



# W3C PROV: A Brief Introduction

Satya S. Sahoo

Division of Medical Informatics

Case Western Reserve University

November 13, 2014

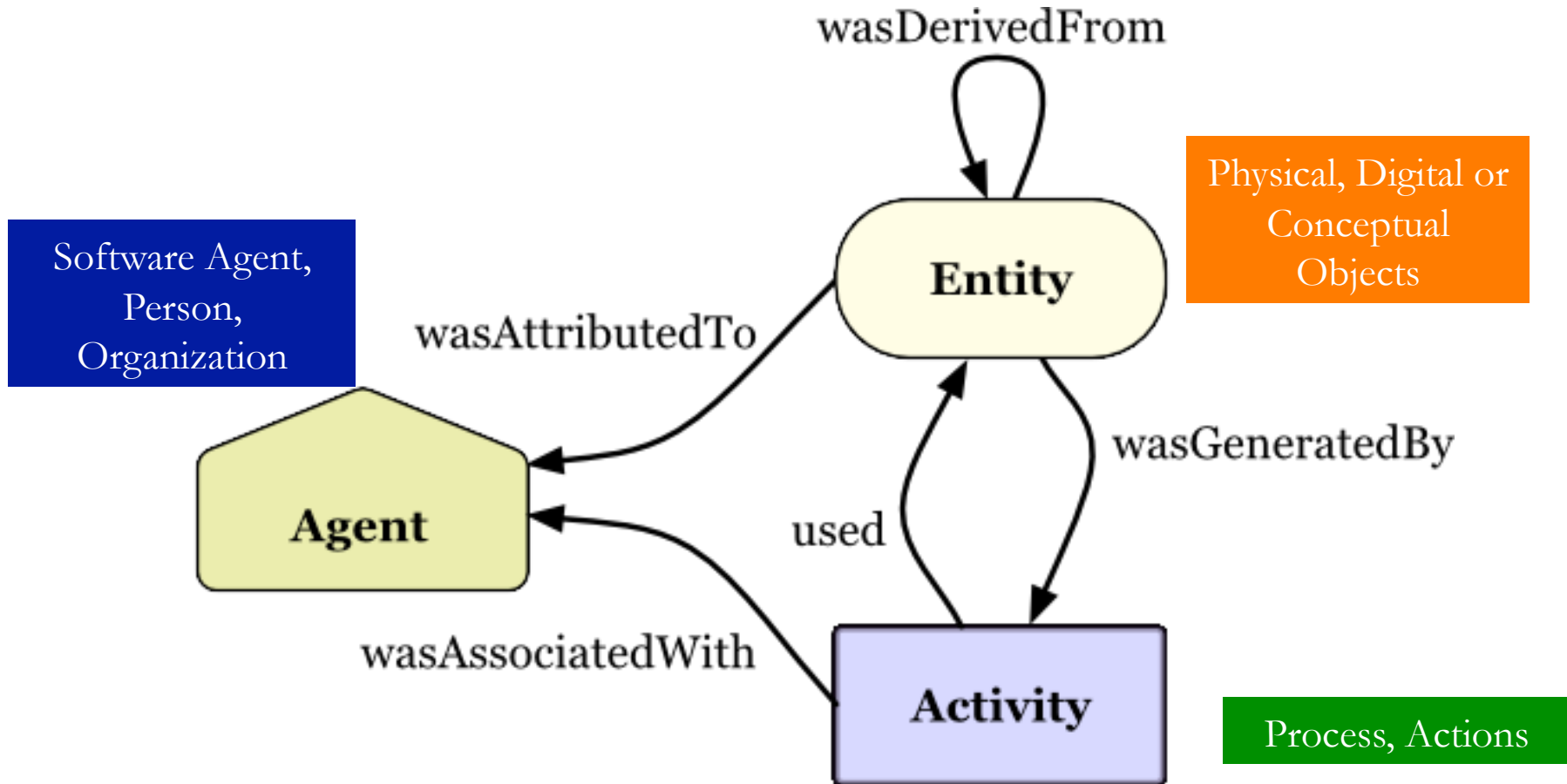
# Background: W3C and the Provenance Working Group

- World Wide Web Consortium (W3C) develops standards for the Web, such as
  - Web Design and Applications: HTML, CSS, Ajax
  - Semantic Web: RDF, OWL2, SPARQL, RIF
  - Web Services: WSDL, SOAP
  - XML Technologies: XML, XML Schema, XSLT
- Provenance Working Group (PWG) was chartered in 2012 to:
  - “...define a language for [exchanging provenance information](#) among applications.”
  - “...[publication and use of provenance information](#) of Web documents, data, and resources”

