

Single Day Event

Huw Mason & Selena Baset,
F. Hoffmann-La Roche Ltd

GDSR Workbench

where clinical data standards are born and raised FAIR!

Introduction: The GDSR Platform

The GDSR Vocabularies platform is a comprehensive suite of tools for hosting FAIR data standards and vocabularies. It facilitates data linking and sharing across the different layers of the Roche Data Commons model.

Powered with strong semantic capabilities, it continues to evolve in a way that meets the dynamically changing needs of a growing community of stakeholders.

In GDSR, FAIR data standards and their defining vocabularies are organized in domains that are readily accessible via the same common GDSR portal but governed independently by the domain business process owners.

Recap: GDSR FAIR (2020)

What makes the GDSR FAIR?

- Unique and persistent identifier
- Rich Metadata
- Searchable (meta)data

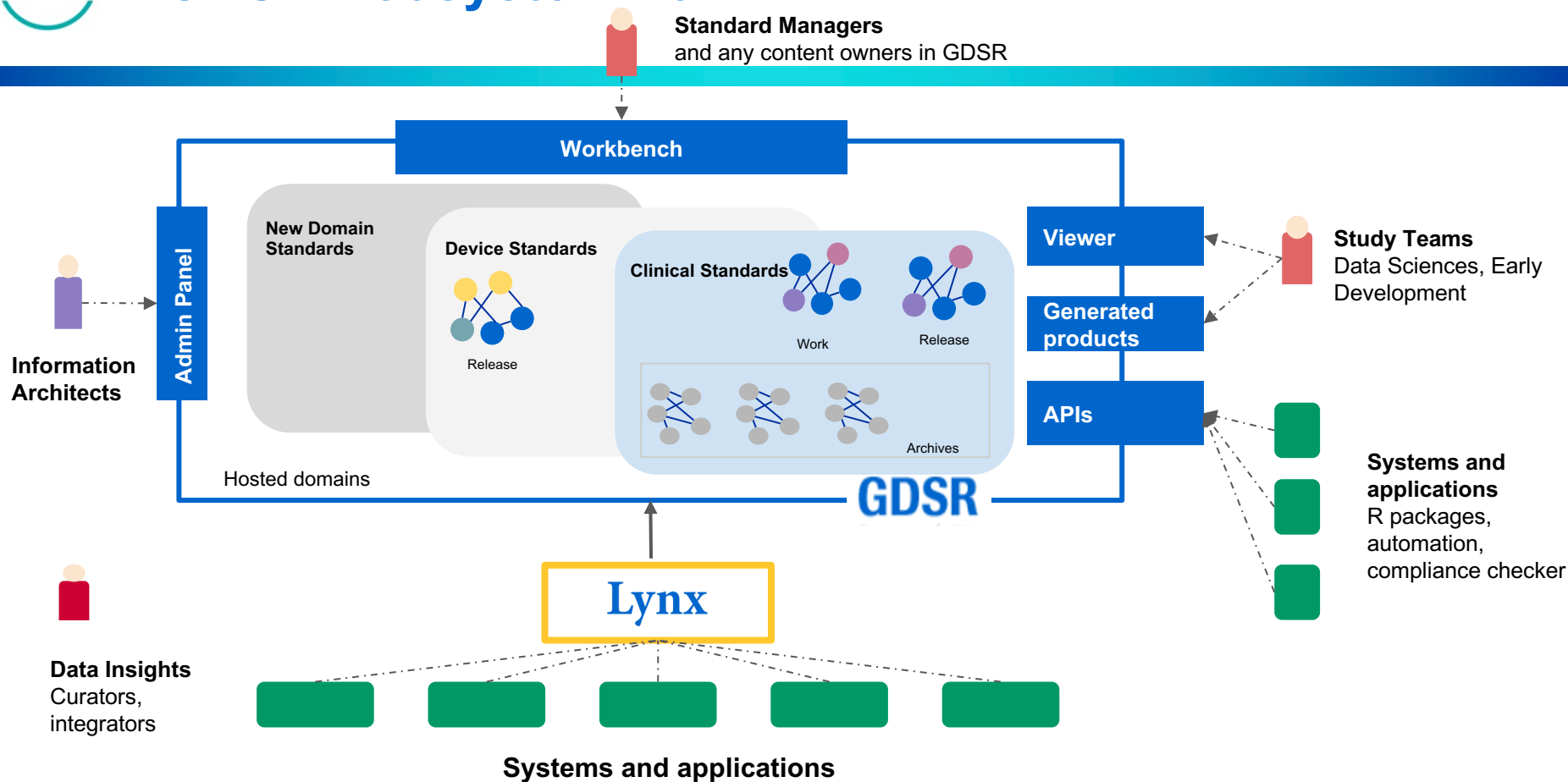


- Retrievable (meta)data via standard protocols
- (Optional) authentication/authorization
- Metadata is always accessible

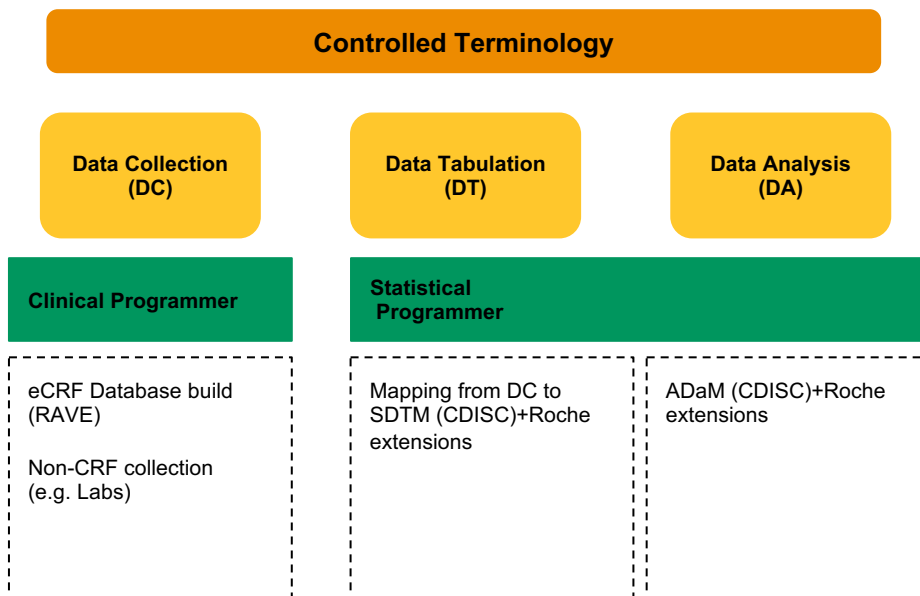
- Rich, plural, accurate meta(data)
- Accessible data license
- Keep the provenance of the information

- meta(data) in a formal and shared language
- external meta(data) are FAIR
- meta(data) points to external (meta)data

GDSR Ecosystem 2021



The Global Clinical Data Standards Domain (within GDSR) is Roche and Genentech's single source of truth for clinical data standards, storing and making them available to users globally in the most business useful format.

