An update on the GQL & SQL/PGQ standards efforts

Keith W. Hare

Convenor, ISO/IEC JTC1 SC32 WG3 Database Languages

June 23, 2023

Sixteenth LDBC TUC Meeting

Who Am I?

- JCC Consulting, Inc.
  - President since August 2019
  - Senior Consultant 1985 – 2019
  - Specialize in
    - High performance database systems
    - Data replication and migration
    - Database Administration

- Standards – SQL and GQL
  - Convenor, ISO/IEC JTC1 SC32 WG3 Database Languages
  - Vice Chair, ANSI INCITS Data Management
- Neo4j Languages, Standards, and Research (LANGSTAR)
  - Standards Diplomacy
  - GQL Strategy
  - Developing the GQL standard

Introduction

- What is a standard?
- SQL and GQL Standards
  - SQL Standards brief history
  - SQL:2023
  - GQL Standard brief history
- SQL, SQL/PGQ, and GQL
  - Graph Pattern Matching
  - GQL Standard Timing
- What happens after GQL V1 is published?
- Summary
- Extra material
  - Related Articles and Web Sites
  - SQL and GQL Artifacts
  - Property Graph Examples – investigative journalism
  - Standards process and participants

What is a standard?

- Standards are agreements between participants on how to do something

  It could be about making a product, managing a process, delivering a service or supplying materials – standards cover a huge range of activities. Standards are the distilled wisdom of people with expertise in their subject matter and who know the needs of the organizations they represent – people such as manufacturers, sellers, buyers, customers, trade associations, users or regulators. https://www.iso.org/standards.html

- **Standards benefit businesses and consumers**

  - **For businesses – easier and cheaper to cooperate and compete**

  - **For consumers – increased choices, quality, and safety, and decreased costs**

Standards Organizations (incomplete list)

- ISO – International Organization for Standardization
- IEC – International Electrotechnical Commission – electronics
- IETF – Internet Engineering Task Force – specifies how the internet works
- IEEE – Institute of Electrical and Electronics Engineers
- ITU – International Telecommunications Union – telephones

Standards Example – Shipping Containers

- Defined by ISO 668:2020 *Series 1 freight containers — Classification, dimensions and ratings*

- Reduce labor to load and unload ships

- Reduce the cost of shipping goods world wide

- Size:
  - 8 feet wide
  - 20 or 40 feet long

- Transported by
  - Ships
  - Trucks
  - Trains

Photo by Ian Taylor

Standards Example – Wireless networks

- IEEE 802.11

  *IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements - Part 11: **Wireless** LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*

- Wireless internet access

- Revised and enhanced over multiple generations
  - 802.11 – 1997
  - 802.11b – 1999
  - 802.11a – 1999
  - 802.11g – 2003
  - 802.11n – 2008
  - 802.11ac – 2014
 - 802.11ax – 2019/2020
  - 802.11be – (2024)

Database Language Standards

- ISO/IEC 9075 *Information technology — Database languages — SQL*
- ISO/IEC 39075 *Information technology — Database languages — GQL*

SQL & GQL Standards

- Developed by ISO/IEC JTC/1 SC/32 WG/3 Database Languages
  - ISO – International Organization for Standardization
  - IEC – International Electrotechnical Commission
  - JTC 1 – Joint Technical Committee 1 – Information Technology standards
  - SC 32 – Sub Committee 32 – Data Management and Interchange
  - WG 3 – Working Group 3 – Database Languages
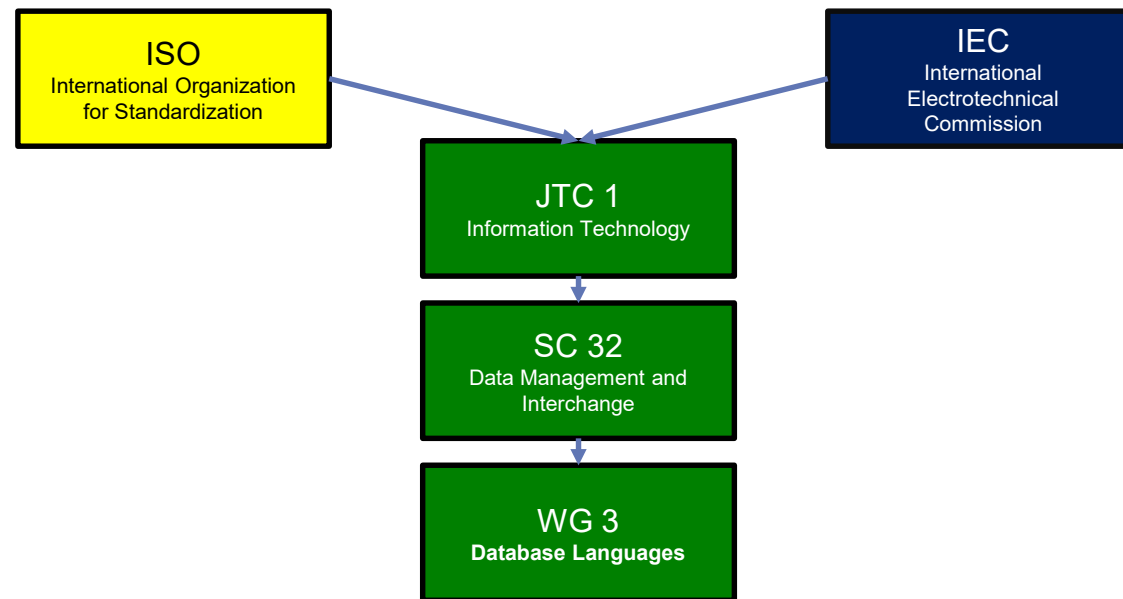
SQL Standards – a brief history