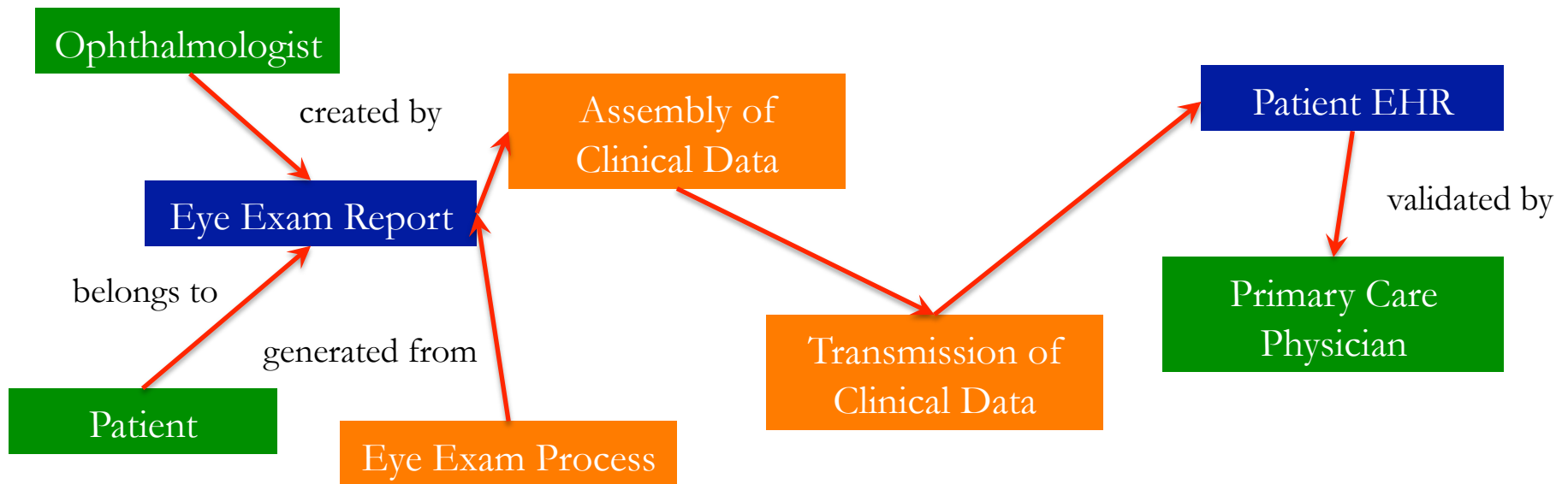
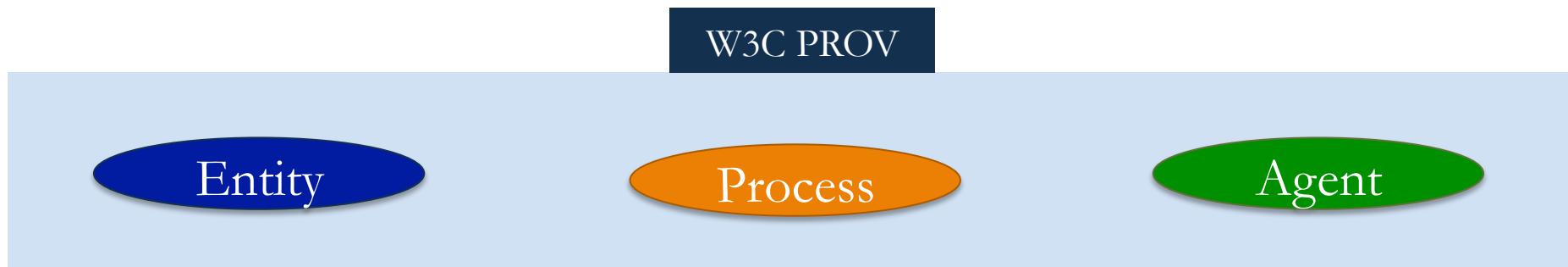
# PROV Family of Specifications

- PROV is designed to model provenance information describing (from PROV Overview):
  - “Understanding how data was collected”
  - “Verifying that the process and steps used to obtain a result”
  - “Determining ownership and rights over an object”
  - “Making judgments about information to determine whether to trust it”
  - “Reproducing how something was generated”
- PROV has four primary components (W3C Recommendations):
  - [PROV Data Model \(PROV-DM\)](#)
  - [PROV Ontology \(PROV-O\)](#)
  - [PROV Constraints](#)
  - [PROV Notation \(PROV-N\)](#)

