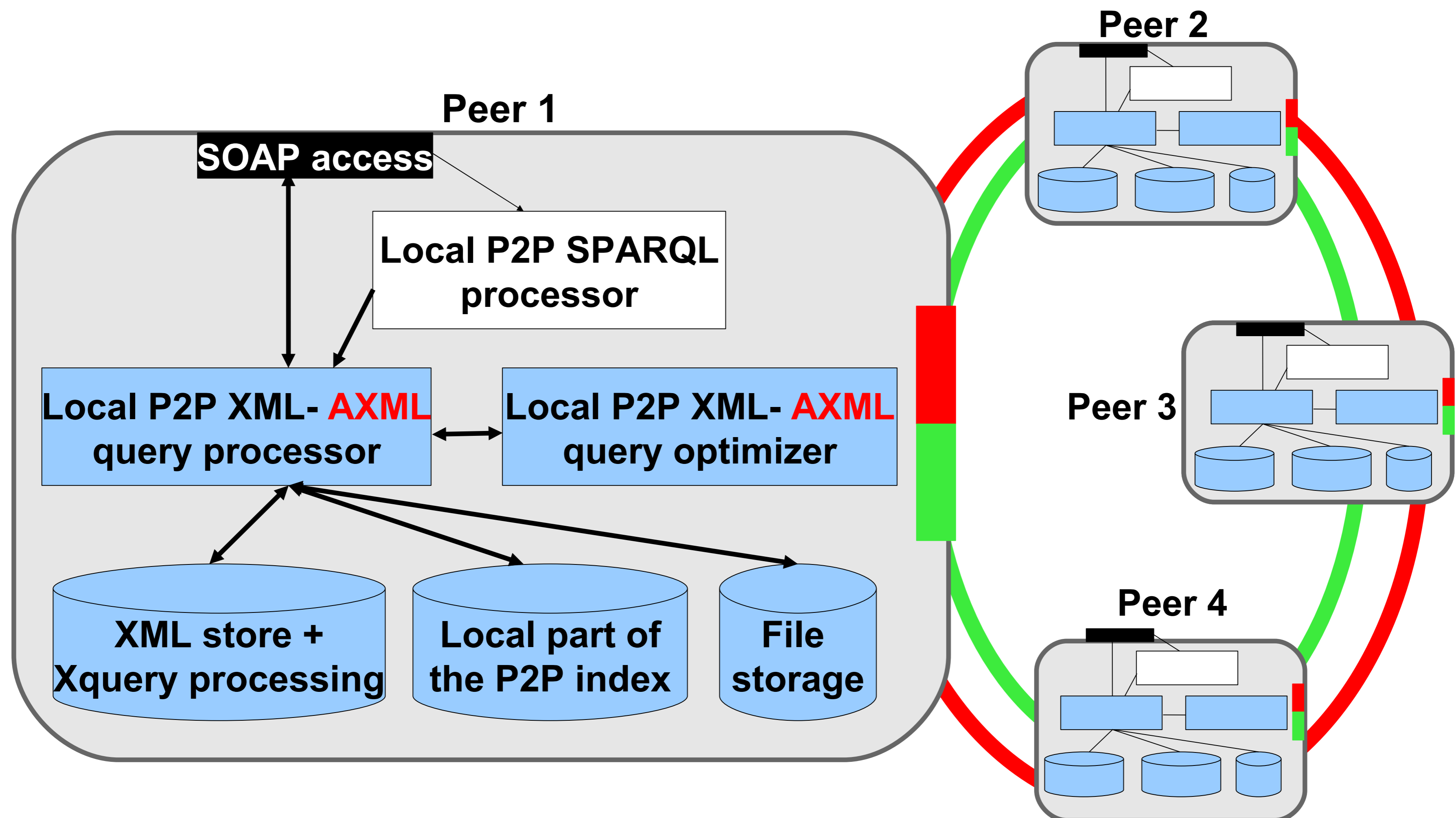


WebContent: Efficient P2P Warehousing of Web Data

S. Abiteboul¹, T. Allard², P. Chatalic¹, G Gardarin², A. Ghitescu¹,
F. Goasdoué¹, I. Manolescu¹, B. Nguyen², M. Ouzara¹, A Somani^{1,4},
N. Travers³, G. Vasile¹ and S. Zoupanos¹

¹INRIA Saclay – Île-de-France & LRI – Univ. Paris-Sud, ²Univ. Versailles-Saint-Quentin, ³CNAM
and ⁴IIT Bombay



Multiple DHT-based XML indexing models:

- Keyword-based indexing based on structure and content (KadoP).
- Path-based indexing supporting value inequalities (PathFinder)

Querying:

- Query decomposition into simpler queries (algebraic XQuery compiler TGV)
- Data flow reorganization to best exploit the index (OptimAX)
- Service invocation and result composition (ActiveXML)

Original Query:

```
for $d1 in //Col1/Book, $d2 in //Col2/Book
where $d1/PageNo < 400
and $d2/Author = "Ullman"
and $d2/Code = $d1/Code
return $d2/Title
```

Join query:

```
for $x_0 in $in_0//res, $x_1 in $in_1//res
where $x_0//Code = $x_1//Code
return $x_0//Title
```

Tree pattern query:

```
<Col2><Book returned="true">
  <Author>Ullman</Author>
  <Code/>
</Book></Col2>
```

Query with inequality:

```
for $d1 in //Col1/Book
where $d1/PageNo < 400
return <res>{$d1/Code}</res>
```