ISO/IEC 9075 Database Language SQL

- SQL-87 – Transactions, Create, Read, Update, Delete

- SQL-89 – Referential Integrity

- SQL-92 – Internationalization, etc.

- SQL:1999 – User Defined Types, triggers

- SQL:2003 – XML & OLAP

- SQL:2008 – XML expansion, "instead of" triggers

- SQL:2011 – Temporal

- SQL:2016 – JSON, RPR, PTF, MDA (2019)

- SQL:2023 – Property Graphs in SQL, JSON enhancements

SQL:2023 includes 11 parts

- Part 1: Framework (SQL/Framework)
- **Part 2: Foundation (SQL/Foundation)**
- Part 3: Call-Level Interface (SQL/CLI)
- Part 4: Persistent stored modules (SQL/PSM)
- Part 9: Management of External Data (SQL/MED)
- Part 10: Object language bindings (SQL/OLB)
- Part 11: Information and definition schemas (SQL/Schemata)
- Part 13: SQL Routines and types using the Java[tm] programming language (SQL/JRT)
- Part 14: XML-Related Specifications (SQL/XML)
- Part 15: Multidimensional arrays (SQL/MDA)
- **Part 16: Property Graph Queries (SQL/PGQ)**

SQL:2023

- Includes
  - SQL/PGQ – Property Graph Queries in SQL
  - Expanded support for JSON
  - Some additions to SQL/Foundation
  - Bug fixes and cleanup
- **Published June 1, 2023!**