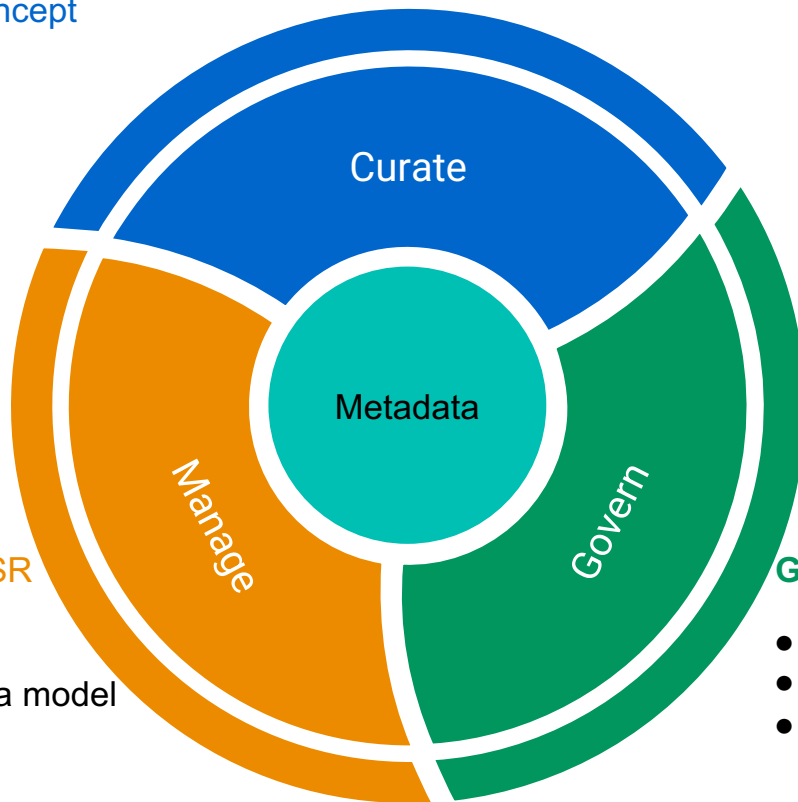


Curate: Develop a metadata concept

- Input from SMEs
- Review guidelines
- Query existing content
- View metadata history

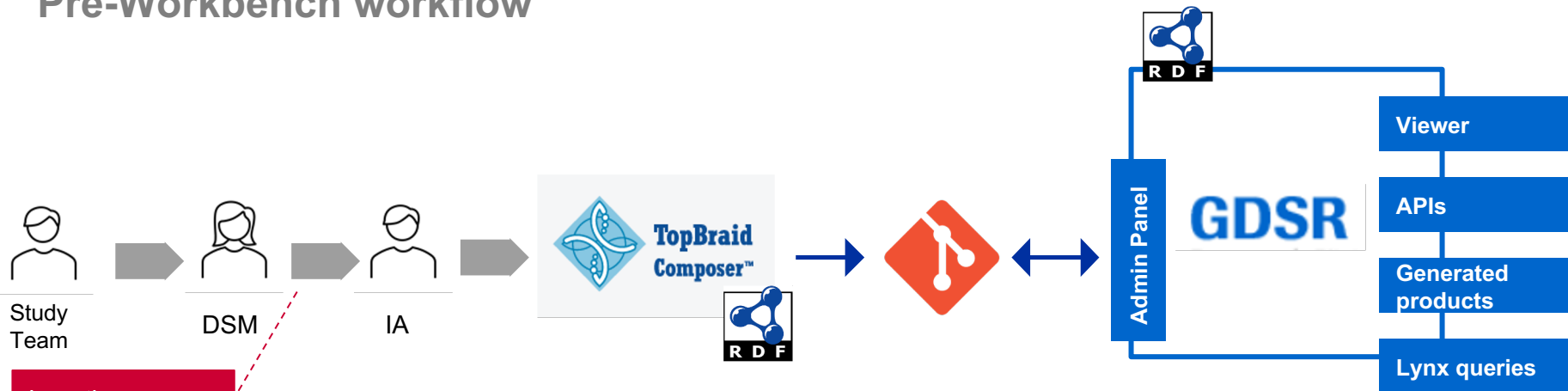
Manage: Own a concept in GDSR

- Add, Edit, Delete metadata
- Current implementation of a model
- Hierarchies & relationships

**Govern:** QC & approve a concept

- Peer review
- Quality Control
- Approval for release to end users

Pre-Workbench workflow



Lengthy process

Requests for clinical standards that are endorsed by DSG are not accessible in GDSR unless an RDF expert (an IA) enters them using a proprietary third-party software, version control them, then upload to GDSR via the AP.

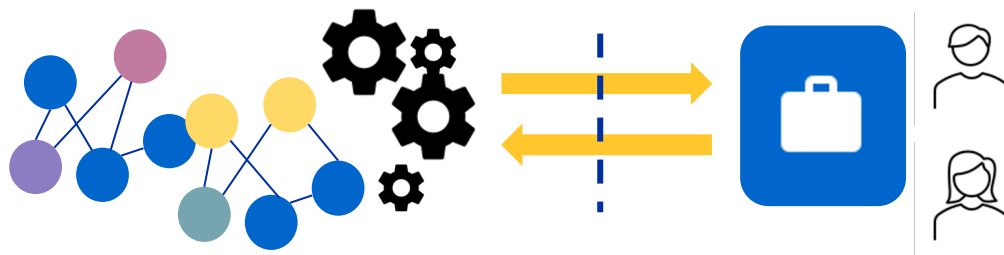
What is the GDSR Workbench?