# PROV Concepts (PROV Primer)



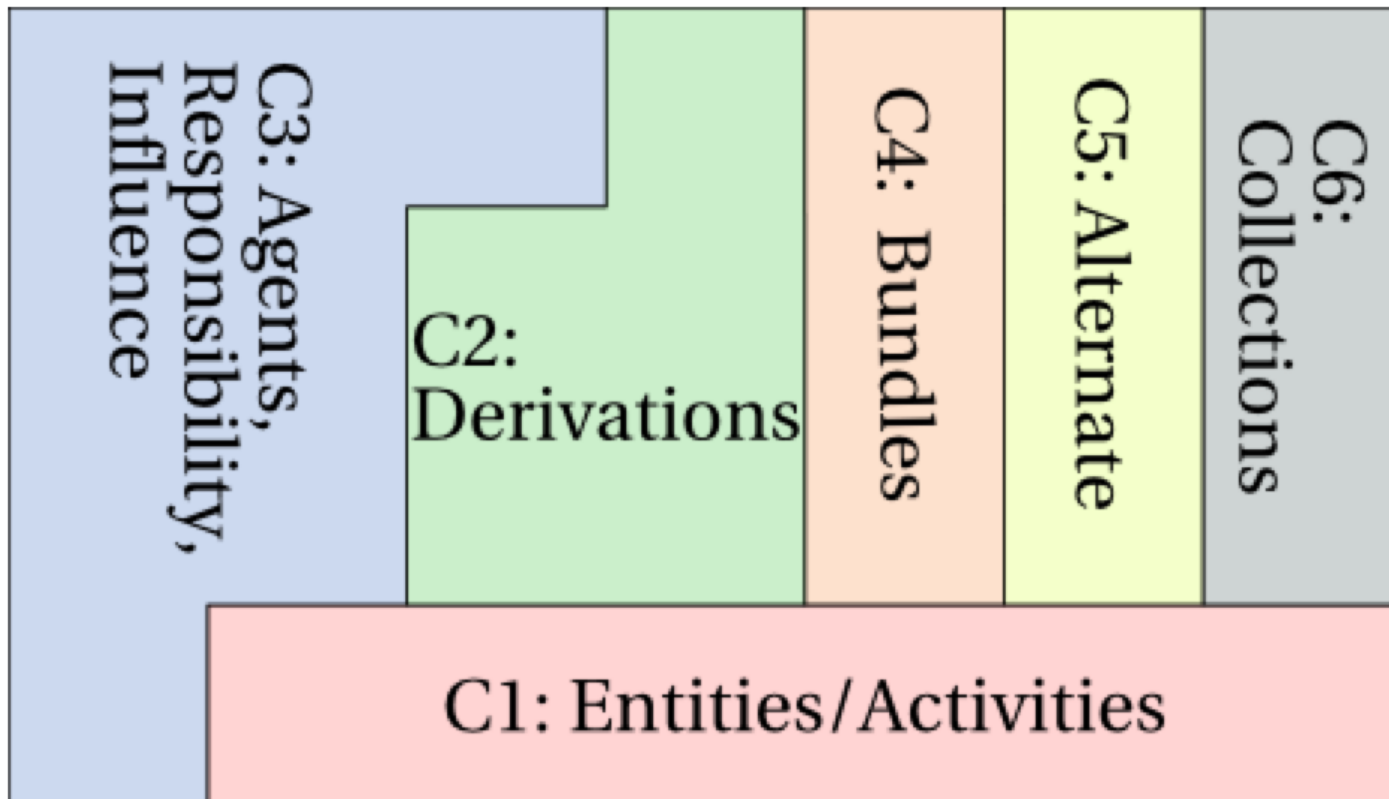
# Using PROV Concepts for Data Provenance Use Case



# Mechanism to Extend PROV Concepts for Application

- Specialization
  - Example: a software agent is special kind of agent
- Extended Relations
  - Binary relations
  - N-ary relations between objects: Time duration of transmission for clinical data
- Meta Provenance: Provenance of Provenance
  - Bundle: Set of provenance descriptions (E.g. provenance bundle of clinical document)
  - Collections: Provides structure to provenance descriptions