GQL Standards – a brief history & timing

ISO/IEC 39075 *Information technology — Database languages — GQL*

- Property Graph Database Language

- Project approved in September 2019

- Draft International Standard (DIS) ballot

  - Initiated May 23, 2023

  - Closes August 15, 2023

- Technically complete now

- Should be published in late 2023 or early 2024

- Graph Pattern Matching

  - Common with SQL/PGQ Graph Pattern Matching

SQL, SQL/PGQ, and GQL

SQL

SQL/PGQ

GPM*

GQL
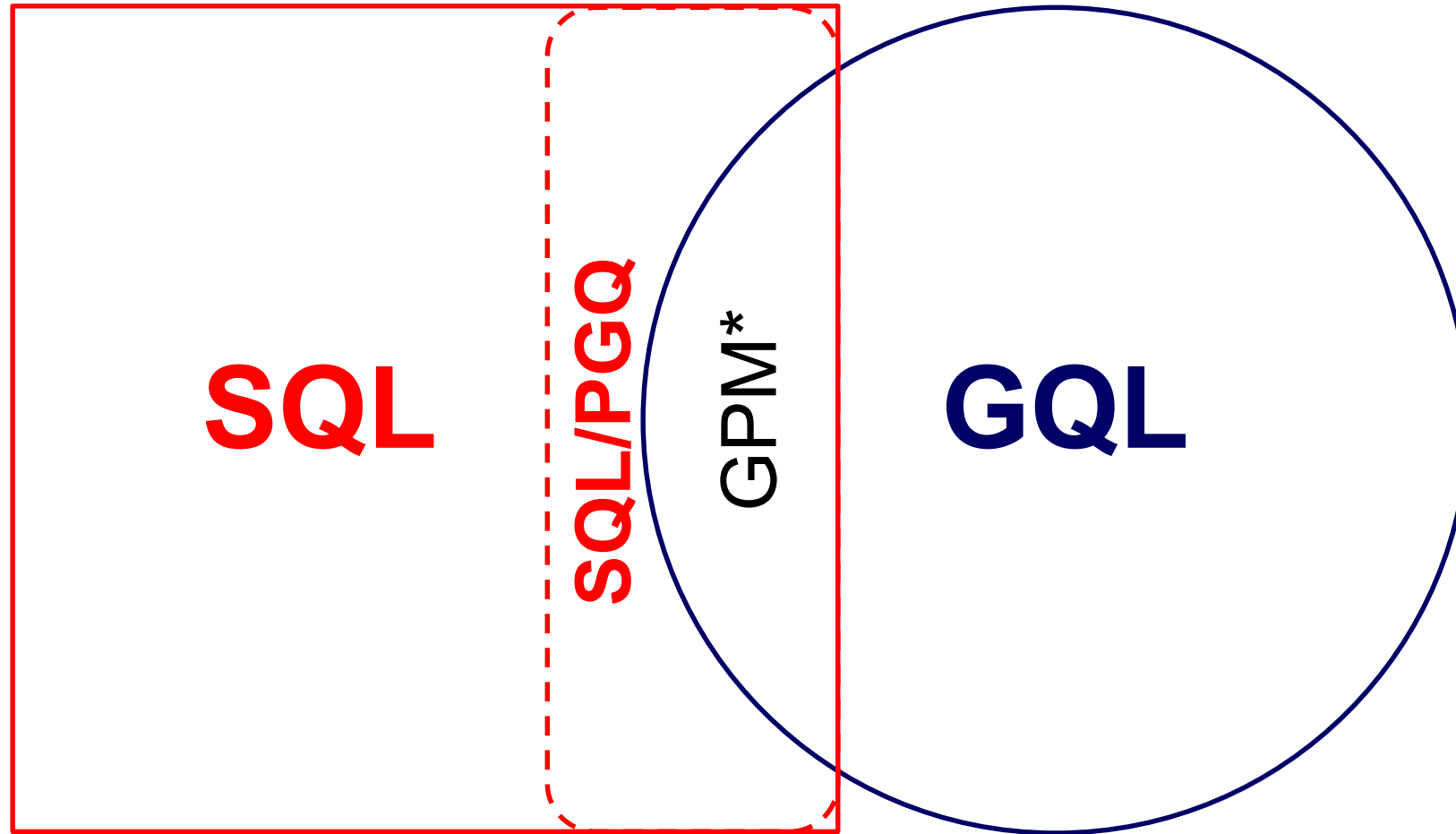
*Graph Pattern Matching

SQL/PGQ – property graph queries in SQL

- ISO/IEC 9075-16 *Information technology — Database languages SQL — Part 16: Property Graph Queries (SQL/PGQ)*

- Additional capability for the SQL standard

- Two major capabilities
  - Define a property graph view on top of existing SQL tables
  - Query a property graph view in an SQL FROM clause

Property Graphs – SQL/PGQ and GQL

SQL/PGQ

- Property Graph views of SQL tables
- **Graph Pattern Matching queries**
  - GRAPH_TABLE() in SQL FROM
  - Supports Reads
- Common foundation with SQL and graph query languages
- Does not support schema-flexible graphs

GQL

- Full DB language
  - DML – Create, Read, Update, Delete
  - DDL – Create Graph, Create Graph Type, Create Graph From Graph Type
- **Graph Pattern Matching queries**
- Leverages common foundation from SQL and property graph languages
- Supports schema-fixed and schema-flexible variants

Graph Pattern Matching – SQL/PGQ versus GQL

SQL/PGQ Example:

```
SELECT * FROM GRAPH_TABLE (students_graph
  MATCH
  (a IS person) -[e IS friends]->
              (b IS person WHERE b.name='Alice')
  WHERE a.name='Mary'
  COLUMNS (a.name AS person_a, b.name AS person_b));
```

GQL Example:

```
  USE students_graph
  MATCH
  (a IS person) -[e IS friends]->
              (b IS person WHERE b.name='Alice')
  WHERE a.name='Mary'
  RETURN a.name AS person_a, b.name AS person_b
```

Graph Pattern Matching

- Powerful capabilities for expressing patterns

- Common between SQL/PGQ and GQL

- GPM has been technically stable since August 2022

- Beginning to appear in products

  - Oracle 23c

  - DuckDB

  - Neo4j 5.9

Expected Dates

- SQL/PGQ
  - DIS Ballot completed – December 2022
  - **Published standard – June 1, 2023**
- GQL V1
  - DIS Ballot Started – May 23, 2023
  - **DIS Ballot completes – August 15, 2023**
  - Resolution of DIS ballot comments – WG3 meeting September 25-29, 2023
  - Potentially an 8-week Final Draft International Standard (FDIS) ballot
  - Published standard – Early 2024
- Note that draft standards are stable by DIS ballot start

What happens after GQL V1 is published?