The GDSR Workbench is a bundle of semantic and user-friendly authoring tools for metadata resources



Provides Data Standards managers with a direct interface to the GDSR metadata



Internally developed and configured



Abstracts the technicalities of the underlying semantic technology and triple-based model

Data managers can now **curate, manage, and govern** their metadata from a single point of access

What's in the GDSR Workbench?

Manage



How can I create, edit, and delete concepts?

Metadata Management



An editing interface for data managers to browse, create, edit or delete resources according to a predefined set of permitted editing scenarios

Curate



How can I find a concept using its identifier?

Identifier Checker



A tool to check the usage of any resource given its unique and permanent Identifier (URI). By looking up the identifier, the user can see whether the resource exists, how it is linked to other resources and track the evolution of the resource across the different publications.

Govern



How can I (peer) review and publish my content?

Governance



The Governance module allows data managers to review and approve all changes introduced to the metadata via the Metadata Management module.

Curate



How can I query GDSR content ?

Ask GDSR



A library of queries to the GDSR content that are expressed in a natural language. The query results can be easily exported lending themselves to different usages such as quality control, report generation or data

What's in the GDSR Workbench?



Identifier Checker

Curate

Findable1. (Meta)data are assigned a globally unique and persistent identifier
Findable3. Metadata clearly and explicitly include the identifier of the data they describe
Findable4. (Meta)data are registered or indexed in a searchable resource
Accessible2. Metadata are accessible, even when the data are no longer available
Interoperable3. (Meta)data include qualified references to other (meta)data

Each metadata concept in the GDSR has its own unique identifier (URI).
The Identifier Checker allows a Data Manager to check if a concept exists or has ever existed, and see the history of that concept across publications.



URI Checker

Check if

GDS Identifier

<http://rdf.cdisc.org/sdtm-terminology#C99079>

All the publications plus IA work and v

All URIs 343 result(s)							EXPORT ALL
Connected data (transposed) 39 result(s)							EXPORT
Publication name/Label	Subject	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://rdf.cdisc.org/mms#inValueDomain	http://rdf.cdisc.org/ct/schema#nciPreferredTerm	http://rdf.cdisc.org/ct/schema#nciCode	http://rdf.cdisc.org/ct/schema#cdiscSubmissionV	
Filter	Filter	Filter	Filter	Filter	Filter	Filter	
archive_2017-04-06T13.31.01.000	http://rdf.cdisc.org/sdtm-terminology#C99079.C16032	http://rdf.cdisc.org/mms#PermissibleValue	http://rdf.cdisc.org/sdtm-terminology#C99079	Long-term Follow-up	C16032	LONG-TERM FOLLOW-UP	
archive_2017-04-06T14.30.31.000	http://rdf.cdisc.org/sdtm-terminology#C99079.C16032	http://rdf.cdisc.org/mms#PermissibleValue	http://rdf.cdisc.org/sdtm-terminology#C99079	Long-term Follow-up	C16032	LONG-TERM FOLLOW-UP	



Ask GDSR

Curate

Findable2. Data are described with rich metadata (defined by R1 below)
Findable3. Metadata clearly and explicitly include the identifier of the data they describe
Findable4. (Meta)data are registered or indexed in a searchable resource
Interoperable3. (Meta)data include qualified references to other (meta)data
Reusable1. (Meta)data are richly described with a plurality of accurate and relevant attributes

Library of (SPARQL) queries by model, expressed in business language, allowing Data Managers to query their content and generate real time report views on the GDSR Workbench data.



Ask GDSR

A curated collection of natural-language queries to be run across GDSR.

Category Filter
GDSR Clinical Data Standards

Subcategory filter
Controlled Terminology: GDSR Clinical Data Standards

Named Query

Return codelists where a keyword appears in a term's definition or submission

Query description: Returns codelists for which the term has a definition or submission value

Query parameters

Term Definition or Submission Value keyword *

Showing up to 1000 results. Some records may not be visible.

