Enhancing Clinical Data Quality and Consistency with Value Level Metadata for non-CRF Data Collection

Zahra Karimaddini, Igor Steiner

Meet the Speakers

Zahra Karimaddini

Title: Senior Data Science Product Leader

Organization: F. Hoffmann-La Roche Ltd

Email: zahra.karimaddini@roche.com

Zahra Karimaddini, PhD, has a background in computational biology, data science and personalized medicine. In her role at the Data Standards and Governance group, she is working on development and enhancement of various non-CRF data models, including digital measures, oncology image based assessments as well as electrode-based assessments.

Igor Steiner

Title: Principal Data Science Product Leader

Organization: F. Hoffmann-La Roche Ltd

Email: igor.steiner@roche.com

Igor Steiner has worked in Clinical Data Management for nearly 20 years, specializing in non-CRF data acquisition. In his current position as a Biomedical Data Standards Specialist, he develops and maintains non-CRF collection standards for various types of data.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. Problem Statement
2. Clinical Data Life Cycle
3. Value Level Metadata
4. Outlook

Problem Statement: Translating Complex Data

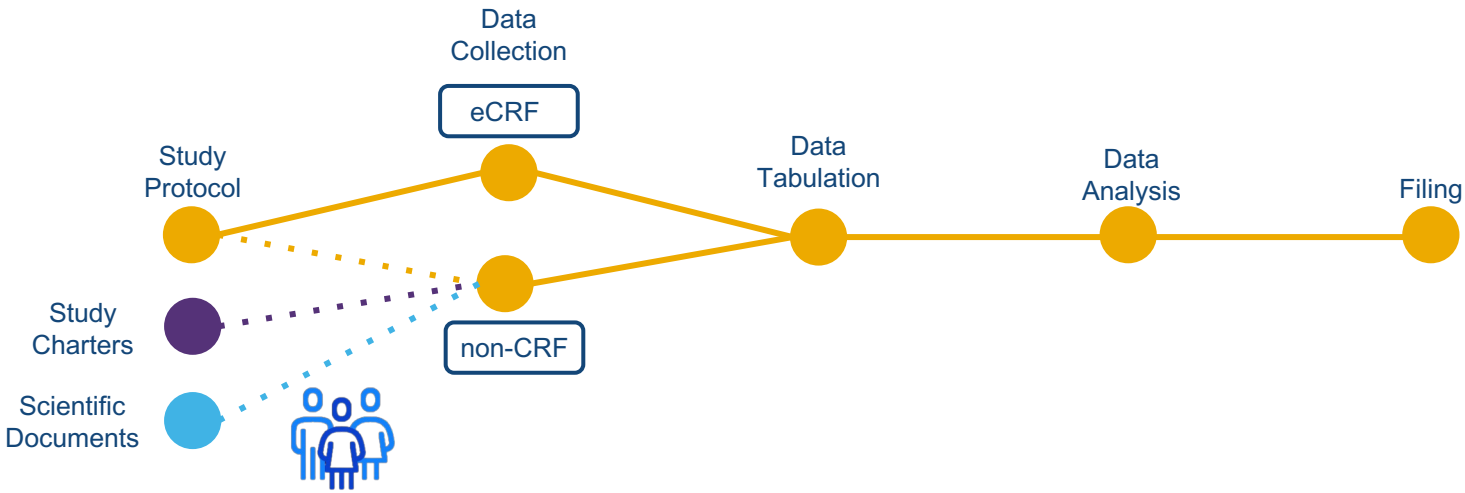
From Source Documents to Structured Gold

A standard approach is needed to reduce risks and challenges in achieving consistent data collection specifications.

Clinical Data Life Cycle

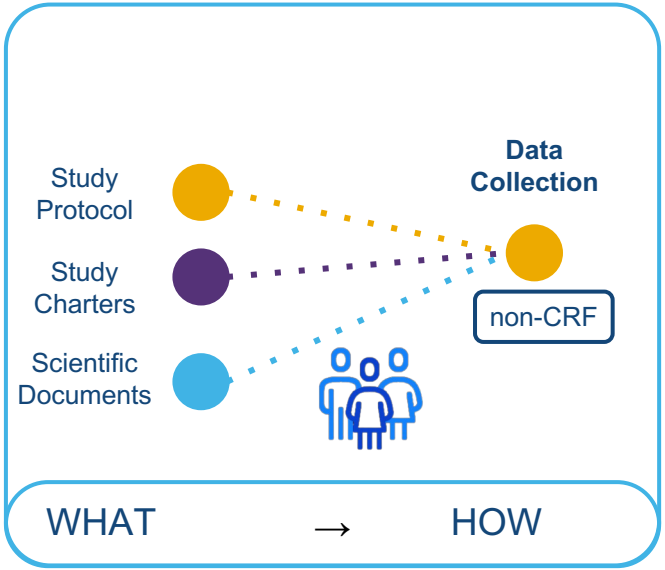
Harmonizing clinical trial data through its life cycle enables automated data processing pipelines