During the June 2023 WG3 meeting (Washington DC USA), we discussed a number of ideas for possible GQL enhancements

- *Ideas for GQL Expansions* (WG3:DCA-031) (LEX-036)

- *LDBC Extended Schema (LEX) Overview* (WG3:DCA-036) (LEX-035)

- *LDBC Extended Schema Working Group - Use Case Collection Read-out* (WG3:DCA-030r1) (LEX-031)

- *PG-Schema* (WG3:DCA-037) (LEX-034)

- *GQL Types, Names, Labels, and Aliases* (WG3:DCA-038r2) (LEX-027r3)

- *JSON Schema and GQL Schema* (WG3:DCA-039r1) (LEX-030)

- *Schema sub-graphs and incremental transactional updates of graph databases* (DCA-045r1) (LEX-033)

Discussions included both WG3 and LDBC participants

GQL Futures Discussions – In Person

GQL Futures Discussions – Web Conference

What happens after GQL V1 is published?

- Lots of ideas

- GQL users are likely to produce more requirements

- Timing of expansions will depend on:

  - Vendor ability implement GQL capabilities in products

  - User ability to absorb GQL capabilities delivered by products

  - Standards participants ability to write new papers

Summary – Database Language Standards

- SQL Standard – nine editions since 1987
  - Incorporates new features over time
  - New edition published June 1, 2023
  - SQL:2023 includes SQL/PGQ – Property Graph Queries in SQL
    - SQL/PGQ Graph Pattern Matching (GPM) identical to GQL GPM
- GQL Standard – new in 2023
  - Full property graph database language
  - Work on first edition of GQL is technically complete
  - GQL Graph Pattern Matching (GPM) identical to SQL/PGQ GPM
  - Should be published in early 2024
  - Already thinking about GQL V2 and later

Questions?

Related Articles

- *PG-Keys: Keys for Property Graphs*, Renzo Angles et al, ACM SIGMOD Proceedings of the 2021 International Conference on Management of Data, https://dl.acm.org/doi/10.1145/3448016.3457561

- *Graph Pattern Matching in GQL and SQL/PGQ*, Alin Deutsch et al, ACM SIGMOD, Proceedings of the 2022 International Conference on Management of Data, June 2022, https://dl.acm.org/doi/abs/10.1145/3514221.3526057

- *GPC: A Pattern Calculus for Property Graphs,* Nadime Francis et al, Accepted for ACM SIGMOD PODS 2023, available as https://arxiv.org/abs/2210.16580

- *Representing Paths in Graph Database Pattern Matching,* Wim Martens et al, https://arxiv.org/abs/2207.13541

- *PG-Schemas: Schemas for Property Graphs,* Renzo Angles et al, accepted for ACM SIGMOD PODS 2023 Industrial Track, https://arxiv.org/abs/2211.10962

- *Expert Perspectives on Student Errors in SQL*, Daphne Miedema, George Fletcher, Efthimia Aivaloglou, ACM Transactions on Computing Education Vol. 23, No. 1, December 29, 2022, https://doi.org/10.1145/3551392

