.Robinson

huw.mason@roche.com | nelia.lasierra@roche.com | didier.clement@roche.com | ivan.robinson@roche.com

How does the GDSR support FAIR Data Standards?

The Data standards available in the GDSR contribute to the I of the FAIR principles, enabling Interoperability. To ensure that data across multiple sites can be easily combined we know that we need to use data standards and controlled vocabularies. Having common formal vocabularies for describing (meta) data assets ensures common understanding and facilitates integration and sharing for machines. FAIRness is also required for our data standards. The GDSR platform relies on RDF, ontologies and the more general Linked Data principles as instruments to enable our data standards to be FAIR.

From the Roche FAIR Data playbook:

“FAIR data are best stored in a system that provides unique data asset IDs, tracking and management of current and historical versions, and facilities for structured annotation with metadata. Suitable systems must support an organized set of annotation tags and metadata that describe and characterize data and permit easy inspection and querying of the metadata by individuals and by other computing systems.”

F - URIs are used to describe every concept registered in the GDSR

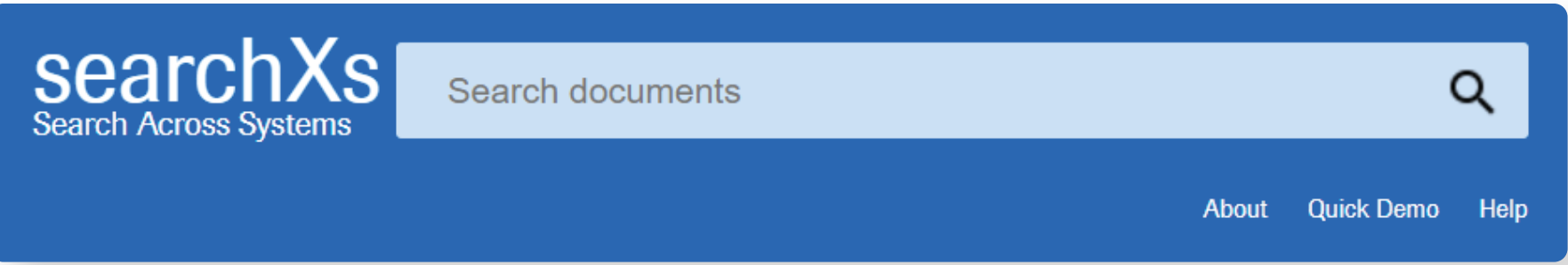
GDSR uses RDF and ontologies to describe the metadata and standard models. Hence, each concept is identified using a URI.

<<http://gdsr.roche.com/cdisc/adamig#Table.ADSL>>

Metadata about the models is provided by a catalog, giving users a summary by publication.

Reports give detail to the components of a model.

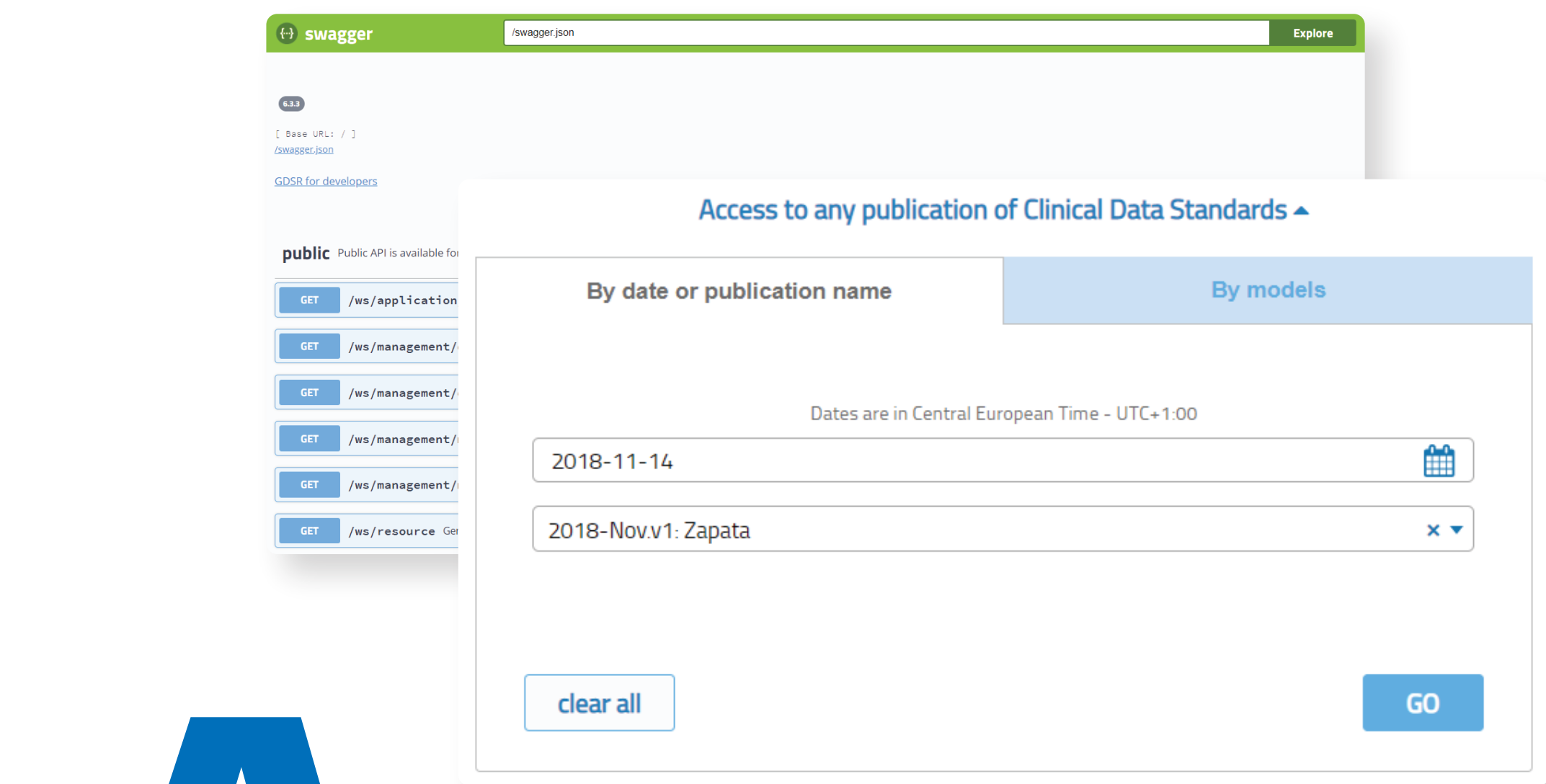
GDSR browser supports indexed searching.