Clinical Data Life Cycle: Data Collection



Non-CRF data collection is often based on more documents than only the study protocol.

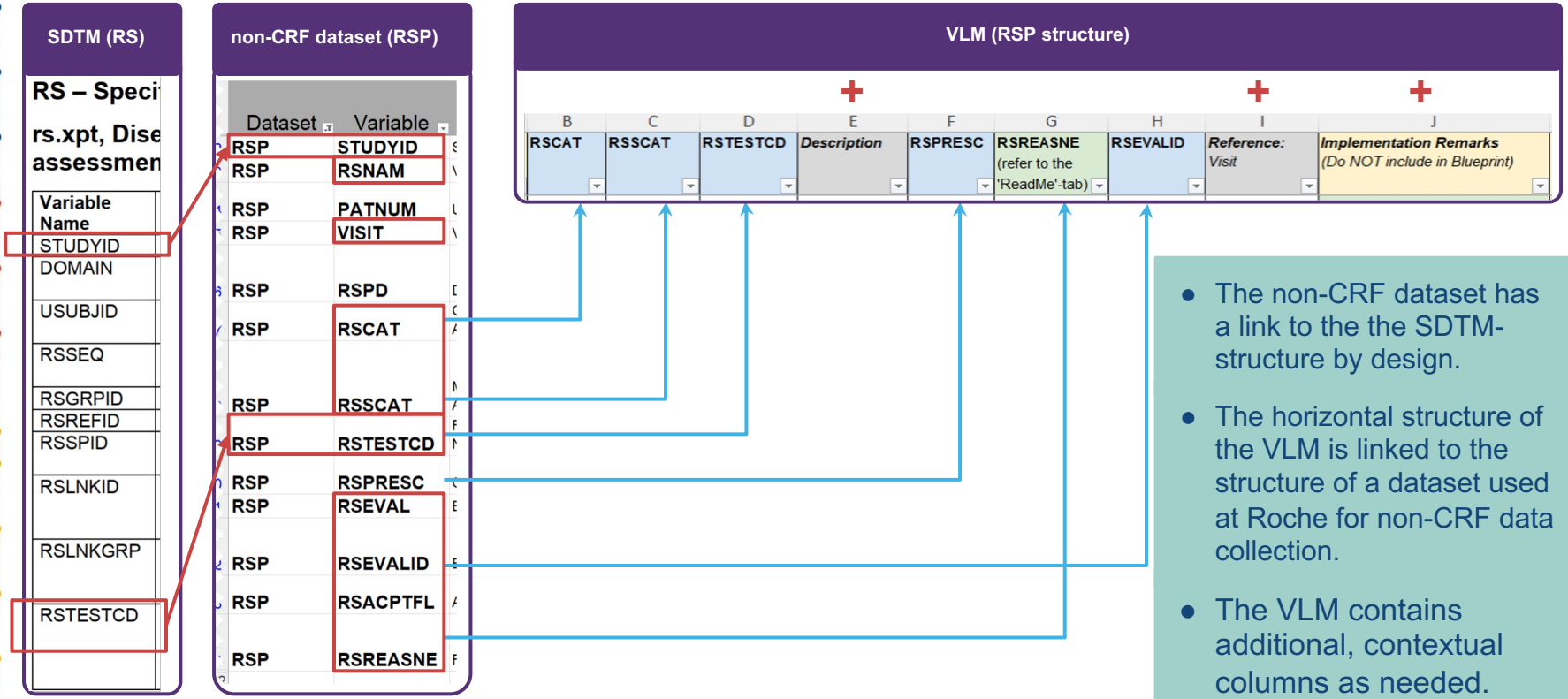
Clinical Data Life Cycle: non-CRF Data Collection



- Data structures & Controlled Terminologies (CDISC Standards + Roche extensions) are used to collect the data points.
- These are provided as valid terminologies for each variable as one-dimensional codelists.
- CDISC Codetable Mappings help to understand the relationship between terminology of a few variables.