Related Web Sites and Downloads

- ISO Standards, ISO/IEC 39075 Information technology — Database languages — GQL

- GQL Standards web page, GQLStandards.org

- SQL and GQL Artifacts

  - Generated from the source standards documents as a part of building the PDFs

    - BNF in TXT and XML

    - Exception Conditions

    - Optional Features

    - Implementation Defined Elements

    - Implementation Dependent Elements

  - Useful for SQL and GQL implementers

SQL and GQL Artifacts – Download Links

- SQL
  - SQL/Framework https://standards.iso.org/iso-iec/9075/-1/ed-6/en/
  - SQL/Foundation https://standards.iso.org/iso-iec/9075/-2/ed-6/en/
  - SQl/CLI https://standards.iso.org/iso-iec/9075/-3/ed-6/en/
  - SQL/PSM https://standards.iso.org/iso-iec/9075/-4/ed-7/en/
  - SQL/MED https://standards.iso.org/iso-iec/9075/-9/ed-5/en/
  - SQL/OLB https://standards.iso.org/iso-iec/9075/-10/ed-5/en/
  - SQL/Schemata https://standards.iso.org/iso-iec/9075/-11/ed-5/en/
  - SQL/JRT https://standards.iso.org/iso-iec/9075/-13/ed-5/en/
  - SQL/XML https://standards.iso.org/iso-iec/9075/-14/ed-6/en/
  - SQL/MDA https://standards.iso.org/iso-iec/9075/-15/ed-2/en/
  - SQL/PGQ https://standards.iso.org/iso-iec/9075/-16/ed-1/en/
- GQL https://standards.iso.org/iso-iec/39075/ed-1/en/

ICIJ Property Graph Data Set

- International Consortium of Investigative Journalists (ICIJ) has used property graphs for a number of investigations, including:
  - The Panama Papers: Exposing the Rogue Offshore Finance Industry – 2016
  - Pandora Papers – 2021
- ICIJ Data Set is available to be queried and downloaded:
  - Offshore Leaks Database
- Panama Papers Documentary

Additional Material

The following slides on "Standards Process and Terminology" are included for your information but will not be presented unless there are questions.