# PROV Concepts (from PROV-DM)

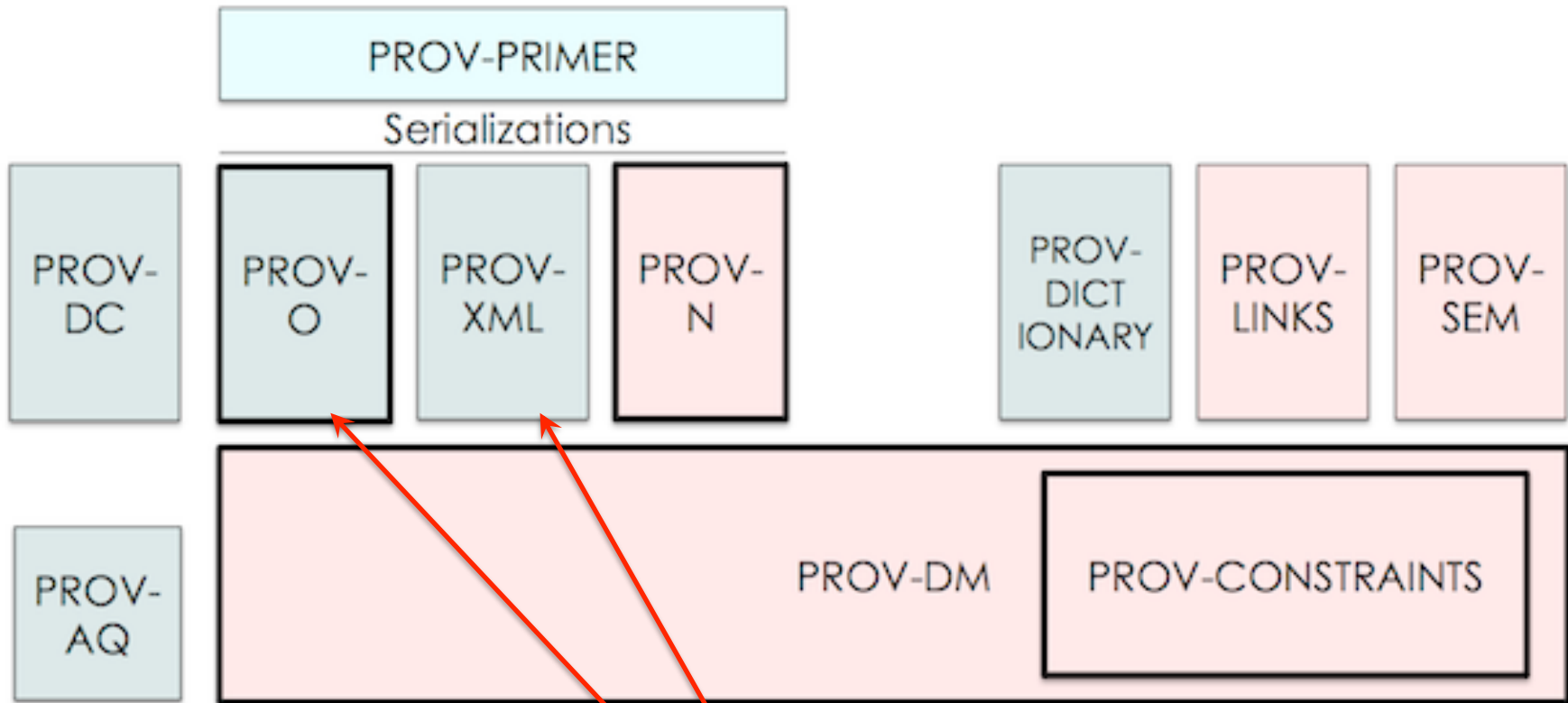


# PROV Relations (from PROV-DM)

|         |          | Object  |  |                                  |                                   |       |
|---------|----------|---|--|----------------------------------|-----------------------------------|-------|
|         |          | Entity  |  | Activity                         |                                   | Agent |
| Subject | Entity   | <b>WasDerivedFrom</b><br><u>Revision</u><br><u>Quotation</u><br><u>PrimarySource</u><br><u>AlternateOf</u><br><u>SpecializationOf</u><br><u>HadMember</u> | <b>WasGeneratedBy</b><br><u>WasInvalidatedBy</u> | <i>R</i><br><i>T</i><br><i>L</i> | <b>WasAttributedTo</b>            |       |
|         | Activity | <b>Used</b><br><u>WasStartedBy</u><br><u>WasEndedBy</u>   | <b>WasInformedBy</b>                             | <i>R</i><br><i>T</i><br><i>L</i> | <b>WasAssociatedWith</b> <i>R</i> |       |
|         | Agent    | —   | —  | —                                | <b>ActedOnBehalfOf</b>            |       |



# Complete Family of PROV Specifications (from PROV Overview)



## Conclusions

- W3C PROV specifications are designed for use in domains with large datasets for scalability and extensibility
  - Web, Healthcare
- Provides a good bridge to other existing W3C recommendations (e.g. XML, RDF)
- Additional Resources:
  - PROV Overview: <http://www.w3.org/TR/prov-overview/>
  - PROV FAQs: <http://www.w3.org/2001/sw/wiki/PROV-FAQ>
  - PROV Tools: <http://www.w3.org/2001/sw/wiki/PROV>
  - PROV Mailing List: [public-prov-comments@w3.org](mailto:public-prov-comments@w3.org)

## Acknowledgements

- W3C Provenance Working Group
- Many of the diagrams and examples used in this presentation are taken from the PROV specifications
- W3C PWG Chairs: Luc Moreau and Paul Groth
- PWG Members: Paolo Missier, Timothy Lebo, Khalid Belhajjame, Graham Klyne, Stian Soiland-Reyes, Stephan Zednic, Yolanda Gil, Simon Miles, Jun Zhao, Daniel Garijo, Helena Deus and all W3C Provenance Working Group Members