The Value Level Metadata are bridging the gap between WHAT should be collected scientifically and HOW this should be collected according to the standards.

Value Level Metadata (@Roche): Structure



- The non-CRF dataset has a link to the the SDTM-structure by design.
- The horizontal structure of the VLM is linked to the structure of a dataset used at Roche for non-CRF data collection.
- The VLM contains additional, contextual columns as needed.

Value Level Metadata (@Roche): Content

Aligned with CDISC's Codetable Mapping Files

VLM

	A	B	C	D	E	F	G	H	I	
1	***sorting (reset) (Do NO***	**RSCAT**	**RSSCAT**	**RSTESTCD**	***Description***	**RSPRESC**	**RSREASNE (refer to the 'ReadMe'-tal**	**RSEVALID**	***Reference: Visit***	***Implementatio (Do NOT includ***
2	1	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		CR	<blank>	RADIOLOGIST RADIOLOGIST 1	>BSL	- Post Baseline - Target lesions
3	2	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		PR	<blank>	RADIOLOGIST RADIOLOGIST 1	>BSL	- Post Baseline -Target lesions
4	3	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		SD	<blank>	RADIOLOGIST RADIOLOGIST 1		
5	4	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		PD	<blank>	RADIOLOGIST RADIOLOGIST 1		
6	5	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		NE	<mandatory, free text>	RADIOLOGIST RADIOLOGIST 1		
7	6	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP		NA	<blank>	RADIOLOGIST RADIOLOGIST 1		
	7	RECIST 1.1	ANATOMICAL MEDICAL	NTRGRESP		CR	<blank>	RADIOLOGIST		

CDISC

	A	B	C	D	E	F	G	H	I
1	**C-code (Concept Code)**	**Category of Oncology Response Assessment (ONCRSCAT) (codelist code = C124298)**		**C-code (Concept Code)**	**Oncology Response Assessment Test Code (ONCRTSCD) (codelist code = C96782)**	**Oncology Response Assessment Test Name (ONCRTS) (codelist code = C96781)**		**C-code (Concept Code)**	**Oncology Response Assessment Result (ONCRSR) (codelist code = C96785)**
22	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C4870	CR
23	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C18058	PR
24	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C18213	SD
25	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C35571	PD
26	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C62222	NE
27	C124415	RECIST 1.1		C94534	TRGRESP	Target Response		C103424	NOT ALL EVALUATED
30									
31									

RS_RECIST1.0 Mapping 2025-03-28 | RS_RECIST1.1 Mapping 2025-03-28 | RS_iRECIST Mapping 2024-03-27 | RS_LUGANO Mapping 2022-

The VLM content
- is developed through cross-functional discussions with the SMEs
- reflects CDISC recommendations if available and possible
- covers more details compared to the CDISC codetable mapping files

Value Level Metadata (@Roche): Content

Added Value

HOW

WHAT

	A	B	C	D	F	G	H	I	K
1	***sorting*** *(reset)* *(Do*	**RSCAT**	**RSSCAT**	**RSTESTCD**	**RSPRESC**	**RSREASNE** (refer to the 'ReadMe'-	**RSEVALID**	***Reference:*** *Visit*	***Scientific Denomination*** *(Do NOT include in Blueprint)*
2	1	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP	CR	<blank>	RADIOLOGIST RADIOLOGIST 1 RADIOLOGIST 2	>BSL	Complete response (CR) of target lesions as assessed by a radiologist according to RECIST 1.1
8	7	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	NTRGRESP	CR	<blank>	RADIOLOGIST RADIOLOGIST 1 RADIOLOGIST 2	>BSL	Complete response (CR) of non-target lesions as assessed by a radiologist according to RECIST 1.1
14	13	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT METABOLIC MEDICAL IMAGE ASSESSMENT ANATOMICAL MEDICAL IMAGE ASSESSMENT AND METABOLIC MEDICAL IMAGE ASSESSMENT	NEWLIND	Y	<blank>	RADIOLOGIST RADIOLOGIST 1 RADIOLOGIST 2	>BSL	New Tumor lesions present at visit as assessed by a radiologist according to RECIST 1.1
15	14	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	OVRLRESP	CR	<blank>	RADIOLOGIST RADIOLOGIST 1 RADIOLOGIST 2	>BSL	Complete overall timepoint response (CR) as assessed by a radiologist according to RECIST 1.1
22	21	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT ANATOMICAL MEDICAL IMAGE ASSESSMENT AND CLINICAL EVALUATION	OVRLRESP	CR	<blank>	ONCOLOGIST ONCOLOGIST 1 ONCOLOGIST 2	>BSL	Complete overall timepoint response (CR) as assessed by an oncologist according to RECIST 1.1