Standards Process and Terminology

SQL & GQL standards developed by

- ISO/IEC JTC/1 SC/32 WG/3 Database Languages
  - ISO – International Organization for Standardization
  - IEC – International Electrotechnical Commission
  - JTC 1 – Joint Technical Committee 1 – Information Technology standards
  - SC 32 – Sub Committee 32 – Data Management and Interchange
  - WG 3 – Working Group 3 – Database Languages
- WG3 Current Projects
  - 9075 Database Language SQL
  - 19075 SQL Guidance Standards
  - 29075 Function Libraries (very preliminary)
  - 39075 Database Language GQL
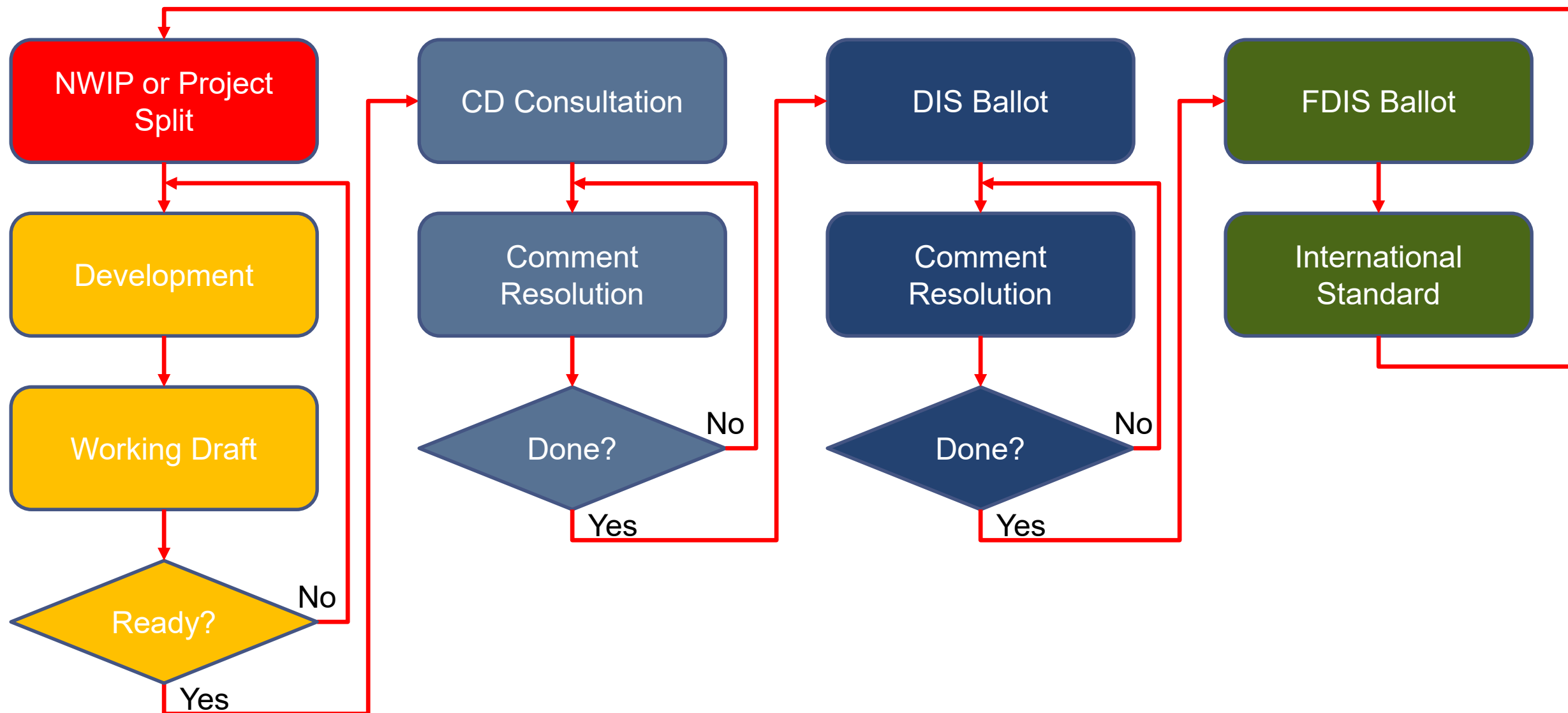
International Standards Hierarchy
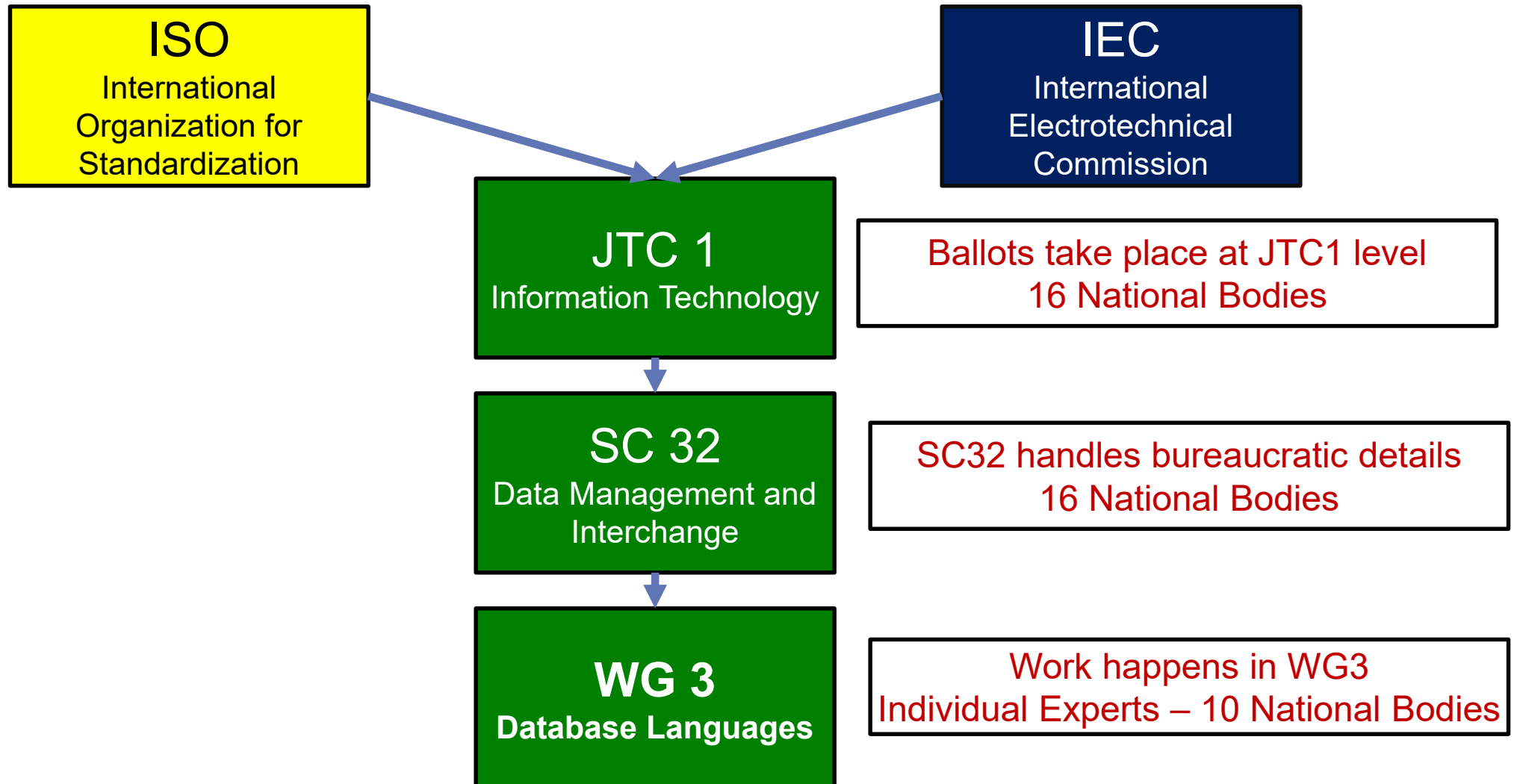
Standardization Steps and Acronyms