A - Access is available through REST API, browser and other endpoints

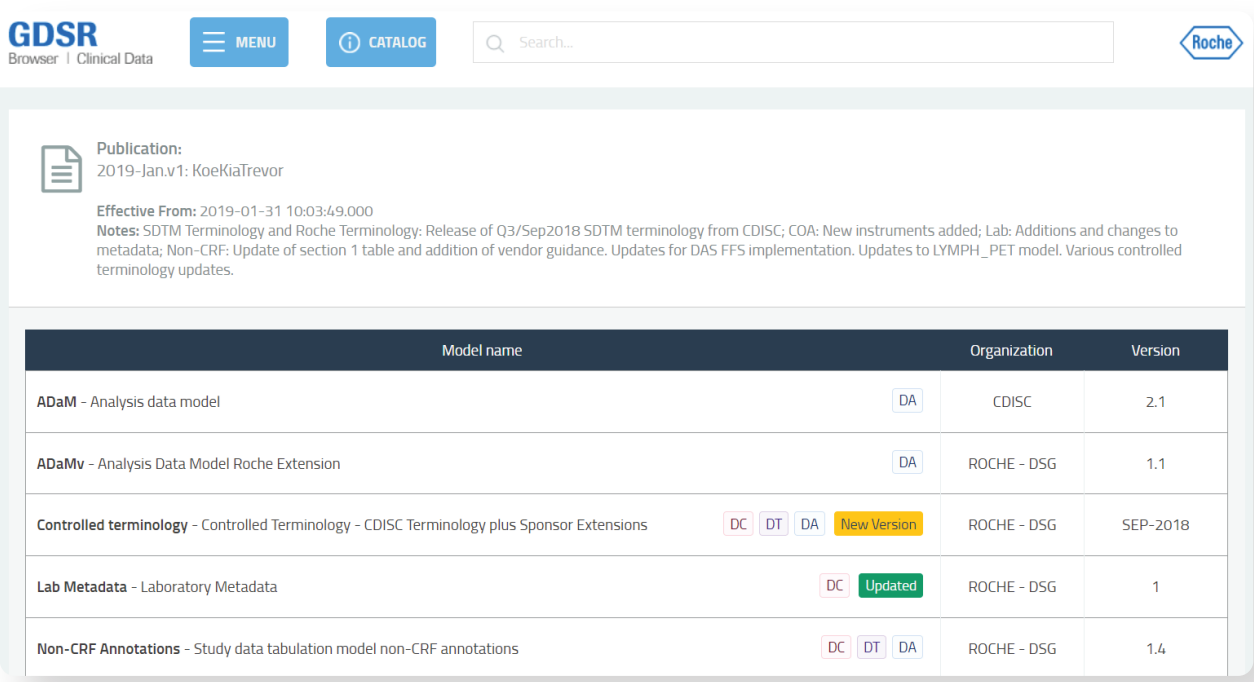
GDSR makes content ACCESSIBLE through a browser and a public REST API, and uses admin roles for authentication.

Publication Versioning allows users to access archives of previous data standards versions.



R - Clinical Data standards are accessible and valid globally

The Catalog contains authorship of Data Standards Models.



The Provenance ontology PROV-O is used to capture information about changes to the models between versions.

GDSR technology follows community trends, it is aligned with Semantic Web (W3) standards and vocabularies, as well as Pharma best practices (CDISC).



I - Ontologies and controlled vocabularies are used to define the standards metadata

GDSR uses RDF/OWL ontologies and RDF for representing metadata.

```
<http://gdsr.roche.com/pd-biometrics/models#Model.AdaM>
rdf:type vmm:Model ;
vmm:isDefinedBy vmm:Clinical_data_interchange_standards_consortium ;
vmm:isModelType vmm:Foundational_model ;
vmm:isUsedFor vmm:data_analysis_clinical_trial ;
skos:altLabel "AdaM" ;
skos:definition "CDISC foundational model describing the underlying rules for AdaM datasets. New versions
skos:note "When published by CDISC and Roche choose to adopt - should use the CDISC version number" ;
skos:prefLabel "Analysis data model" ;
```

All metadata and schema in the GDSR are described in ontologies.

Standards stored in the GDSR link to external datasets e.g. CDISC, enabling the exploration of connected vocabularies.