The Value Level Metadata allows define a meaningful combination of standard terminology and to link it with the scientific statement

Coverage and Advantages Today

Advantages

- Common ground
- Facilitates discussions
- Documents agreements

- Less questions
- Higher quality *
- Faster implementation

- Resolves ambiguities
- Sets boundaries
- Clarifies expectations

* design as well as implementation of standards

Covered by VLMs

- **Oncology Image Assessments**
- Digital Measures
- Electrode Based Assessments
  (ECG, EEG, Polysomnography)
- QRS
- LAB
- Musculoskeletal System Findings
- Central Nervous System Imaging
  (AD and MS MRI, PET)
- Biomarker (stable draft)
- Ophthalmology (stable draft)
- and more to come

Image source: Gemini or Roche internal

Challenges Today

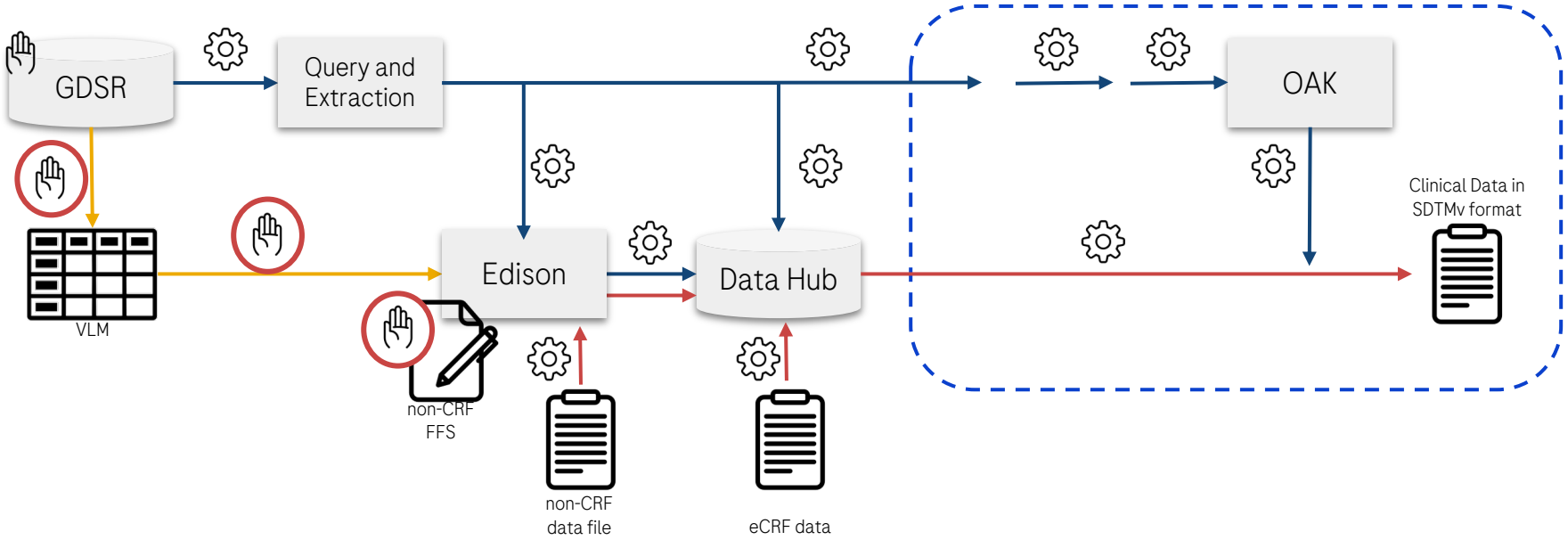
The VLMs bring advantages but at the moment they are created and maintained manually as stand-alone Excel documents.