- New Work Item Proposal – NWIP

- Working Draft – WD

- Committee Draft – CD

- Draft International Standard – DIS

- Final Draft International Standard – FDIS

- International Standard – IS

ISO/IEC JTC1 Standardization Process

Who participates – SC32 WG3?

Experts from the following national bodies participate in SC32 WG3:

1. China
2. Denmark
3. Finland
4. Germany
5. Japan
6. Korea
7. Netherlands
8. Sweden
9. United Kingdom
10. United States

Organizations Participating in National Bodies

- China
  - Ant Financial
  - Boray Data
  - CESI
  - Huawei
- Denmark
  - TF Informatik
- Finland
  - Profium
- Germany
  - EDB
  - Oracle
- Japan
  - Hitachi
  - Tokyo Metropolitan University

Note: This list is probably incomplete

- Korea
  - Bundang Hospital
  - CnTechSystems
- Netherlands
  - Cannan Consultancy
  - EDB
- Sweden
  - Neo4j
- United Kingdom
  - PR Brown
  - University of Edinburgh
- USA (see later slide)

International Hierarchy mirrored in the US
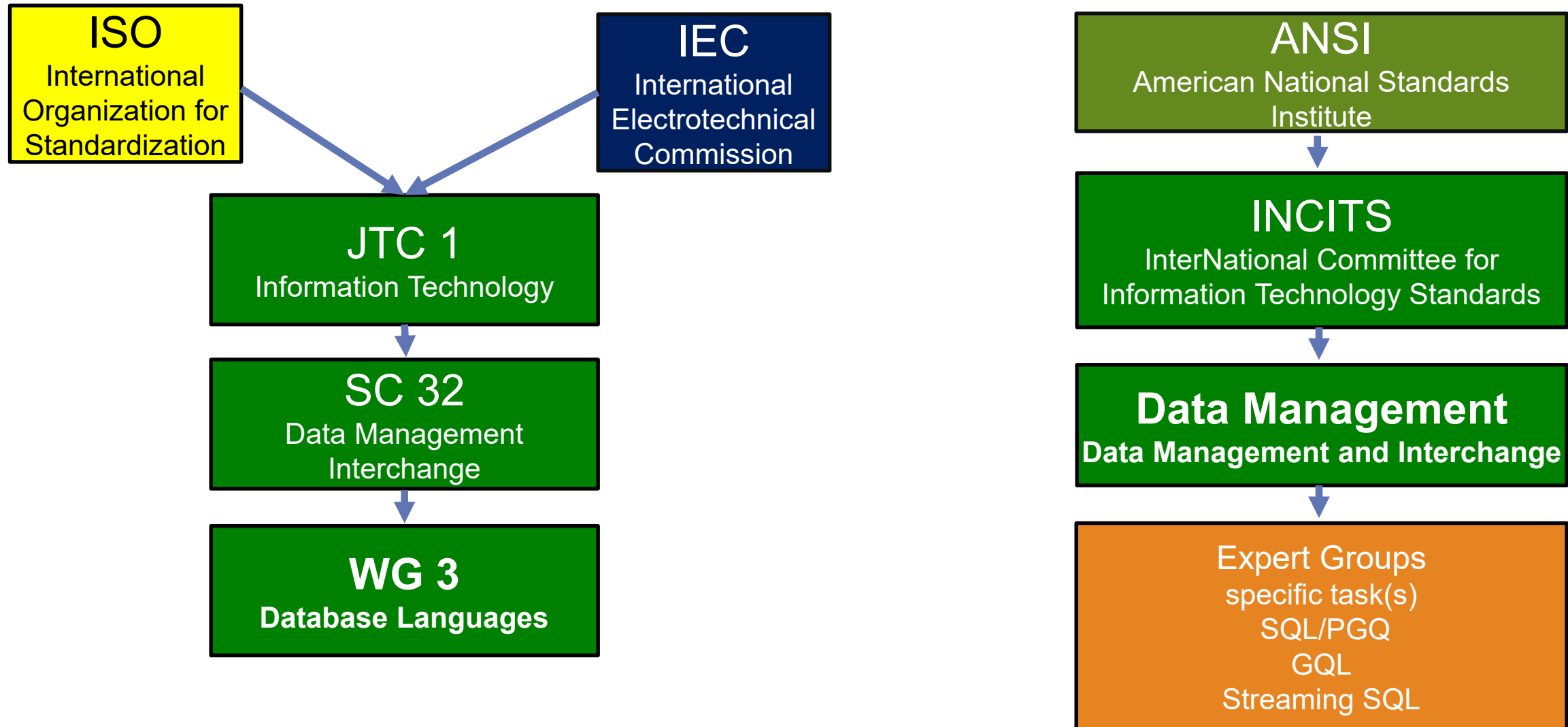
USA Participants – INCITS Data Management