Term URI	Codelist URI	SDTM/Sponsor Term Value Code	SDTM Origin Codelist Code	SDTM/Sponsor Term Submission Value	NCI Preferred
Filter	Filter	Filter	Filter	Filter	Filter
http://rdf.cdisc.org/sdtm-terminology#C71620.C96619	http://rdf.cdisc.org/sdtm-terminology#C71620	C96619	C71620	/HPF	Per Hig
http://rdf.cdisc.org/sdtm-terminology#C71620.C44256	http://rdf.cdisc.org/sdtm-terminology#C71620	C44256	C71620	RATIO	Ratio

What's in the GDSR Workbench?



Metadata Management

Manage

Findable1. (Meta)data are assigned a globally unique and persistent identifier

Findable2. Data are described with rich metadata (defined by R1 below)

Findable3. Metadata clearly and explicitly include the identifier of the data they describe

Findable4. (Meta)data are registered or indexed in a searchable resource

Reusable1. (Meta)data are richly described with a plurality of accurate and relevant attributes

Metadata Management allows Data Managers to add, edit, delete metadata, and view those concepts and their relationships in the context of a given data standards model.

Dataset name *

ADSL

Dataset label *

Subject Level Analysis Dataset

Ordinal *

1

Dataset structure *

One record per subject

ADaM class *

ADSL

Context *

adamv

Publisher *

Roche

Not implemented in STREAM (true if not yet implemented, else blank)

stored in *

YES

NO

UNSET

ADSL

ADaMv Column Extension variable (68)

Show columns (11/11)

CLEAR FILTERS

+ CREATE NEW

Label (URI)	Name	Ordinal	Label	Length	Sponsor core	Sponsor ordinal key	SAS Format	Context
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
STUDYID - Study Identifier (http://gdsr.roche.com/pd-biometrics/adam-ADSL#Column)	STUDYID	10	Study Identifier	8	true	1		

Name *

STUDYID

Ordinal *

10

Label *

Study Identifier

Length *

8

Sponsor core

YES

NO

UNSET

Sponsor ordinal key

1

Derivation type *

Predecessor

Sponsor compliance *

Required Variable

10 data element *

STUDYID

What's in the GDSR Workbench?



Governance

Govern

Findable2. Data are described with rich metadata (defined by R1 below)

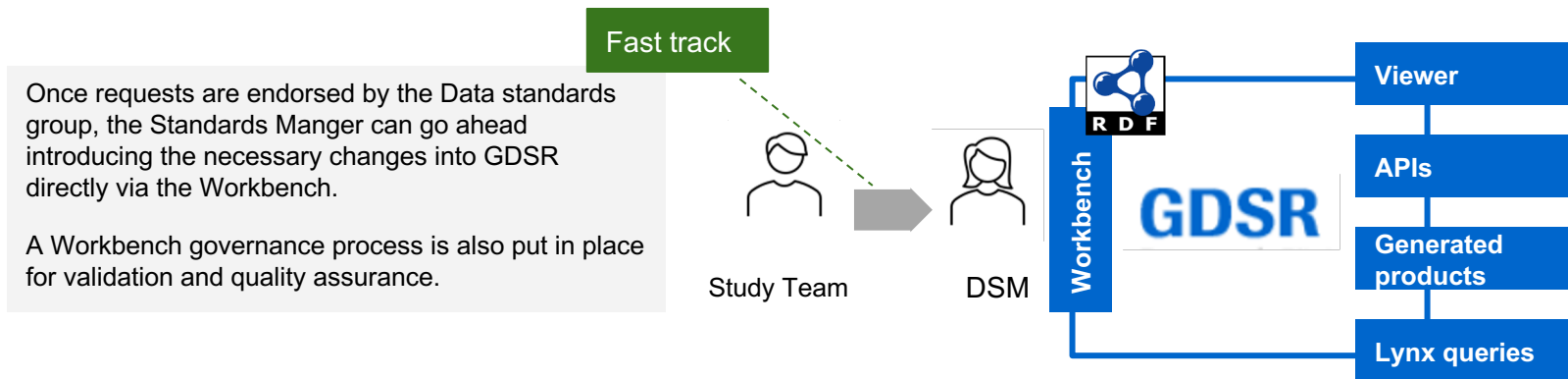
Findable4. (Meta)data are registered or indexed in a searchable resource