This leads to :

- manual alignment of the content with the existing standard data structures and terminologies
- manual downstream usage of the VLMs

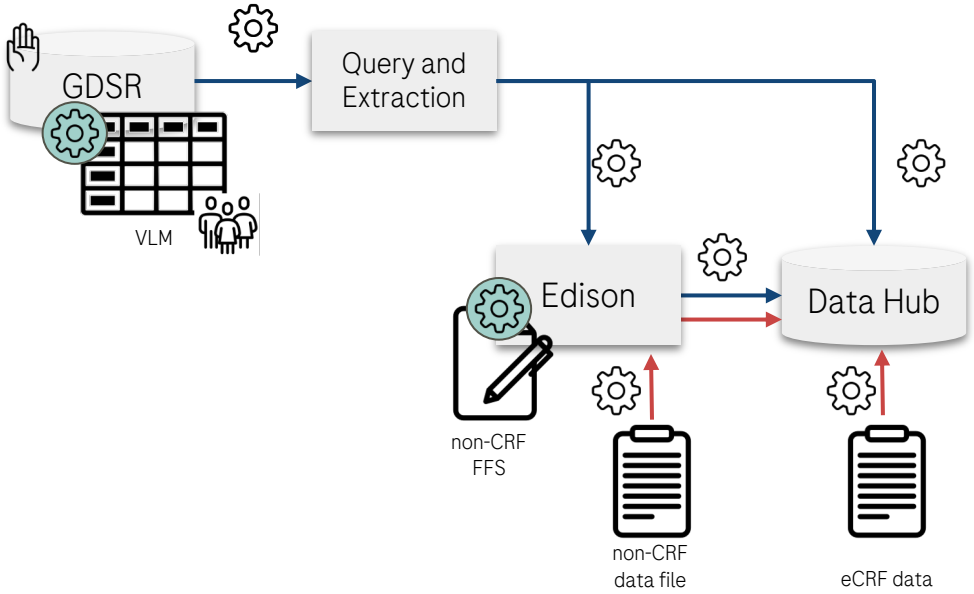
That means extensive maintenance cost (lots of time and resources)!

Outlook Technical Solution: Current Situation



GDSR: Global Data Standards Repository at Roche
FFS: File Format Specification
Edison: Tool for creation of non-CRF FFS and ingestion of non-CRF data platform at Roche.
OAK: R-based solution to automate SDTM mapping developed by Roche and available by CDISC

Outlook Technical Solution: Future



- The VLM will be integrated to GDSR

  GDSR: semantic model driven framework with ontologies, metadata etc

  For the VLM we select variables from datasets to build the structure . . .

  . . . and terminology from codelists as defined in GDSR to specify combinations of terms

- The exposure of the VLM content via the 'query and extraction' tool to create non-CRF FFS
  - allows selection of the standardized combination of terms.
  - prevents room for creation of non-standard and invalid combination of terms.
  - offers implementation of corner cases, e.g., study-specific terminologies, using the implementation guidances.

Summary

The VLMs . . .

- can be of great value to help translate scientifically complex and unstructured information into SDTM-aligned non-CRF specification
- currently are developed and maintained in alignment to, but outside of, our metadata repository
- are planned to be fully integrated into the automated metadata flow for non-CRF data

. . . may have potential values beyond what they were created for.

Potential Expansion of VLM

The VLM resembles the implementation layer of the Biomedical Concepts.

VSTEST	VSTESTCD	VSORRES	VSUNIT	VSLOC
Temperature	TEMP	101.3	F	ORAL

Implementation Layer

And we've seen that it's possible to add a scientific denomination to every row of the VLM.

	A	B	C	D	F	G	H	I	K
1	*sorting (reset) (Do*	RSCAT	RSSCAT	RSTESTCD	RSPRESC	RSREASNE (refer to the 'ReadMe'-	RSEVALID	*Reference: Visit*	*Scientific Denomination (Do NOT include in Blueprint)*
2	1	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	TRGRESP	CR	<blank>	RADIOLOGIST RADIOLOGIST 1 RADIOLOGIST 2	>BSL	Complete response (CR) of target lesions as assessed by a radiologist according to RECIST 1.1
	7	RECIST 1.1	ANATOMICAL MEDICAL IMAGE ASSESSMENT	NTRGRESP	CR	<blank>	RADIOLOGIST	>BSL	Complete response (CR) of non-

Hence,

- there might be a potential usage of the VLMs when designing BCs
- the VLMs could potentially be used directly in the 'Protocol-Driven Automation' process

Thank You!

Acknowledgments:
Philipp Thomas Ernst
Selena Baset
Didier Clément
Javier D. Fernandez

Contacts:
zahra.karimaddini@roche.com
igor.steiner@roche.com