Mostly SQL

Actian Corporation
IBM Corporation
Intersystems Corporation
Microsoft Corporation
Oracle
SAP

Mostly GQL

ArangoDB Inc
DataStax Inc
FairCom USA
Google
Intel
JCC Consulting Inc
Katana Graph
Neo4j Inc
Redis Labs
RelationalAI
TigerGraph

Mostly Streaming SQL

Alibaba Group
Amazon Web Services
Boray Data
Confluent
dbtLabs
Hazelcast
Materialize
Snowflake
Timeplus Inc.

Mostly Metadata

Farance Inc
William McCarthy
National Cancer Institute
Nurocor

Mostly Data Usage
Department of
    Commerce – NIST

Working In INCITS Data Management

- Work done by interested parties in the Expert Groups

  - Property Graph Queries in SQL

  - GQL

- Expert Groups have weekly 2 hour calls

- Discussions in the Expert Groups is based on written proposals

- Proposals are either

  - Concrete change proposals

  - Discussion papers

    - Basis for discussion of designs or alternatives

    - Discussion will eventually lead to a Change Proposal

- Ballot responses approved in INCITS Data Management

Working In ISO/IEC JTC1 SC32 WG3

- Meetings
  - Week-long meetings two to three times a year – now with remote access
  - Monthly web conferences – two or three 3-hour sessions
- International group of national standards bodies
- Participants operate as individual experts
- Concrete change proposals
- Final decisions are made in WG3
- Editors apply approved change proposals to draft standards
- In practice much work happens within US Expert Groups

WG3 Meetings

Year	In-Person	Web Conference
2019	3	0
2020	1	7
2021	0	12
2022	3	6
2023	3	1 (so far)

- 2023 Meetings
  - February 6-20 – Zeist Netherlands
  - March 13 & 15 – Web Conference
  - June 12-16 – Washington DC
  - September 25-29 – Casablanca Morocco
- Web conference challenge – participants from many time zones
- In-person meetings – go to interesting places and sit in conference rooms

Zeist, Netherlands – February 2023
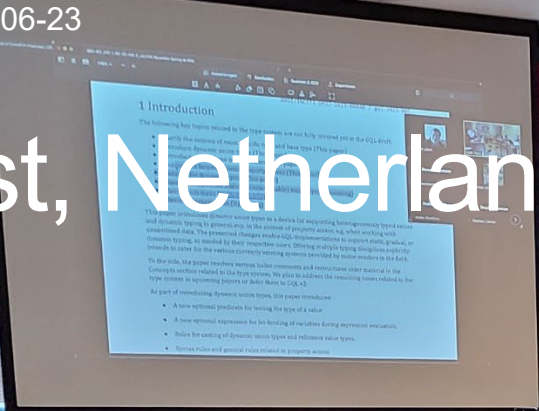
Washington DC, USA – June 2023

SC32 WG3 Formal Liaison Relationships

- LDBC (Linked Data Benchmark Council) — liaison since 2017
  - Focused on property graph work (PGQ & GQL)
    - Benchmarking
    - Existing Languages, Property Graph Schema, GQL Formal Semantics working groups
  - Support/strengthen WG3 standards
    - Review of WG3 documents
    - Contributions to WG3 (critique/corrections, feature suggestions)
    - Requirements for future versions (i.e. GQL Schemas)
  - An evolving bi-directional process for collaboration
  - https://ldbcouncil.org/
- Other liaisons
  - ISO/IEC JTC 1/SC 42 – AI and Big Data
  - OGC (Open Geospatial Consortium)
    - Requirements for supporting spatial data in GQL?

Standards Process Summary

- Iterative, collaborative process
- Compromises between vendors, philosophies, and technologies
- Some amount of standards bureaucracy
- Tedious at times
- Results are pretty good