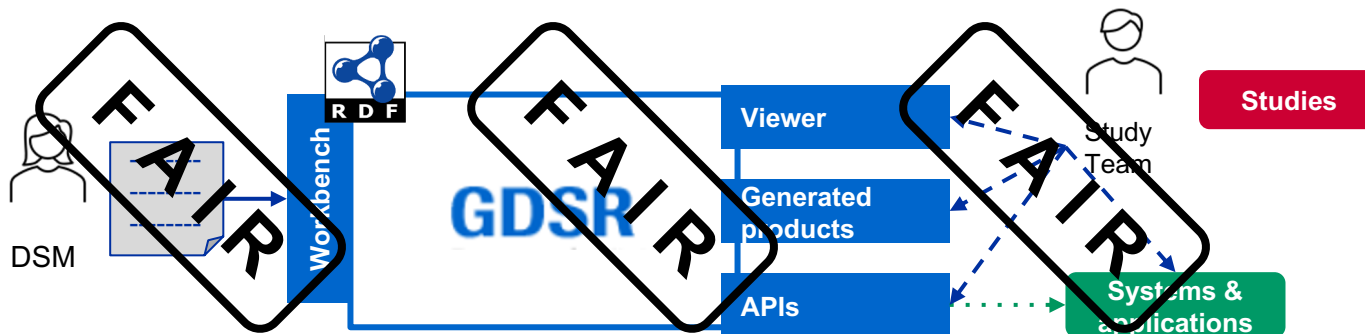
Reusable1. (Meta)data are richly described with a plurality of accurate and relevant attributes

The Governance module allows Data Managers to review and approve for Publication updates and additions made to metadata in Metadata Management. It also provides Data Managers with a complete overview, across models, of all updates for a Publication.

Uri	Type	Label	Status	Comment	Actions
☐ Filter	☐ Filter	☐ Filter	☐ Filter	☐ Filter	
☐ http://uri.Workbench.item	Not editable in this Workbench	SCD - Screening Failure for CDx	Draft	Moved to Draft due to creation of Non-CRF Model Guidance	⋮
☐ http://uri.Workbench.item2	Not editable in this Workbench	STA - Solid Tumor Assessments	Draft	Moved to draft due to modification of Non-CRF Model	⋮
☐ http://uri.Workbench.item3	Roche Codelist	L00003 - Adverse Event Pathogen	Approved	QC'd HM	⋮
☐ http://uri.Workbench.item4	Roche Codelist	L00028 - Method of Tumor Response Assessment	Draft	Moved to Draft due to creation of Roche permissible value	⋮

Workflow with Workbench





FAIR GDSR and Workbench aren't just applicable to the Clinical Data Standards domain, but all domain tenants on the GDSR



Workbench Architects:	Didier Clement, Elvira Mitraka, Huw Mason
Process Owner:	Eddy Foster
WB Domain Owners:	Philipp Thomas Ernst (Clinical Standards), Andreas Thalhammer (RIS)
Product Owner:	Selena Baset
Business Support:	Nelia Lasierra, Javier D. Fernández, Ivan Robinson
IT System Owner:	Christophe Baumgaertner
Solution Architects:	Adam kozak, Michal Pietrusinski
Business Analyst:	Katarzyna Ulmer
PMs:	Ireneusz Lebiodzki, Monika Machera-Zahorowska
UX:	Paulina Frischke
Developers:	Sebastian Choina, Michal Kalinowski, Jacek Caban, Szymon Tyka, Piotr Duleba
Testers:	Celso Pereira, Bartosz Grab, Dariusz Stepień
Validation Leads:	Anna Zukowska, Marta Zalewska
Operations:	Edward Wharton, Witold Borkowski



Questions



Doing now what patients need next



The Global Healthcare Data Science Community

Contact Channels

- 📞 UK +44 1843 609600
- 📞 USA +1 609 514 5105
- ✉️ office@phuse.global
- 🌐 phuse.global

Social Media

- 🐦 [@phusetwitta](https://twitter.com/phusetwitta)
- 📘 [/phusebook](https://facebook.com/phusebook)
- ▶️ [/phusetube](https://youtube.com/phusetube)
- 🌐 [/company/phuse](https://linkedin.com/company/phuse)